Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Data-Driven Identification of High-Risk Patient Groups for Falls Using SHAP-Space Profiling

**Matthias Schulte-Althoff**
Charité - University Medicine Berlin
**Peter Krappen**
Freie Universität Berlin
**Felix Biessmann**
Berliner Hochschule für Technik
**Sebastian Jäger**
Berliner Hochschule für Technik
**Rahel Gubser**
Charité - University Medicine Berlin
**Armin Hauß**
Charité - University Medicine Berlin
**Felix Balzer**
Charité - University Medicine Berlin
**Tim Kilgus**
Freie Universität Berlin
**Daniel Fürstenau**
daniel.fuerstenau@fu-berlin.de

Freie Universität Berlin

**Research Article**

**Keywords:** inpatient falls, risk prediction, electronic health records, explainable AI, SHAP, patient stratification

**Posted Date:** March 17th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9074928/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# DATA-DRIVEN IDENTIFICATION OF HIGH-RISK PATIENT GROUPS FOR FALLS USING SHAP-SPACE PROFILING

Matthias Schulte-Althoff[1,2] (ORCID: 0000-0002-6096-0906); Peter Krappen[2] (ORCID: 0009-0006-7758-1095); Felix Biessmann[3,4] (ORCID: 0000-0002-3422-1026); Sebastian Jäger[3] (ORCID: 0000-0001-9420-8571); Rahel Gubser[1,2] (ORCID: 0000-0002-1231-1987); Armin Haußֿ[5] (ORCID: 0000-0003-3666-4127); Felix Balzer[1] (ORCID: 0000-0003-1575-2056); Tim Kilgus[2] (ORCID: 0009-0002-2908-8272); Daniel Fürstenau[1,2] (ORCID: 0000-0001-8490-7707)

**Affiliations**

1 Institute of Medical Informatics, Charité – Universitätsmedizin Berlin (corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin), Berlin, Germany

2 Department of Information Systems, Freie Universität Berlin, School of Business & Economics, Berlin, Germany

3 Berlin University of Applied Sciences and Technology (BHT), Berlin, Germany

4 Einstein Center Digital Future (ECDF), Berlin, Germany

5 Business Division Nursing, Administrative Department for Academic Health Professions in Practice – Nursing Science and Quality Development, Charité – Universitätsmedizin Berlin, Berlin, Germany

**Corresponding author**

Daniel Fürstenau, Dr. rer. pol.

Department of Information Systems, Freie Universität Berlin, School of Business & Economics

Garystraße 21, 14195 Berlin, Germany

Email: daniel.fuerstenau@fu-berlin.de

**Funding**

This work was part of the KIP-SDM project funded by the Federal Ministry for Research, Technology and Space (BMFTR), Grant No. 16SV8854.

**Abstract.** Hospital falls are common and costly, yet fall risk is heterogeneous, limiting one-size-fits-all prevention. We aimed to derive interpretable, deployable fall-risk profiles from electronic health records (EHRs) to support subgroup-specific interventions. In a retrospective observational study using EHR data from a large German university hospital (2016–2022; N=932,102), we trained an ensemble model to predict inpatient falls and computed patient-level Shapley Additive Explanations (SHAP). We then clustered fall-risk-flagged patients in SHAP space to obtain clinically interpretable risk profiles, benchmarked this approach against alternative clustering strategies, and assessed threshold trade-offs for high-risk classification. The ensemble achieved excellent discrimination (AUROC 0.95) and strong performance for the rare outcome (AUPRC 0.36) on a held-out test set. The model outperforms a guideline-features-only baseline and is competitive with prior falls prediction models in the literature. SHAP-space clustering yielded nine distinct, interpretable risk groups with specific risk factors. This allows actionable mapping to guideline domains and enabling tailored interventions based on established risk profiles. Overall, we provide an approach of how explainable, outcome-guided clustering can translate EHR-based fall prediction into tailored interventions in inpatient care.

