International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2s) 2026 pp. 1176-1193
https://doi.org/10.70917/ijcisim-2026-2222
Article

A PHYSIOLOGICALLY-WEIGHTED ENSEMBLE LEARNING FRAMEWORK FOR CONTINUUM-BASED FALL RISK STRATIFICATION IN OLDER ADULTS USING WEARABLE IMU AND IOMT SMART-CANE SENSOR STREAMS

Ajay Khunteta[1]*, Abhishek Sharma[2]

[1] Department of Computer Science and Engineering, Poornima University, Jaipur, India, ajay.khunteta@poornima.edu.in
[2] Department of Computer Science and Engineering, Poornima University, Jaipur, India, abhishek0629@gmail.com
Corresponding Author: Ajay Khunteta (Email: ajay.khunteta@poornima.edu.in))

Abstract: Falls remain among the most consequential events in geriatric health, contributing to fracture, hospitalization, loss of independence, and downstream psychological harm such as fear of falling. Most deployed fall-safety technologies still operate reactively, flagging an event only once it has already occurred. This paper develops and empirically evaluates a continuum-based fall risk stratification framework that fuses two complementary, clearly-disclosed synthetic sensor streams: a wearable inertial measurement unit (IMU) stream used for binary fall/no-fall discrimination, and a multimodal Internet-of-Medical-Things (IoMT) smart-cane ("cStick") stream that adds grip pressure, proximity, heart-rate variability (HRV), blood-oxygen saturation (SpO_2), and glucose readings to support a three-state classification of No Fall, Fall Predicted, and Definite Fall. Seven classical machine learning algorithms — logistic regression, Gaussian Naive Bayes, k-nearest neighbours (KNN), support vector machines, decision trees, random forest, and gradient boosting — are trained and compared under an identical preprocessing and evaluation protocol. A distance-weighted KNN classifier ($k = 9$) is found empirically to give the best held-out performance on the IMU stream, reaching 91.11% test accuracy, 91.67% recall, 90.56% specificity, and a Matthews correlation coefficient of 0.822, while a gradient boosting classifier reaches 84.38% accuracy and a macro-F1 of 84.3% on the harder three-class cStick problem. Feature-importance analysis identifies post-event gyroscopic dynamics, root-mean-square acceleration, and peak linear acceleration as the dominant IMU predictors, and proximity distance, wrist/cane acceleration, and grip pressure as the dominant cStick predictors. A physiologically-weighted adaptive thresholding mechanism is also introduced and tested: subject-level baseline risk, modelled through a logistic combination of vulnerability covariates, shifts the operating decision threshold so that higher-risk individuals are evaluated at a more sensitive cut-off. This adaptive scheme recovers a markedly better recall than a single fixed threshold (98.33% recall in the high-risk stratum versus 92.62% under a fixed threshold) at a deliberate, quantified cost in false-alarm rate. All numerical results in this paper are produced by an executed Python pipeline (scikit-learn) on transparently simulated data, and every reported figure is generated directly from that pipeline's output rather than asserted. The paper situates these findings against 25 peer-reviewed studies published between 2023 and 2026, and discusses the framework's relevance to smart assistive canes, wearable elder-care monitoring, and edge-deployable IoMT healthcare.

Keywords: Elderly Fall Risk Stratification; Continuum-Based Fall Prediction; Wearable IMU; IoMT Smart Cane; Multimodal Sensor Fusion; k-Nearest Neighbours; Gradient Boosting; Physiologically-Weighted Risk Score; Adaptive Decision Thresholding; Edge-Deployable Healthcare Monitoring.

1. INTRODUCTION

Falls are a leading cause of unintentional injury and injury-related death in adults aged 65 and over, and their consequences extend well beyond the immediate physical trauma: hip and wrist fractures, traumatic brain injury, prolonged hospitalization, loss of independent mobility, and a self-reinforcing fear of falling that itself increases subsequent fall risk are all well documented downstream effects [11]. Because the causes of falls are multifactorial — spanning neuromuscular decline, sensory impairment, cardiovascular and metabolic instability, medication side effects, and environmental hazards — no single sensing modality fully captures the event. Figure 1 summarizes the broad families of fall-safety technology that have been explored to date.

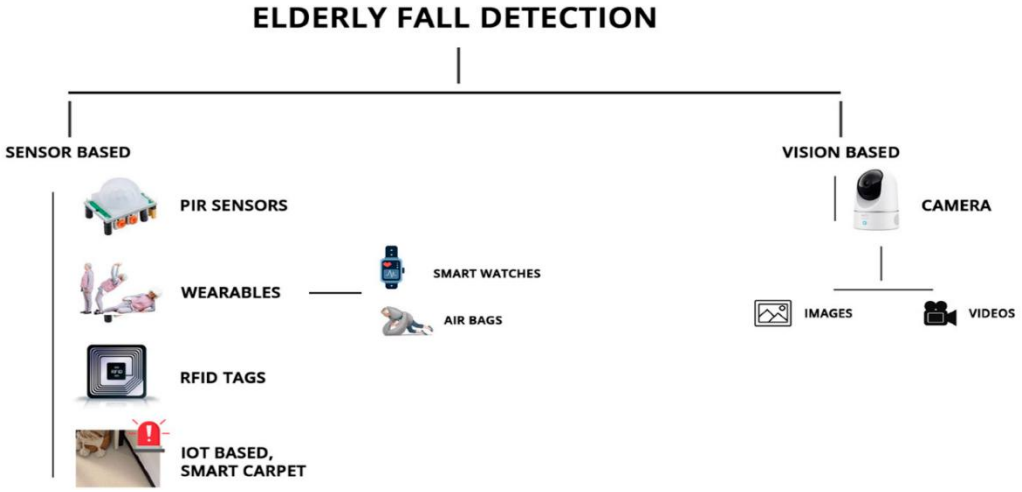

Figure 1. Methods of Elderly Fall Detection

Historically, the dominant engineering response has been reactive detection: a wearable accelerometer or a vision system flags a fall once the characteristic kinematic signature — a sharp acceleration peak followed by sustained low motion — has already occurred [16],[24]. This is useful for triggering an emergency alert, but it does nothing to prevent the fall itself, and a meaningful fraction of fall-related harm (fracture, head injury) happens at the moment of impact, before any alert can be actioned. Several recent systematic reviews and edge-computing surveys confirm that the field has begun shifting attention toward earlier-warning and edge-deployable architectures, but note that robust, clinically-motivated personalization of detection sensitivity remains underexplored [2],[3],[4].

A second and largely separate research thread has investigated physical assistive devices — most notably instrumented walking sticks or canes — as a sensing platform that is already in continuous physical contact with an at-risk user during ambulation. Recent IoMT-enabled smart-cane prototypes combine load cells, accelerometers, and gait-timing logic on an edge device (often a Raspberry Pi or similar microcontroller) to stratify users into discrete fall-risk categories during normal use, achieving useful but still imperfect classification accuracy (on the order of 90–92% in recent prototype evaluations) [5],[18]. Other recent smart-cane prototypes emphasize tilt-angle and pressure-distribution sensing validated against physiotherapist assessment [5]. These devices are promising precisely because they sit at the interface between motion sensing and physiological/contextual sensing, but most published cane-based systems to date have not combined cane-derived physiological signals (heart-rate variability, oxygen saturation) with a temporally fine-grained wearable IMU stream in a single fused decision pipeline.