**Keywords:** inpatient falls, risk prediction, electronic health records, explainable AI, SHAP, patient stratification

# 1 Introduction

Falls are a major concern: globally, they account for around 684,000 deaths per year, with approximately 37.3 million incidents requiring medical intervention [1]. Between 0.85% and 1.5% of total healthcare costs are attributable to falls [2], with up to 42% of falls deemed preventable through the implementation of multicomponent nonpharmacological interventions [3]. Preventing falls among hospitalized patients is one critical concern for patient safety and resource allocation. Traditional fall risk assessments are administered at admission and throughout hospitalization [4]. However, these approaches lack high predictive validity to differentiate between high and low fall risk [4], as they do not consider a patient's detailed medical history. In practice, they favor sensitivity, i.e., they capture most patients who might fall, at the expense of specificity, resulting in nearly all patients being labeled “high risk” [5]. This leads to interventions that are too broad and could be better tailored to individual patients' needs, which may strain clinical resources.

Recent global guidelines have underscored the need to refocus fall prevention on personalized, multifactorial interventions. The World Falls Guidelines [6] recommend that patients identified as high risk for falling should receive a comprehensive multifactorial falls risk assessment with personalized “tailored” interventions. In other words, after a general risk flag, clinicians must drill down into each individual’s unique combination of risk factors, from mobility and cognition to medications and environment, and tailor interventions accordingly. Given the heterogeneous nature of these risk factors, individualizing care plans is difficult in practice. There is a clear need for data-driven tools to support this process by stratifying high-risk patients into meaningful subgroups and identifying which factors to target for each subgroup.

Prior work has shown that fall-risk prediction is feasible across clinical settings, including inpatient EHR-based prediction and longer-term risk forecasting [7,8,9]. Yet integrating such models into routine clinical workflows remains challenging, because prevention workflows must operationalize guideline recommendations under resource constraints [5,6]. In line with recent research calls [10], our work incorporates large-scale predictive analytics into healthcare, using patient records to predict fall risks and generate actionable insights. Notably, the growing digitization of health records has made such analyses possible on an unprecedented scale [11]. However, building effective solutions requires careful balancing between model generalization and patient-specific specifics [12]. What’s more, integrating such models into routine clinical workflows remains challenging. Recent work on deployable EHR-based machine learning systems shows the importance of closing the loop from prediction to actionable clinical outputs, such as risk

monitoring dashboards [10]. However, building effective solutions requires careful balancing between model generalization and patient-specific nuance [12].

Recent developments in Explainable Artificial Intelligence (XAI) highlight promising strategies to unpack how diverse variables combine to influence a single outcome, such as hospital-based falls [13,14]. However, existing approaches either (i) stratify patients using coarse checklist-based tools that prioritize sensitivity and therefore over-alert while providing limited specificity [4,15–18], (ii) cluster patients in outcome-agnostic feature space, which can reveal latent structure but is not necessarily tied to falls outcomes or intervention relevance [19,20], or (iii) provide local explanations without aggregating them into reusable subgroup summaries that are actionable [13,20–22]. At the same time, clinician-facing explanation formats can materially affect trust and decision behavior [14]. An approach is needed that combines risk prediction with explanation-driven subgrouping to derive clinically interpretable risk profiles that can be mapped to tailored interventions [6,23].

This paper presents a methodological framework for *identifying fall risk factors at the patient group level while retaining interpretability for clinical use.* A patient group is defined as a subset of the patient population who share specific characteristics identified by the method. We train machine learning models on an EHR dataset covering 2016–2022 to predict inpatient falls. Using SHapley Additive exPlanations (SHAP) for model explainability, we then perform outcome-guided clustering of patients to reveal distinct risk profiles. In contrast to clustering patient records independent of the risk assessment, the clustering of local explanations allows to identify risk profiles associated with a particular target variable, such as fall risk. And in contrast to local explanations alone, the grouping via clustering methods allows to identify commonalities between risk profiles of individual patients.

Our main contributions are as follows:

- Identify salient fall risk factors using machine learning models on large-scale EHR data that exceed or are competitive with prior falls prediction models in the literature.
- Apply SHAP values to provide local transparency into each prediction.
- Generate patient subgroups via SHAP-based clustering, unveiling different combinations of risk factors that lead to falls.
- Evaluate whether the risk profiles stratify observed fall incidence in a held-out evaluation cohort compared with feature-space clustering baselines.

- Integrate the model into clinical decision-making by mapping new patients to these risk profiles in real-time and suggesting cluster-specific preventive measures based on literature.

# 2 Research Background

## 2.1 Fall Risk Assessment

Fall risk research has ventured in three main areas: (i) identifying fall indicators, (ii) improving fall risk assessment tools, and (iii) developing interventions to prevent falls in hospitals. Comprehensive guidelines, such as the World Fall Guidelines [6] or national guidelines, including the Expert Standard on Falls in Germany [24], consolidate this knowledge. They serve as foundational resources for tailored, institution-specific patient assessment strategies. The guidelines stress multifactorial approaches that examine person-specific risk factors (age, history of falls, fear of falling, cognitive impairment, depression, etc.), medication-related risk factors (polypharmacy, especially psychotropics), and environmental risk factors (e.g., poor lighting, obstacles, inadequate footwear). Extensive research has identified specific person-related conditions that contribute to increased fall risk, including diabetes mellitus [25], pointing in the direction of more nuanced explanations for fall risk. Recent studies have highlighted the role of frailty [26,27], pain, and malnutrition [28].

The implementation of fall risk assessment tools in clinical practice are manual and checklist-based. Common instruments like the Morse Fall Scale [29], STRATIFY [30], and Hendrich II [31] cover a range of person- and medication-specific risk factors: they include nursing observations (e.g., mobility or the Barthel Index for activities of daily living), fall history, fear-of-falling scales, and basic balance or gait tests, such as Timed Up and Go [32] and Tinetti [33]. These practices have been shown to have limited predictive accuracy and broad applicability [15, 16]. In a systematic review, Park [4] emphasized the use of two assessment tools rather than one, as the included assessment tools did not demonstrate sufficiently high predictive validity to differentiate between high and low fall risk. In practice, as noted, most of these tools err on the side of caution (high sensitivity), flagging a large proportion of patients as at risk while lacking specificity. For instance, many hospitals find that well over half of inpatients trigger a fall risk alert under current protocols, which can lead to “alert fatigue” and less targeted prevention measures [18].

## 2.2 Machine Learning and Heterogeneity in Fall Risk

Recent machine-learning studies have used diverse data sources and settings for fall-risk modeling, including acute inpatient EHRs [7,9,34,35], long-term care data [36], disease-specific cohorts [37], sensor-based trajectories [38], and

explainable geriatric assessments [13]. Ossai et al. [34] proposed a real-time risk profiling approach in acute care that (a) used multivariable logistic regression to identify key risk factors of falls, pressure injuries, and combined vulnerability, and (b) evaluated multiple classifiers to predict discrete risk categories at admission, with an additional illustrative cost-savings calculation based on length-of-stay assumptions. Chu et al. [35] and Thapa et al. [36] similarly illustrate that combining multi-domain clinical information, such as comorbidities, can yield richer risk characterizations than single-score tools, while Huang et al. [37] exemplify risk modeling within a specific neurological subgroup (Parkinson's disease). Complementary lines of work leverage sensor signals to derive sequential risk patterns [38] or explainability on assessment-based features, such as SHAP on geriatric measures [13]. Due to methodological constraints and data limitations, the above studies tend to be based on small cohorts and include mostly known risk factors.

Crucially, research has increasingly recognized that "one size fits all" interventions may not work equally for all patients. In a meta-analysis, Morris et al. [5] discuss numerous intervention methods to reduce fall rates in a hospital setting. The authors cite patient and staff education as an important lever to reduce hospital falls, ultimately emphasizing the importance of multifactorial interventions. The authors noted that interventions need to be evidence-based and targeted, given the variability in what causes falls for different patients. Morris et al. [5] also point out the need for evidence-based treatment and intervention plans, given the conflicting results of various studies. In particular, the fall risk assessment instrument tool must balance a trade-off between its sensitivity, i.e., capturing all potential fall-risk patients, and either precision or specificity, i.e., implementing targeted interventions for those confirmed to be at greatest risk to ensure efficient resource allocation [17].

This is consistent with the idea of heterogeneous treatment effects in healthcare, whereby patients or subgroups may respond very differently to the same intervention. Traditionally, clinical research and guidelines examine average effects [39]. Crucially, both fall-risk drivers and responses to prevention measures can vary substantially across patients, consistent with heterogeneous treatment effects in clinical care [40-41]. Yet estimating highly granular causal effects for individuals from observational EHR data is difficult in practice, especially under high dimensionality and noise [42,43]. We therefore pursue a pragmatic middle ground: we cluster patient-level explanation patterns among alerted patients to derive risk profiles with distinct combinations of risk drivers that can be mapped to subgroup-tailored prevention strategies.