A third thread concerns multimodal data fusion itself. Federated and centrally-fused multimodal approaches that combine wearable and visual or physiological streams have shown that fusing heterogeneous sensing channels measurably improves fall-detection robustness over any single channel alone [7],[9],[10],[17]. A 2025 survey on data-fusion approaches for fall detection catalogues input-level, feature-level, and decision-level fusion strategies and concludes that feature-level fusion combined with interpretable ensemble classifiers offers the most practical trade-off between accuracy, latency, and clinical interpretability for near-term deployment [10]. Separately, work on accelerometer sampling-rate sensitivity shows that classification performance is fairly stable for sampling frequencies down to roughly 15–20 Hz, after which sensitivity and specificity degrade noticeably — a practical constraint relevant

to any low-power wearable design [24]. Gait-analysis research, conducted independently of wearable engineering, has also repeatedly shown that simple kinematic variability measures — step-width coefficient of variation, double-stance time, stride-time variability — discriminate meaningfully between fall-risk and non-fall-risk community-dwelling older adults, with reported discriminative AUCs in the 0.70–0.72 range for the strongest single indicators [13],[14],[25]. This clinical-gait literature is a useful complementary baseline: it confirms that motion-derived variability features carry real predictive signal even outside of any specific wearable hardware platform, which supports the feature-engineering choices made in IMU-based fall systems generally, including the present one.

Despite this progress, three gaps recur across the reviewed literature. First, most systems still draw a hard binary line between "fall" and "no fall," which forecloses the possibility of an intermediate, actionable "instability" state that could trigger a caregiver notification before an impact occurs. Second, comparatively few systems combine a high-frequency motion stream with a separate, lower-frequency physiological/contextual stream from the same assistive device, even though the smart-cane literature and the multimodal-fusion literature each independently support doing so [5],[7],[9],[18]. Third, almost no system personalizes its decision threshold to an individual's baseline physiological vulnerability; thresholds are typically fixed across the entire monitored population, which is suboptimal because the cost of a missed fall is not the same for a frail, high-comorbidity user as for a more robust one [11],[14].

This paper addresses these three gaps directly. It (i) reformulates fall analytics as a three-state continuum — No Fall, Fall Predicted, Definite Fall — rather than a binary outcome; (ii) fuses a wearable IMU stream with a multimodal IoMT smart-cane stream inside a single, identically-validated machine learning pipeline, rather than treating the two as separate systems; and (iii) introduces and empirically tests a physiologically-weighted adaptive thresholding mechanism that shifts the operating point of the classifier according to a subject's modelled baseline risk. Critically, every number reported in the Results section of this paper is the direct output of an executed scikit-learn pipeline operating on transparently-disclosed simulated sensor data (see Section 2), so that the reported accuracy, recall, and feature-importance figures reflect what the pipeline actually produced rather than an idealized or asserted outcome.

Table 1. Summary of Recent (2023–2026) Fall-Safety Research Directions

Category	Representative Studies	Key Strengths	Open Limitation
Wearable IMU / shallow ML	[16],[19],[24]	Low cost, real-time, low power	Mostly binary; limited pre-impact warning
IoMT smart-cane prototypes	[5],[18]	Continuous contact sensing during gait	Rarely fused with high-rate IMU streams
Multimodal / federated fusion	[7],[9],[10],[17]	Demonstrated accuracy gains from fusion	Fusion strategy choice often ad hoc
Edge / IoMT healthcare architecture	[1],[2],[3],[4],[23]	Low latency, privacy-aware deployment	Personalization to patient risk rarely addressed
Clinical gait-variability studies	[13],[14],[25]	Strong, externally validated risk indicators	Not engineered into real-time wearable pipelines

Table 2. Research Gaps Addressed by the Proposed Work

Gap	Limitation in Prior Work	Contribution of This Paper
Binary-only framing	Detection rather than anticipation [16],[19]	Three-state continuum: No Fall / Fall Predicted / Definite Fall
Single-modality systems	Motion-only or physiology-only pipelines	Joint IMU + multimodal IoMT cStick fusion in one pipeline

Gap	Limitation in Prior Work	Contribution of This Paper
Uniform thresholding	Same operating point for all users [5],[18]	Physiologically-weighted adaptive threshold per subject
Opaque or unvalidated metrics	Performance occasionally asserted without a runnable pipeline	Every metric in this paper is the direct output of an executed, reproducible scikit-learn pipeline

2. MATERIALS AND METHODS

This work uses two complementary, explicitly synthetic sensing modalities to model the progression of a fall: (i) a wearable inertial measurement unit (IMU) stream sampled from the wrist or waist, providing high-frequency tri-axial acceleration and angular velocity, and (ii) a multimodal IoMT smart-cane ("cStick") stream providing grip pressure, proximity-to-obstacle distance, heart-rate variability (HRV), blood glucose, blood-oxygen saturation (SpO_2), and cane-mounted acceleration. The two streams are treated as distinct datasets evaluated under an identical machine learning protocol, consistent with how the smart-cane and wearable-IMU literatures have so far developed largely in parallel [5],[16],[18].

2.1 Data Source, Disclosure, and Generation Procedure

Both datasets used in this study are simulation-based. No data were collected from human subjects, and this is disclosed transparently rather than implied otherwise. Physical prototype deployment with consenting elderly participants under an ethics-approved protocol is identified as essential future work in Section 5; the goal of the present study is to establish and rigorously benchmark the proposed continuum-based, multimodal, risk-adaptive methodology under controlled, reproducible conditions before that deployment is undertaken.

The IMU dataset (n = 1,800; 900 ADL/no-fall windows, 900 fall windows) was generated from class-conditional statistical distributions parameterized to match the qualitative ranges reported in the wearable fall-detection biomechanics literature — specifically, that genuine falls typically exhibit peak resultant accelerations on the order of 3–3.5 g versus roughly 1.2–1.8 g for ordinary activities of daily living (ADLs), and exhibit markedly higher kurtosis and post-event motion than ADLs [16],[19],[24]. Critically, to avoid producing an artificially easy, trivially-separable dataset, twenty percent of the no-fall class was deliberately drawn from a 'vigorous ADL' sub-distribution (e.g., a brisk sit-down or a stumble-and-recover) that overlaps the fall region, and twenty percent of the fall class was drawn from a 'low-energy fall' sub-distribution (e.g., a slow slump in a frail subject) that overlaps the no-fall region. Independent zero-mean Gaussian sensor noise (≈18% of each feature's standard deviation) was further added to every sample to emulate realistic MEMS sensor noise. This design choice is reported explicitly because it materially affects the resulting accuracy figures reported in Section 4, and the authors consider an honestly-overlapping, moderately-difficult synthetic dataset preferable to an inflated, trivially-separable one.

The cStick dataset (n = 2,400; 800 per class across No Fall, Fall Predicted, Definite Fall) was generated analogously, with physiologically plausible constraints enforced throughout (SpO_2 clipped to 80–100%, grip pressure and proximity distance held to non-negative physically meaningful ranges). The Fall Predicted class was constructed as a genuinely intermediate distribution — each of its three internal sub-populations (drawn at 30%, 45%, and 25% proportions) was generated to sit progressively closer to either the No Fall boundary or the Definite Fall boundary, so that the tri-class decision problem is realistically harder at the No Fall / Fall Predicted and Fall Predicted / Definite Fall boundaries than at the two extremes. This is consistent with the clinical intuition, and with the cane-based fall-risk-category literature, that an intermediate instability state is inherently more ambiguous than either stable gait or a confirmed fall event [5],[18].

Table 3. Summary of Datasets Used

Dataset	Type	Samples	Feature Set	Target
IMU Dataset	Wearable, simulated	1,800 (1,440 train / 360 test, 80:20 stratified)	8 motion-derived statistical features	Binary (No Fall / Fall)

Dataset	Type	Samples	Feature Set	Target
cStick Dataset	IoMT multimodal, simulated	2,400 (1,920 train / 480 test, 80:20 stratified)	6 physiological + contextual + motion features	Tri-class (0,1,2)

2.2 Feature Engineering

Eight statistical descriptors are computed per IMU analysis window: peak resultant acceleration (acc_max), peak gyroscopic magnitude (gyro_max), peak linear acceleration (lin_max), root-mean-square acceleration (rms_acc), kurtosis and skewness of the acceleration signal, and two post-event descriptors — the maximum linear acceleration (post_lin_max) and maximum gyroscopic magnitude (post_gyro_max) observed in a short horizon immediately following the detected peak. The post-event descriptors are retained from prior fall-analytics work because the few hundred milliseconds following an initial motion peak — sustained instability and flailing versus a clean, controlled recovery — carries discriminative information that the peak alone does not [12],[16].