This advance, which can be described as the combination of unsupervised and supervised machine learning models to improve transparency and guide outcomes, have gained some promising traction: Miotto et al. [19] demonstrated that unsupervised deep representations of EHR data ("*Deep Patient*" model) could identify patient profiles, though not explicitly tied to outcomes. A similar clustering approach by Lee et al. [44] also lacks interpretability. More recently, Li et al. [22] and Carr et al. [23] have shown the value of combining clustering with explainability or outcome guidance. Li et al. [22] introduced an interpretable clustering framework for clinical time-series, where local feature importance was used to form clusters that are both data-driven and clinically meaningful. Carr et al. [23] similarly argued that incorporating outcome information (like whether a fall occurred) during clustering can produce patient groups that better capture disease-relevant features than purely unsupervised methods. Loftus et al. [20] provide a narrative review on clinical subgroup clustering in healthcare and stress the importance of interpretability for clinical adoption. Clinicians need to understand why patients are grouped together and what risk factors define each subgroup. Grouping local explanations such as SHAP values or decision rules to describe each cluster, can bridge that gap. Here we leverage well established clustering methods to group risk factors of individual patients' predictions. This approach provides both a global perspective (a cluster-level synopsis of risk factors) and a local perspective (patient-specific justifications) to enhance confidence and ease of use in regulated clinical contexts.

Building on this background, our study uses an outcome-guided clustering framework that integrates EHR data, including nursing documentation, to capture important fall risk factors. By identifying risk profiles and linking them to targeted intervention strategies, we aim to operationalize the personalized approach advocated by fall prevention guidelines. The modeling idea is sketched in Fig. 1.