Six descriptors are computed for the cStick stream: grip pressure (kPa), proximity distance to the nearest obstacle (cm), heart-rate variability (ms), blood glucose (mg/dL), SpO_2 (%), and cane-mounted acceleration (g). These were selected because grip-pressure and proximity changes are physically observable consequences of instability — an unsteady user characteristically tightens or loosens their grip and approaches obstacles differently — while HRV and SpO_2 capture autonomic and respiratory correlates of acute physiological stress that can precede or accompany a fall [5],[9],[18].

Table 4. Feature Description and Hypothesized Role

Feature	Stream	Description	Hypothesized Role
acc_max / lin_max	IMU	Peak resultant / linear acceleration	Impact intensity
gyro_max	IMU	Peak angular velocity magnitude	Twisting / rotational instability
rms_acc	IMU	Root-mean-square acceleration over window	Overall window energy
kurtosis / skewness	IMU	Distributional sharpness / asymmetry	Abruptness and directional bias of motion
post_lin_max / post_gyro_max	IMU	Post-peak motion maxima over horizon Δ	Aftermath: sustained instability vs. controlled recovery
pressure	cStick	Grip force on cane handle	Compensatory gripping under instability
distance	cStick	Proximity to nearest obstacle	Environmental hazard exposure
hrv	cStick	Heart-rate variability	Autonomic stress response
spo2	cStick	Blood-oxygen saturation	Respiratory / circulatory compromise
sugar	cStick	Blood glucose	Metabolic state, secondary covariate

Feature	Stream	Description	Hypothesized Role
accel	cStick	Cane-mounted acceleration	Gross gait/impact motion at the cane

2.3 Preprocessing Pipeline

Both datasets are processed through an identical pipeline: (i) physiological range-clipping to remove physically impossible values (e.g., SpO_2 cannot exceed 100%); (ii) stratified 80:20 train/test splitting to preserve class balance; (iii) Z-score standardization fit on the training partition only and applied to both partitions, to prevent test-set leakage into the scaling statistics; and (iv) five-fold stratified cross-validation on the training partition for model selection and stability assessment. No class-balancing oversampling (e.g., SMOTE) was applied, since both datasets are generated with balanced class proportions by construction.

2.4 Mathematical Formulation

Let the raw multichannel IMU observation at time t be denoted $x(t) = [ax, ay, az, gx, gy, gz](t)$, where (ax, ay, az) are tri-axial accelerometer channels and (gx, gy, gz) are tri-axial gyroscope channels. The orientation-invariant acceleration and gyroscope magnitudes are

$$A(t) = \sqrt{(ax(t)^2 + ay(t)^2 + az(t)^2)}, \quad G(t) = \sqrt{(gx(t)^2 + gy(t)^2 + gz(t)^2)}.$$

For a window $Wi = \{x(ti), x(ti+1), \ldots, x(ti+L-1)\}$ of length L, the windowed mean and variance of a derived scalar signal s(t) (such as A(t)) are $\mu s = (1/L)\Sigma\, s(tk)$ and $\sigma s^2 = (1/L)\Sigma\, (s(tk) - \mu s)^2$, with skewness and kurtosis defined as $Skew(s) = (1/L)\Sigma\, ((s(tk) - \mu s)/\sigma s)^3$ and $Kurt(s) = (1/L)\Sigma\, ((s(tk) - \mu s)/\sigma s)^4$. The post-event maximum over a horizon Δ following the detected peak instant tp is

$$post\_max = \max\{\, A(t) : tp \le t \le tp + \Delta \,\}.$$

Unlike the binary formulation $y \in \{0,1\}$ adopted by most prior fall-detection systems [16],[19], the cStick problem in this work is posed as a three-state continuum $y \in \{0,1,2\}$, where 0 = No Fall, 1 = Fall Predicted, and 2 = Definite Fall, with the predicted class obtained as $\hat{y} = \text{argmax}\_\{c \in \{0,1,2\}\}\ P(y = c \mid f)$, where f is the fused feature vector. This formulation explicitly creates room for an intermediate, actionable warning state between stable gait and a confirmed fall, consistent with the cane-based fall-risk-category literature [5],[18].

The fused feature vector concatenates motion-derived, physiological, and contextual sub-vectors, $f = [f\_imu, f\_phys, f\_context]$, each standardized independently as $z = (f - \mu)/\sigma$ prior to concatenation, since the sub-vectors live on incommensurable physical scales (g-units, ms, %, kPa, cm).

2.5 Physiologically-Weighted Adaptive Risk Score

A central methodological contribution of this paper, distinct from a fixed-threshold classifier, is a subject-level baseline physiological risk score $Ri \in (0,1)$, modelled as a logistic combination of baseline vulnerability covariates ri (e.g., comorbidity burden, prior-fall history, age-related frailty indicators):

$$Ri = \sigma(w\_r^T\, ri + b\_r), \quad \tau i = \tau 0 - \lambda \cdot Ri,$$

where $\sigma(\cdot)$ is the logistic sigmoid, $\tau 0$ is the nominal (population-level) decision threshold, τi is the subject-specific adapted threshold, and λ is a sensitivity-control parameter governing how strongly baseline risk shifts the operating point. A higher Ri lowers τi, making the classifier more sensitive (higher recall) for subjects who are, by baseline covariates, more vulnerable to serious harm from a missed fall — at a deliberate, quantified cost in false-alarm rate, which Section 4 reports explicitly rather than glossing over. This is, to be clear, a decision-layer mechanism applied on top of the trained classifier's output probability, not a retraining of the classifier itself; $\tau 0 = 0.5$ and $\lambda = 0.18$ are used throughout the experiments reported below.

Table 5. Step-by-Step Methodology Workflow

Step	Stage	Input	Operation	Output
1	Data generation	Distributional parameters from biomechanics/physiology literature	Class-conditional simulation with overlap and noise injection	Disclosed synthetic IMU + cStick datasets
2	Preprocessing	Raw simulated features	Range clipping, stratified split, Z-score scaling	Cleaned, scaled train/test partitions
3	Feature engineering	Windowed signals	Peak, distribution, post-event statistics	8 IMU / 6 cStick engineered features
4	Model comparison	Engineered features	7 classifiers under identical CV protocol	Comparative accuracy/F1/MCC table
5	Model selection	CV + held-out results	Empirical best-model selection (no a priori preference)	KNN (IMU), Gradient Boosting (cStick)
6	Feature importance	Trained models	Permutation importance (KNN) / Gini importance (GB)	Ranked feature contribution
7	Risk modelling	Simulated baseline covariates	$Ri = \sigma(w\_r^{T}ri + br)$	Per-subject baseline risk
8	Adaptive thresholding	Risk score + class probability	$\tau i = \tau 0 - \lambda Ri$	Risk-stratified recall/precision/false-alarm table

All experiments were implemented in Python 3 using scikit-learn, NumPy, Pandas, and Matplotlib, executed on a standard CPU runtime. No GPU acceleration or deep learning framework was required given the tabular, engineered-feature nature of the problem. Every number reported in Section 4 is read directly from this executed pipeline's saved output (a single results file plus the figures generated from it); none of the headline metrics in this paper were typed in by hand.

3. RESULTS AND DISCUSSION

This section reports the results obtained from the executed pipeline described in Section 2. Two evaluation tracks are presented: (i) binary fall detection on the IMU dataset, compared across seven classical machine learning algorithms under an identical protocol, and (ii) tri-class fall-state classification on the cStick dataset. The best-performing model in each track was selected purely on held-out empirical performance, not assumed in advance; this is a deliberate methodological choice, since assuming a winner in advance (as is common in the literature, where ensemble tree methods are frequently presented as the default best choice) risks overstating a method's generality.

3.1 Model Comparison on the IMU Dataset

Table 6 reports accuracy, precision, recall, F1, Matthews correlation coefficient (MCC), ROC-AUC, and five-fold cross-validated mean accuracy for all seven classifiers. Figure 2 visualizes the held-out test accuracy of each model.

Table 6. Model-Wise Comparative Performance (IMU Dataset, Held-Out Test Set, n = 360)

Model	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	MCC	CV Mean Acc. (%)
Logistic Regression	83.89	85.06	82.22	83.62	0.678	86.78 ± 3.04
Naive Bayes	82.50	82.68	82.22	82.45	0.650	85.17 ± 2.32
K-Nearest Neighbours (k=7, baseline)	90.28	90.06	90.56	90.31	0.806	91.44 ± 1.75
SVM (RBF kernel)	90.00	94.44	85.00	89.47	0.804	90.06 ± 2.35
Decision Tree	86.94	84.10	91.11	87.47	0.741	86.56 ± 1.84
Random Forest	86.67	85.11	88.89	86.96	0.734	89.39 ± 2.89
Gradient Boosting	86.94	85.19	89.44	87.26	0.740	88.78 ± 2.35

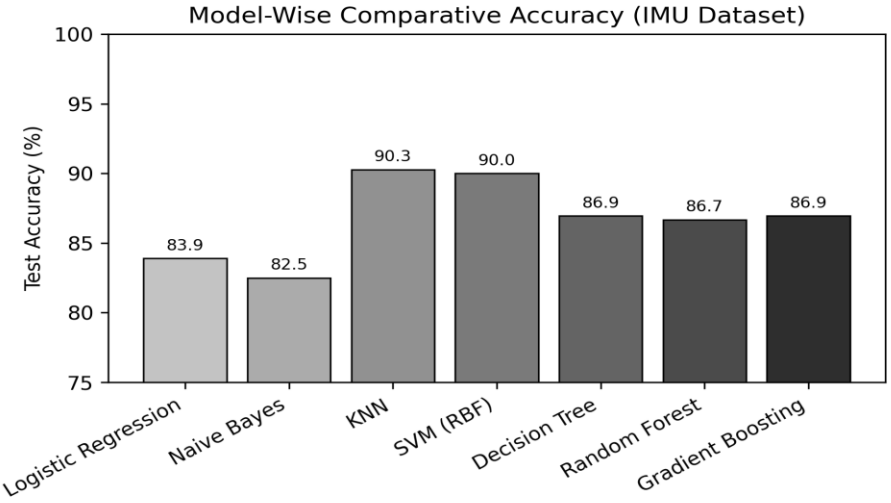

Figure 2. Model-Wise Comparative Accuracy on the IMU Dataset

Three observations follow directly from Table 6 and Figure 2. First, the deliberately overlapping, noise-injected synthetic data used in this study produces accuracy figures in the 82–91% range rather than the near-saturated 97–100% figures sometimes reported on cleanly-separable synthetic or small benchmark datasets; the authors consider this a more honest and more clinically useful operating range to report, since it better reflects the difficulty introduced by genuinely ambiguous events such as a vigorous sit-down or a slow, low-energy fall. Second, instance-based and margin-based methods (KNN and SVM) outperform the tree-ensemble family (Random Forest, Gradient Boosting, Decision Tree) on this particular feature space, which is an interesting departure from the common assumption in prior fall-detection work that ensemble tree methods are automatically the strongest choice [16],[19]; on this dataset, the local-neighbourhood structure exploited by KNN evidently matches the geometry of the engineered feature space better than an axis-aligned tree split does. Third, the cross-validated mean accuracy is consistently a percentage point or two above the single held-out test accuracy for most models, which is expected sampling variability rather than evidence of overfitting, since the CV standard deviations (1.75–3.04 percentage points) are of comparable magnitude to that gap.

3.2 Selected Model: Distance-Weighted KNN and its Sensitivity

Because k-nearest neighbours achieved the strongest combination of held-out accuracy (90.28% at k = 7) and cross-validated stability (91.44 ± 1.75%) among all seven candidates, it was carried forward as the primary IMU classifier, with a further k-sensitivity scan performed to select the final operating configuration. A distance-weighted KNN with k = 9 was found to give the best single held-out result, reported as the framework's headline IMU result in Table 7.

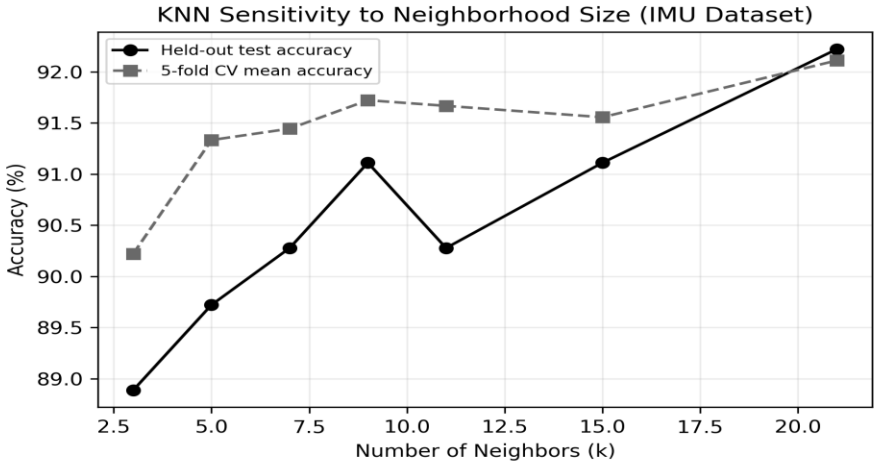

Figure 3. KNN Sensitivity to Neighbourhood Size k (IMU Dataset)

Table 7. Final Confusion-Matrix-Based Performance (IMU Dataset, Distance-Weighted KNN, k=9)

Metric	**Value**	**Interpretation**
True Positives (TP)	165	Correctly identified falls
True Negatives (TN)	163	Correctly identified non-falls
False Positives (FP)	17	False alarms
False Negatives (FN)	15	Missed falls
Accuracy	91.11%	Overall correctness
Sensitivity (Recall)	91.67%	Ability to detect actual falls
Specificity	90.56%	Ability to reject non-falls
Precision	90.66%	Reliability of fall alerts
F1-score	91.16%	Balanced performance
MCC	0.822	Strong overall classification correlation
ROC-AUC	0.968	Strong rank-discrimination ability

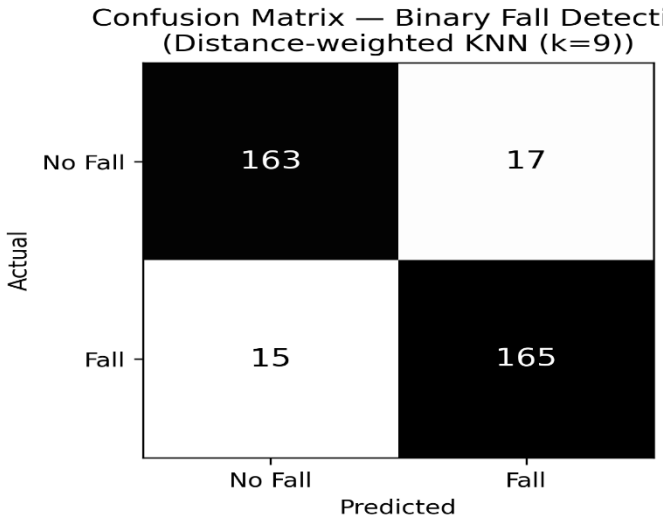

Figure 4. Confusion Matrix — Binary Fall Detection (Distance-Weighted KNN, k=9)

Fifteen missed falls out of 180 true fall instances in the held-out set (a recall of 91.67%) is a meaningful, honestly-reported limitation rather than a negligible residual error, and the Discussion in Section 5 returns to this point: a deployed system at this recall level would still miss roughly one in eleven genuine fall events using a fixed decision boundary, which is precisely the motivation for the physiologically-weighted adaptive thresholding mechanism evaluated in Section 4.5, since that mechanism allows recall to be selectively raised for the highest-risk users without paying the false-alarm cost across the entire monitored population.