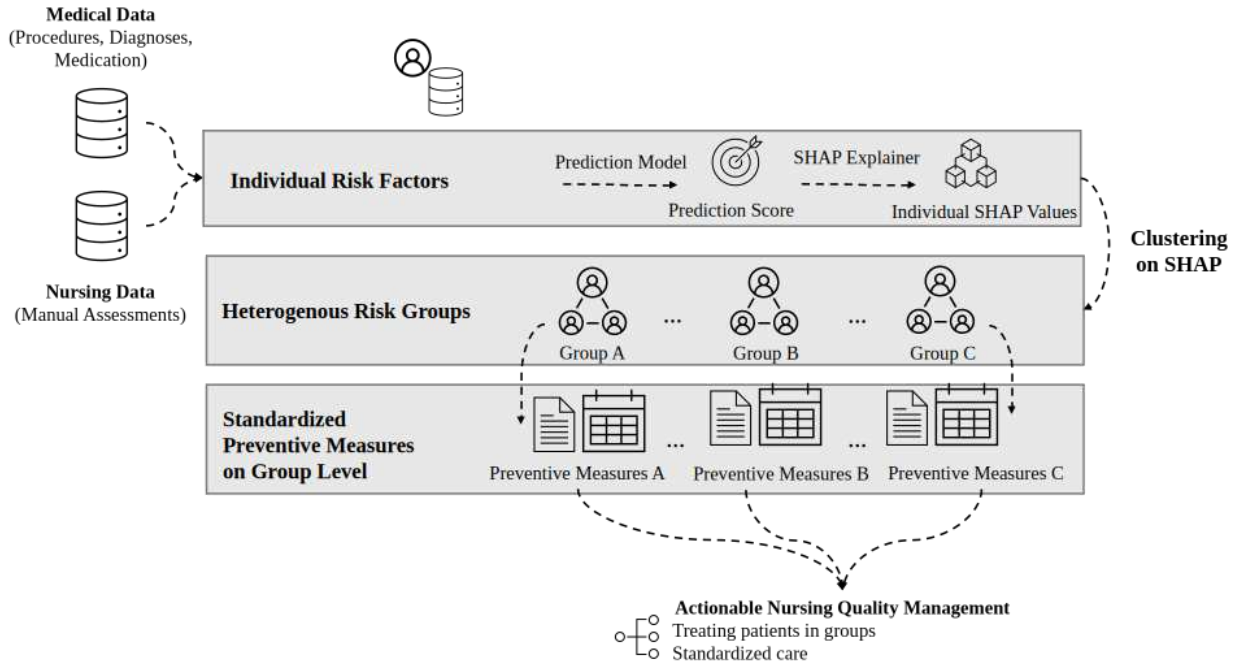



**Fig. 1** Identification of patient risk profiles using machine learning, SHAP values, and clustering.

# 3 Methods and Data

## 3.1 Data Sources and Coding

We conducted a retrospective observational study using pseudonymized EHR data from a large German university hospital. Inpatient admissions from 2016 to 2022 were considered. To enrich the data for fall events, we included all patients who experienced one or more falls during the period (as documented in incident reports), as well as a sample of patients who did not fall; inclusion among all patients was limited to adults (age ≥18) and to ensure a broad representation across hospital services, non-fallers were drawn from the 20 departments with the highest fall counts (e.g., oncology, geriatrics, neurology, neurosurgery, internal medicine, psychiatry; excluding intensive care units)[1]. The final dataset comprised 932,102 unique patients, of whom 10,443 (1.12%) had at least one recorded fall during their hospital stay. This resulted in a highly imbalanced dataset (Table 1). We address class imbalance through appropriate choice of evaluation metrics, such as AUPRC and F1. For each patient, we extracted demographics (age, sex), diagnoses, procedures, nursing care data[2], which encompasses, among others, mobility, gait, medication, and fall

[1] Our evaluation of the university hospital's fall risk model demonstrated varying levels of accuracy, precision, and recall, as detailed in Appendix A. Based on logistic regression, patients labeled "+" (increased fall risk) are over four times (OR 4.17) more likely to fall, while those with "++" (strongly increased fall risk) have over six times (OR 6.31) the risk compared to those marked "0".

[2] In some cases, diagnoses may not have been coded until the end of the hospital stay. In these cases, the diagnosis date is too late and is sometimes disregarded for machine learning. However, procedures necessary for billing purposes were subject to more precise dating.

risk assessment scores documented by nursing staff at admission and every 7 days thereafter (the hospital uses a three-level scale: "0" = no risk, "+" = some risk, "++" = high risk).

Diagnosis data included ~12.18 million entries coded in ICD-10-GM (the German modification of the International Classification of Diseases, 10th Revision). Procedure data included ~5.72 million entries coded in OPS (Operationen- und Prozedurenschlüssel, the German procedure coding system). Both coding systems are hierarchical, allowing aggregation into broader categories when needed. Based on clinical consultation, we applied a data cleaning step: diagnoses marked with low certainty (about 4.1% of diagnosis entries) were removed to avoid spurious information. Missing age data for a small fraction of non-fallers (<0.05%) were imputed with the median age (57 years), with no age data missing for fallers. Nursing care data was absent in 76% of all patients, primarily because routine nursing documentation was not initiated for patients who were screened at admission as sufficiently independent/fit and therefore did not require a formal nursing care assessment within the nursing care module. We also derived from hospital records whether each patient had any documented falls prior to the first inpatient fall (for fallers) or during their hospital stays (for non-fallers), and we recorded the initial nursing risk assessment score for falls at admission. These serve as baseline comparisons for our model's predictive performance. To prevent data leakage, only data available before a fall incident or before discharge for non-fallers was used via the clinic's timestamps.

**Table 1** Patient attributes and faller vs. non-faller comparison from 2016 to 2022. Values are presented as mean ± standard deviation or count (percentage).

| **Characteristic** | **All Patients** (n = 932,102, 100%) | **Non-Faller** (n = 921,658, 98.8%) | **Faller** (n = 10,443, 1.12%) |
|---|---|---|---|
| **Male (m)** | 435,324 (46.71%) | 429,578 (46.62%) | 5,746 (55.02%) |
| **Female (w)** | 496,588 (53.29%) | 491,891 (53.38%) | 4,697 (44.98%) |
| **Age (years)** | 57.44 ± 19.76 | 57.23 ± 19.70 | 76.03 ± 15.72 |
| **Diagnoses / patient** | 4.82 ± 6.73 | 4.67 ± 6.48 | 17.96 ± 12.46 |
| **Procedures / patient** | 1.96 ± 3.46 | 1.88 ± 3.34 | 8.90 ± 6.12 |