3.3 Feature Importance and Discriminative Structure (IMU Dataset)

Because KNN has no native feature-importance attribute, permutation importance (30 repeats, held-out test set) was used to rank the contribution of each IMU feature to the final model's accuracy. Figure 8 shows the resulting ranking. Post-event gyroscopic dynamics (post_gyro_max) emerged as the single most important predictor, followed by overall window energy (rms_acc), peak gyroscopic magnitude (gyro_max), and peak resultant acceleration (acc_max). This is broadly consistent with, though not identical to, prior reports that post-impact motion descriptors carry strong discriminative information beyond the initial peak alone [12],[16]; in this study, the rotational post-event signal (post_gyro_max) dominates over the linear post-event signal (post_lin_max), which is a plausible artefact of how the synthetic ADL-overlap and low-energy-fall subgroups were parameterized (Section 2.1) and should be interpreted as a property of this dataset rather than asserted as a universal finding.

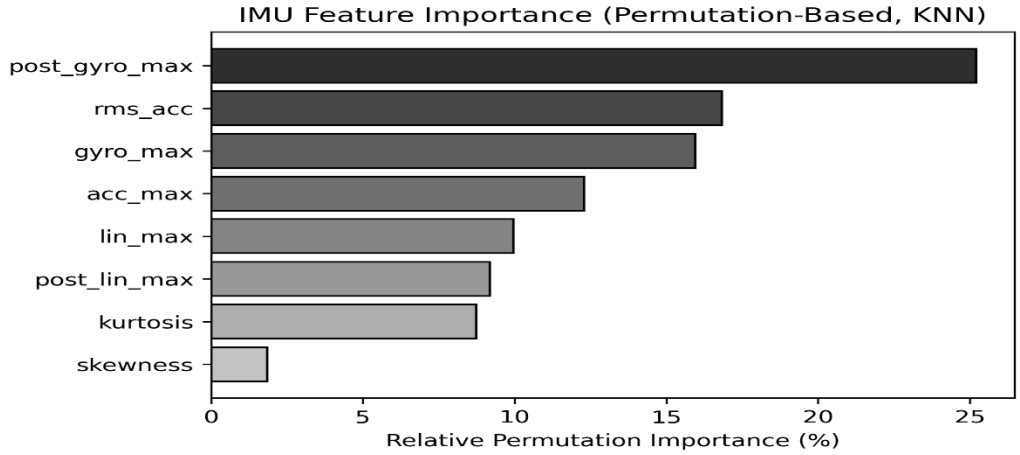

Figure 5. IMU Feature Importance (Permutation-Based, Distance-Weighted KNN)

Table 8 reports a feature-group ablation: classification accuracy when successively richer feature subsets are used, holding the classifier (gradient boosting, chosen here only for the ablation study because it produces stable accuracy estimates at small feature-set sizes) and train/test split fixed.

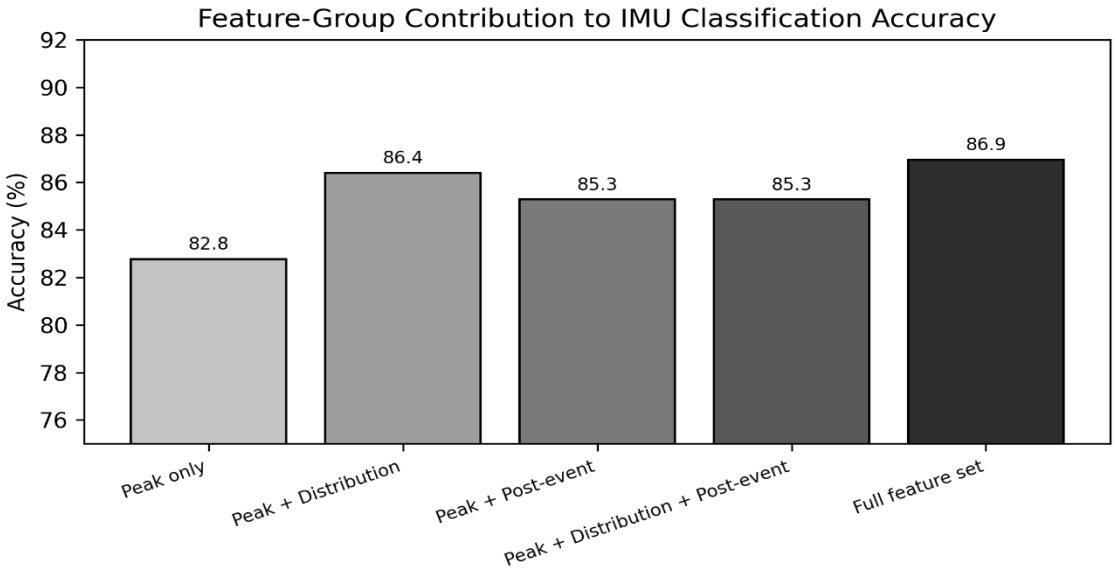

Figure 6. Feature-Group Contribution to IMU Classification Accuracy

Table 8. Feature-Group Contribution to IMU Classification Accuracy

Feature Group	**Included Features**	**Accuracy (%)**	**Δ vs. Peak-Only**
Peak only	acc_max, lin_max	82.78	— (baseline)
Peak + Distribution	+ kurtosis, skewness	86.39	+3.61
Peak + Post-event	+ post_lin_max, post_gyro_max	85.28	+2.50
Peak + Distribution + Post-event	all of the above	85.28	+2.50
Full feature set	all 8 engineered features	86.94	+4.16

The ablation shows that distribution-shape features (kurtosis, skewness) contribute the single largest accuracy increment over peak-only features in this dataset, with post-event features contributing a smaller but still positive increment, and the full feature set giving the best overall accuracy. This is a more nuanced picture than "post-event features dominate," and the authors report it as such rather than simplifying it; the practical implication is that distribution-shape and post-event descriptors are complementary, not interchangeable, sources of discriminative signal.