Fallers were significantly older (Cohen's d = 0.956) and had more complex medical histories. In our data, the five most common diagnosis categories among fallers were (1) family history of health risk factors, (2) ischemic heart disease, (3) metabolic disorders (e.g., diabetes), (4) hypertension, and (5) musculoskeletal disorders (see Appendix B). Non-fallers shared many of the top diagnoses (reflecting overall hospital case mix), but fallers had higher rates of cardiovascular and metabolic issues, likely due to their higher age. On the procedure side, fallers most frequently underwent imaging, circulatory system procedures, e.g., catheterizations, and oncologic treatments, such as radiation

or chemotherapy. Two procedure categories notably more prevalent in fallers were *computed tomography* (72% vs. 16%) and *patient monitoring* (56% vs. 12%).

## 3.2 Predictive Modeling

We formulated fall prediction as a binary classification problem: fall vs. no-fall during the hospital stay. Features were constructed from the patient's data available up to the time of fall for fallers, or up to the end of the hospital stay for non-fallers (see Fig. 2). In practice, for patients who fell, we included all diagnoses and procedures recorded before their first fall event to avoid labeling consequences of the fall as causes. Guideline-specified fall risk factors included: fall during current stay; walking impairment; balance/gait impairment; use of a walking aid; gait assessment; fall within the last 12 months; number of falls in the last 12 months; transfer impairment; transfer assistance; cognitive impairment; disorientation (time/place/own person); agitation; confusion; bed-mobility impairment; bed-mobility assessment; skin condition relevant to mobility; presence of medical items (e.g., infusion, catheter); impairment of excretions; urinary/faecal incontinence; nocturia; use of psychotropic or sedative medications; decubitus; and fall during transfer. For patients who did not fall, we included data from their entire hospitalization. Each patient thus has a feature vector summarizing their demographics, diagnoses, procedures, and nursing care data prior to fall or prior to discharge for non-fallers. Categorical features (ICD-10-GM and OPS) were one-hot encoded using codes on the group level. A higher number of digits indicates a more specific diagnosis, e.g., "J45" indicates asthma, while "J45.0" indicates predominantly allergic asthma. Like the ICD, OPS codes have a hierarchical structure, e.g., "8-98" indicates various complex treatments, while "8-981" indicates a neurological complex treatment. The first two characters of the codes were used for grouping.

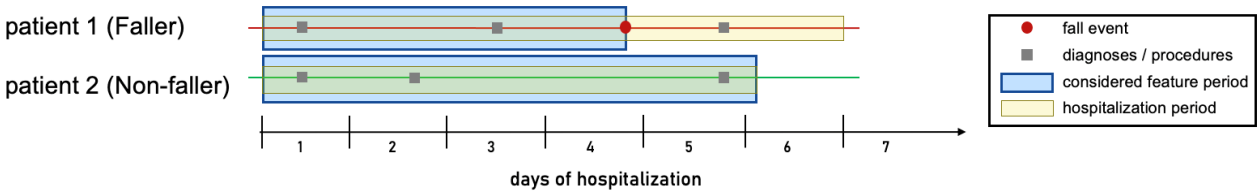



**Fig. 2** Modeling of medical events during the patient's hospital stay.

We leveraged an AutoML library (AutoGluon) to explore a suite of models—logistic regression, multilayer perceptron (neural network), random forest, gradient-boosted trees (XGBoost and LightGBM), and stacked ensembles. We performed a patient-wise random split into a training set used for cross-validation and model selection (80%;

n=745,682) and a final held-out test set for all reported performance and threshold analyses (20%; n=186,420). Model selection and hyperparameter tuning were performed using 5-fold cross-validation on the training set, with the area under the precision–recall curve (AUPRC) as the primary metric for optimization. AUPRC is well-suited for imbalanced data because it focuses on performance in the minority class and is less dominated by the abundance of negative cases. It also evaluates model performance across all possible prediction thresholds, aligning with our interest in how different threshold choices affect precision and recall. For comparability, we additionally report macro-F1 and operating-point metrics (sensitivity/recall, specificity, precision/PPV) at selected probability thresholds.