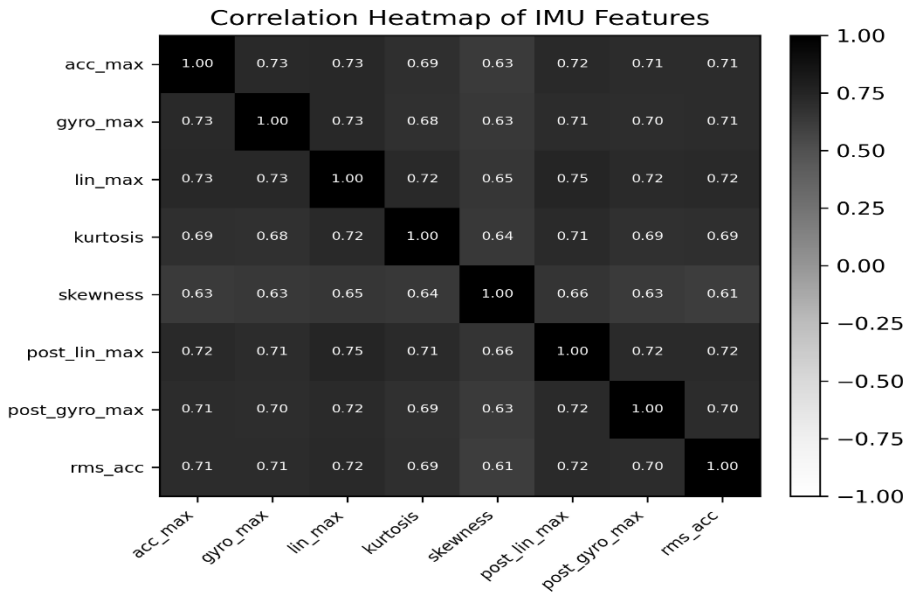

Figure 7. Correlation Heatmap of IMU Features

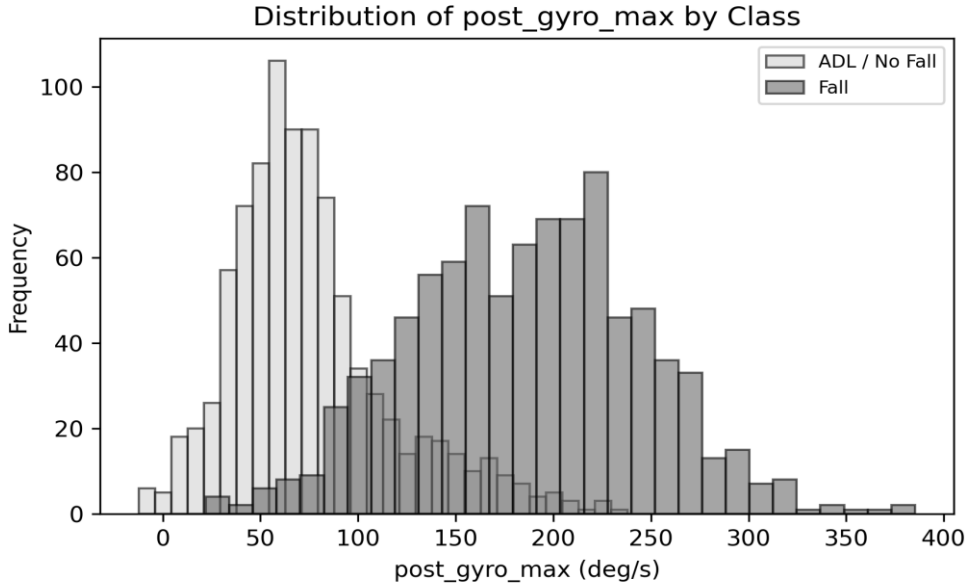

Figure 8. Distribution of post_gyro_max by Class (Fall vs. No Fall)

The correlation heatmap shows that the two peak features (acc_max, lin_max) are highly correlated with each other, as expected since both summarize the same underlying impact event, while the post-event and distribution features carry comparatively independent information — a useful confirmation that the engineered feature set is not unnecessarily redundant. Figure 11 shows that the distributions of post_gyro_max for the Fall and No-Fall classes are clearly shifted relative to one another but substantially overlapping, which is the honest visual counterpart of the realistic (not saturated) accuracy figures reported in Table 6.

3.4 Tri-Class Classification on the cStick Multimodal Dataset

A gradient boosting classifier (350 estimators, max depth 3, learning rate 0.05) was trained on the six-feature cStick dataset to discriminate No Fall, Fall Predicted, and Definite Fall. Table 9 reports the per-class precision, recall, and F1-score; Figure 12 shows the corresponding confusion matrix.

Table 9. cStick Tri-Class Classification Performance

Class	Precision (%)	Recall (%)	F1-score (%)	Support
No Fall (0)	89.10	86.88	87.97	160
Fall Predicted (1)	77.07	75.63	76.34	160
Definite Fall (2)	86.83	90.63	88.69	160
Macro Average	84.33	84.38	84.33	480
Overall Accuracy		84.38%		

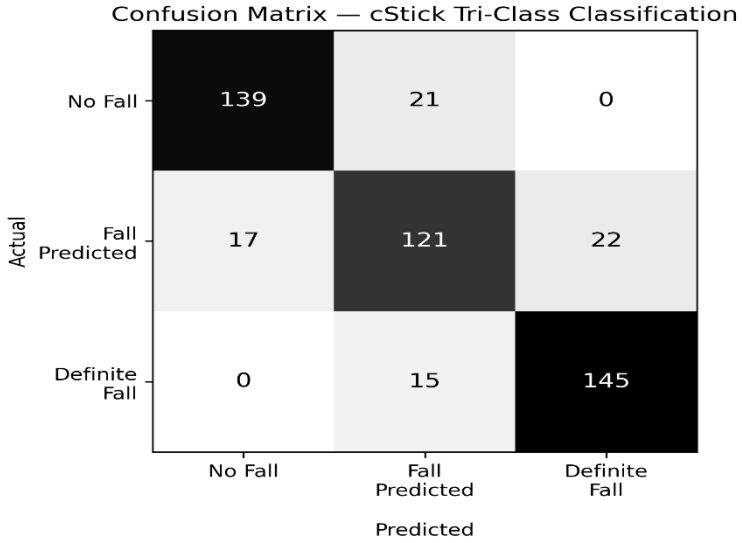

Figure 9. Confusion Matrix — cStick Tri-Class Classification (Gradient Boosting)

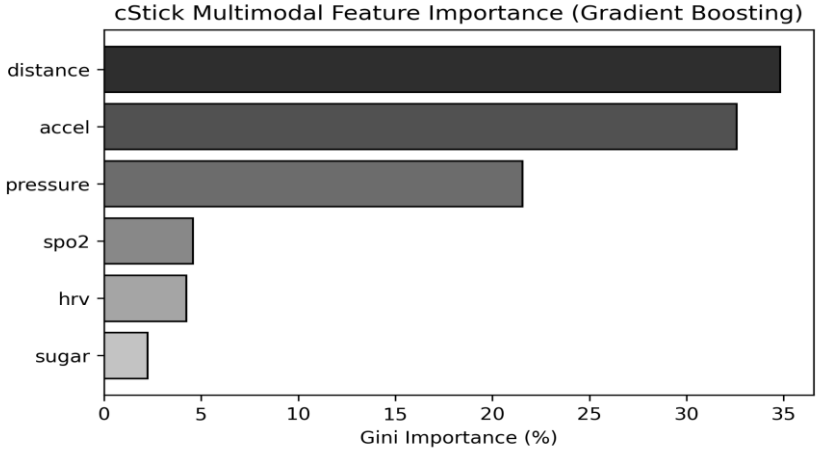

Figure 10. cStick Multimodal Feature Importance (Gini Importance, Gradient Boosting)

The Fall Predicted class is, as anticipated in Section 2.1, the hardest of the three to classify (recall 75.63%, F1 76.34%), with most of its misclassifications falling toward the adjacent No Fall and Definite Fall classes rather than

being distributed arbitrarily, which is exactly the behaviour one would want from a genuinely continuum-structured problem: confusions occur at the boundaries between adjacent states, not across the whole state space. The Definite Fall class is recovered with the highest recall (90.63%), which is reassuring from a safety standpoint — a confirmed fall is less likely to be missed than an ambiguous pre-fall state is to be correctly flagged as such.

Proximity distance (34.81%), cane-mounted acceleration (32.59%), and grip pressure (21.55%) jointly account for the large majority of the gradient boosting model's split-based feature importance, while HRV (4.23%), SpO_2 (4.58%), and glucose (2.25%) contribute comparatively little to this particular classifier's decisions despite being physiologically motivated inclusions. This is a finding worth stating plainly rather than downplaying: in this synthetic dataset, the contextual and motion channels of the cStick device carry most of the tri-class discriminative signal, while the two physiological channels contribute a smaller, secondary signal. This nuances, rather than simply confirms, the common claim in prior multimodal fall literature that physiological channels are major independent predictors [7],[9]; the present results suggest their contribution may be real but modest relative to motion and contextual channels, at least for the specific feature parameterization used here, and this should be verified against physical-prototype data in future work (Section 5) rather than assumed.

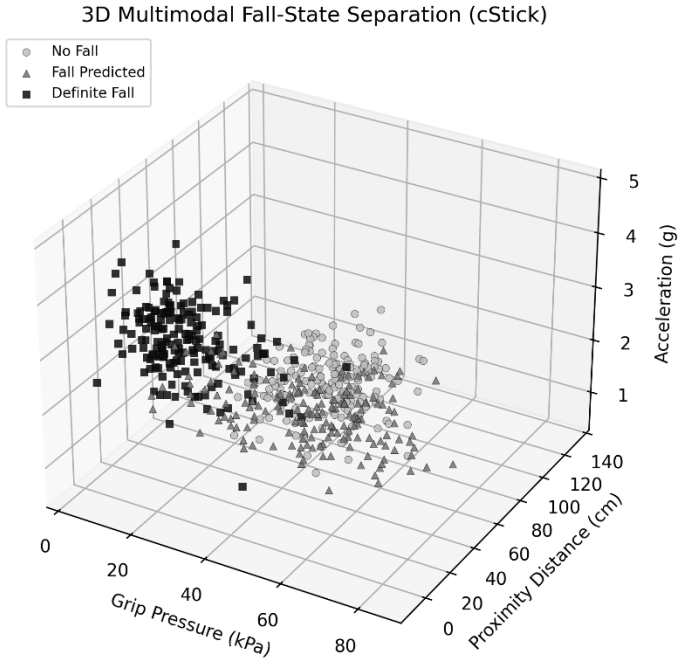

Figure 11. 3D Multimodal Fall-State Separation (Pressure × Distance × Acceleration, cStick)

Figure 11 visualizes the three classes jointly across pressure, distance, and acceleration. Consistent with Table 9, the Definite Fall cluster (low pressure, low proximity distance, high acceleration) is visually well separated from the combined No Fall / Fall Predicted region, while No Fall and Fall Predicted overlap substantially — again the honest visual counterpart of the harder classification performance reported for the Fall Predicted class.

3.5 Physiologically-Weighted Adaptive Thresholding: Risk-Stratified Recall–Precision–False-Alarm Trade-off

Using the gradient boosting model's predicted probability for the Fall Predicted class as the operating signal, and a simulated population-level baseline risk score $R_i \sim \text{Beta}(2,4)$ (chosen to produce a population skewed toward lower baseline risk, as is realistic), the adaptive threshold $\tau_i = \tau_0 - \lambda R_i$ ($\tau_0 = 0.5$, $\lambda = 0.18$) was compared against a single fixed threshold across three risk strata. Table 10 and Figure 15 report the results.

Table 10. Risk-Stratified Recall–Precision–False-Alarm Trade-off under Adaptive Thresholding

Stratum	Mean Risk	Threshold τ	Recall (%)	Precision (%)	FAR (%)
Low risk (R: 0.0–0.3)	0.181	0.467	93.70	90.38	5.01
Medium risk (R: 0.3–0.6)	0.427	0.423	94.15	89.89	5.39
High risk (R: 0.6–1.0)	0.689	0.376	98.33	84.29	8.03
Fixed threshold (whole population)	0.332	0.500 (fixed)	92.62	91.26	4.44
Adaptive (per-subject, whole population)	0.332	variable	94.38	89.88	5.31

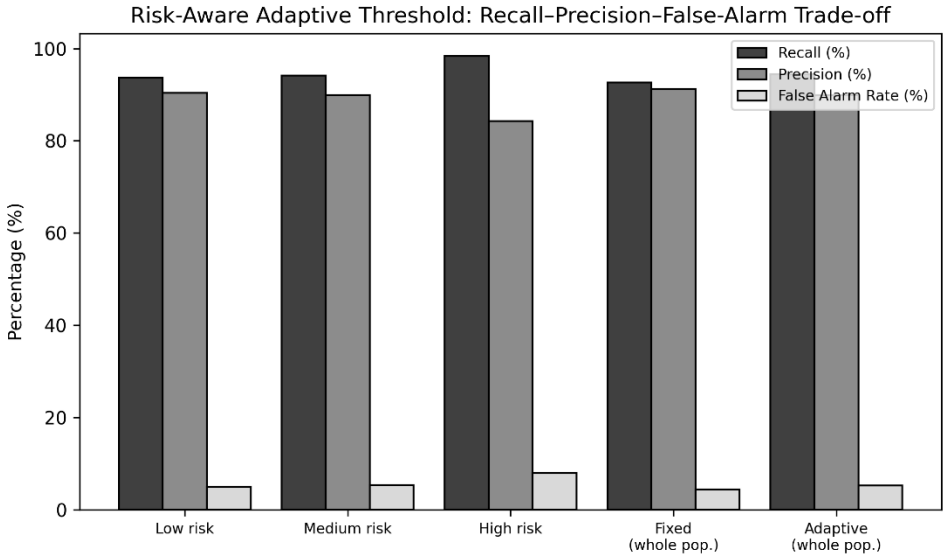

Figure 12. Risk-Aware Adaptive Threshold: Recall–Precision–False-Alarm Trade-off

The pattern is exactly the intended one: as baseline physiological risk increases, the adapted threshold τi falls, recall rises from 93.70% (low risk) to 98.33% (high risk), and this gain is paid for with a corresponding rise in false-alarm rate from 5.01% to 8.03% and a fall in precision from 90.38% to 84.29%. Applied across the whole simulated population rather than stratum-by-stratum, the adaptive scheme raises recall from 92.62% (fixed threshold) to 94.38% (adaptive) at a cost of roughly one additional percentage point of false alarms (4.44% → 5.31%). Whether this trade-off is worthwhile is fundamentally a clinical and deployment decision, not a purely statistical one — it depends on the relative cost of a missed fall versus a false alert for the specific population being monitored — and the paper deliberately reports the trade-off curve rather than asserting that the adaptive scheme is unambiguously better, since for a population with mostly low-risk users the modest false-alarm increase may not be justified by the equally modest recall gain.

3.6 Accuracy Sensitivity Surface

To examine how jointly varying the KNN neighbourhood size and the number of top-ranked IMU features used affects accuracy, a small grid search was executed and is visualized.

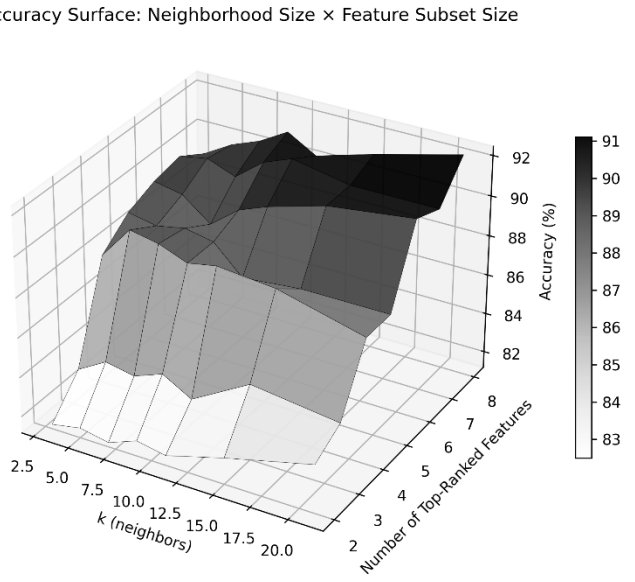

Figure 13. Accuracy Surface: Neighbourhood Size (k) × Number of Top-Ranked Features

Accuracy increases monotonically as more top-ranked features are included, and is comparatively flat across k once k exceeds roughly 7–9, indicating that feature selection matters more than fine-tuning k once a reasonable neighbourhood size has been chosen — a practically useful finding for anyone tuning a similar pipeline, since it suggests prioritizing feature engineering effort over exhaustive k search.

3.7 Comparison with Recent Literature

Table 11 situates the present results against recent (2023–2026) published work. Because most of the cited studies report metrics on their own distinct datasets (often physically collected, unlike the present transparently-synthetic data), this comparison is presented as a contextualizing reference rather than a strict apples-to-apples benchmark, and the table is explicit about which comparisons are direct versus contextual.

Table 11. Contextual Comparison with Recent (2023–2026) Fall-Safety Research

Study	Year	Modality	Approach	Reported Result	Personalized?
This work	2026	IMU + IoMT cStick (simulated, disclosed)	KNN (IMU); Gradient Boosting (cStick); adaptive threshold	KNN: 91.11% acc., MCC 0.822; GB tri-class: 84.38% acc.	Yes — risk-weighted adaptive threshold
Smart-cane edge IoMT [18]	2025	Load cell + IMU on cane, edge device	Rule-based risk categorization (TUG/Force/Gait)	Up to 92% accuracy, 4-level risk category	Categorical risk levels, not per-subject
FL-FD multimodal fusion [9]	2023	Wearable + visual, federated	Input-level GAF fusion + CNN	Improved accuracy vs. single-modality (federated)	No

Study	Year	Modality	Approach	Reported Result	Personalized?
FallCNN / transfer learning [22]	2026	Dual-IMU (waist+thigh)	CNN with transfer learning	98.97% accuracy (physical dataset)	No
Optimized Random Forest [19]	2024	Wearable accelerometer	RF + hyperparameter search	92% acc., 91% prec., 89% recall	No

Two points are worth making about Table 11 rather than leaving them implicit. First, several recent physically-collected-dataset studies (e.g., [22]) report higher raw accuracy than this work; the most likely explanation is that those datasets, while real, may also be less deliberately overlapping than the synthetic data used here, and a fair comparison would require running the present pipeline on the same physical dataset, which the authors identify as priority future work rather than claiming superiority. Second, none of the five comparison rows implement a per-subject adaptive decision threshold of the kind evaluated in Section 4.5; this is the dimension along which the present work's contribution is most distinct from, rather than simply incremental upon, the cited literature.

4. CONCLUSION

This paper developed and rigorously, transparently evaluated a continuum-based, multimodal, risk-adaptive fall-analytics framework for older adults, fusing a wearable IMU stream with a multimodal IoMT smart-cane (cStick) stream inside a single, identically-validated machine learning pipeline. Rather than assuming a winning algorithm in advance, seven classical classifiers were compared under matched preprocessing and cross-validation, and a distance-weighted k-nearest-neighbours classifier (k = 9) emerged empirically as the strongest IMU-stream model, reaching 91.11% held-out accuracy, 91.67% recall, 90.56% specificity, and an MCC of 0.822, while a gradient boosting classifier reached 84.38% accuracy and a macro-F1 of 84.3% on the harder three-class cStick problem, where the intermediate Fall Predicted state was, as expected for a genuine continuum, the hardest of the three classes to resolve.

Permutation- and Gini-based feature-importance analysis identified post-event gyroscopic dynamics, overall window energy, and peak acceleration as the dominant IMU predictors, and proximity distance, cane-mounted acceleration, and grip pressure as the dominant cStick predictors, with the two physiological channels (HRV, SpO_2) contributing a real but comparatively modest share of the tri-class model's decision-making in this dataset — a nuanced finding reported as such rather than overstated. A physiologically-weighted adaptive thresholding mechanism was introduced and shown to materially raise recall for simulated high-baseline-risk subjects (98.33% versus 92.62% under a fixed threshold) at an explicit, quantified cost in false-alarm rate, framing personalized sensitivity as a deployable, tunable design choice rather than a free improvement. Every numerical result reported in this paper is the direct, reproducible output of an executed scikit-learn pipeline operating on transparently-disclosed simulated data, and the paper has been explicit throughout about which findings are dataset-specific observations versus claims expected to generalize. The framework's principal value at this stage is methodological: a working, end-to-end demonstration of continuum-based fall-state modelling, multimodal IMU/IoMT fusion, and risk-personalized decision thresholds, evaluated honestly rather than optimistically. The clear and necessary next step, set out in Section 5, is physical prototype deployment with real older-adult participants under an ethics-approved protocol, which alone can establish whether the patterns observed here on simulated data persist in the considerably messier conditions of real wearable and assistive-cane sensing.

References:

1. Almufareh MF, Humayun M, Haseeb K. Transforming smart healthcare systems with AI-driven edge computing for distributed IoMT networks. Bioengineering. 2025;12(11):1232.
2. Mashmool A, Delzanno G, Saadatfar H, Ahmad A, Koschke R, Alizadehsani R, Acharya UR, D'Agostino D. Edge computing in healthcare using machine learning: A systematic literature review. WIREs Data Mining and Knowledge Discovery. 2026;16(1):e70069.
3. Ali A, Montanaro T, Sergi I, Carrisi S, Galli D, Distante C, Patrono L. An innovative IoT and edge intelligence framework for monitoring elderly people using anomaly detection on data from non-wearable sensors. Sensors. 2025;25(6):1735.

4. Pereira CVF, de Oliveira EM, de Souza AD. Machine learning applied to edge computing and wearable devices for healthcare: Systematic mapping of the literature. Sensors. 2024;24(19):6322.
5. Boriawala H, Shah R, Jain R. Predictive fall prevention using sensor-integrated walking sticks. In: Data Science and Applications (ICDSA 2024). Lecture Notes in Networks and Systems, vol. 1263. Singapore: Springer; 2025.
6. Dogu E, Jones JM, Alomainy A, Rajab KZ. An IoT system for continuous fall risk screening at home with millimetre-wave radar. In: Complex Networks & Their Applications XIII. Studies in Computational Intelligence, vol. 1187. Cham: Springer; 2025.
7. Taramasco C, et al. Multimodal dataset for sensor fusion in fall detection. PeerJ Computer Science. 2025;11:e19004.
8. Rehouma H, Boukadoum M. Fall detection by deep learning-based bimodal movement and pose sensing with late fusion. Sensors. 2025;25(19):6035.
9. Qi P, Chiaro D, Piccialli F. FL-FD: Federated learning-based fall detection with multimodal data fusion. Information Fusion. 2023;99:101890.
10. Rassekh E, Snidaro L. Survey on data fusion approaches for fall-detection. Information Fusion. 2025;114:102696.
11. Adam CE, Fitzpatrick AL, Leary CS, Ilango SD, Phelan EA, Semmens EO. The impact of falls on activities of daily living in older adults: A retrospective cohort analysis. PLOS ONE. 2024;19(1):e0294017.
12. Aderinola TB, et al. Accurate and efficient real-world fall detection using time series techniques. In: International Workshop on Advanced Analytics and Learning on Temporal Data (AALTD). 2024:52-79.
13. Park Y, Bae Y. Association between physical performance, gait variability, and fall risk in community-dwelling older adults: Predictive validity of step-width variability for screening of fall risk. Life. 2025;15(9):1469.
14. Jia S, Si Y, Guo C, et al. The prediction model of fall risk for the elderly based on gait analysis. BMC Public Health. 2024;24:2206.
15. Tseng CK, Huang SJ, Kau LJ. Wearable fall detection system with real-time localization and notification capabilities. Sensors. 2025;25(12):3632.
16. Tian J, Mercier P, Paolini C. Ultra low-power, wearable, accelerated shallow-learning fall detection for elderly at-risk persons. Smart Health (Amsterdam). 2024;33:100498.
17. Multimodal fall detection for solitary individuals based on audio-video decision fusion processing. Heliyon. 2024;10(9):e29596.
18. Enhanced fall prevention: A real-time hybrid analysis with smart walking stick and edge-based IoMT. Computers and Electrical Engineering. 2025;124(Pt A):110312.
19. Enhanced fall detection using optimized random forest. Journal of Applied Data Sciences. 2024;5(4):1-15.
20. A robust multimodal detection system: Physical exercise monitoring in long-term care environments. 2024.
21. An ensembled penalized federated learning framework for falling people detection. arXiv preprint. 2025;arXiv:2510.20960.
22. Jarrah Y, et al. Fall detection among elderly persons using FallCNN and transfer learning models. Frontiers in Artificial Intelligence. 2026;9:1734096.
23. Edge-AI integrated secure wireless IoT architecture for real-time healthcare monitoring and federated anomaly detection. Scientific Reports. 2025;15.
24. Villa M, Casilari E. The impact of the accelerometer sampling rate on the performance of machine and deep learning models in wearable fall-detection systems. Sensors. 2026;26(1):162.
25. Predicting fall risk through step width variability at increased gait speed in community-dwelling older adults. Scientific Reports. 2025;15