To address severe class imbalance, we optimized model selection using AUPRC and evaluated all final results on the held-out test set with the natural outcome distribution. We avoided using oversampling techniques, such as SMOTE, during training to keep the training data as representative as possible. To improve the understanding of the model predictions and to investigate the underlying fall risk factors in detail, we used SHAP, a game-theoretic technique that can interpret model outcomes [21]. The resulting values, known as SHAP values, indicate the mean deviation from the mean prediction when a conditioning on a single feature, thereby revealing the impact of individual features on model outcomes. By relating the magnitude of features to their relevance in the model predictions, SHAP values can provide insight into different risk factors and latent interactions at the instance level [45]. For the implementation, we used FastSHAP [46] which has become a popular, more efficient variant of the original implementation.

## 3.3 Threshold Analysis for Model Deployment

The envisioned use of our model is as a clinical decision support tool that runs in parallel with standard care processes. During patient admission and evaluation, the model would generate a fall risk prediction (with a probability score). If the predicted risk is above a chosen threshold (see Threshold Analysis below), the patient is flagged as high-risk for falls. At that point, the system provides two levels of explanation: (1) an individual SHAP explanation highlighting the patient's top risk-increasing and risk-decreasing factors, and (2) a cluster assignment suggestion indicating if the patient matches one of the pre-identified risk clusters. Fig. 1 illustrates this integration conceptually. In practice, a clinician reviewing the alert would see something like: "High fall risk predicted. Key factors: age 85 (+), recent surgery (+), use of sedative medications (+), independent mobility (-). Patient fits Cluster 2— 'Frail patients with rehabilitation needs'." The clinician can then confirm the relevance of this profile and initiate a tailored care plan accordingly. Each cluster is linked to a standardized set of recommended interventions.

A practical aspect is choosing the probability threshold at which to classify a patient as "high risk." In training, we optimized for overall AUPRC, considering all possible thresholds simultaneously. Depending on hospital priorities, one might prefer a sensitive or a more specific model. For deployment, the model outputs a fall-risk probability score per admission that can be converted into an alert via a probability threshold. We therefore conducted a post-hoc threshold analysis on the held-out test set using candidate thresholds and reporting sensitivity, specificity, precision (PPV), and the number needed to evaluate (NNE = 1/PPV). The PPV can characterize the workload–benefit trade-off for clinic administration and nursing care personal.

### 3.4 SHAP-Based Patient Clustering

Our next step was to identify clusters of high-risk patients with similar risk-factor patterns in their local explanations, rather than clustering raw features. In this representation, each patient is described by a vector of feature-attribution values, where larger positive values indicate stronger contributions toward the model predicting a fall. Clustering in SHAP space groups patients by the model-attributed drivers of risk (e.g., mobility, cognition, medications), rather than by feature similarity alone, yields subgroup-level "why" profiles that are globally interpretable, align directly with guideline-based multifactorial assessment, and can be translated into tailored intervention bundles for each cluster. To ensure an unbiased summary, we clustered only the held-out test-set predictions and their SHAP vectors, thus preventing information leakage from training/validation into subgroup discovery. We constructed a vector of SHAP values for each patient using the final model: $Sp = (s_1, s_2, \ldots, s_m)$, where $m$ is the number of features. Here $s_i$ is the SHAP value for feature $i$, representing how much that feature pushed the model's prediction for patient $p$ above or below the baseline fall probability.

Patients who have similar SHAP value profiles can be considered to have similar risk factor contributions, even if their raw feature values differ. For example, two patients may differ in absolute age or diagnosis codes, yet both may be 'high-risk' primarily due to prior falls and sedating medication exposure, leading to similar SHAP patterns. We applied K-means clustering to the matrix of SHAP value vectors for patients who were predicted by the model to be at risk of falling. To determine an appropriate number of clusters $k$, we used the elbow method on the clustering distortion—the within-cluster sum of squared distances. As shown in Fig. 3, the elbow plot suggested an optimal clustering around $k = 9$ clusters for our data. Increasing beyond 9 yielded diminishing returns (clusters became too small or too similar), while fewer than 9 clusters would merge distinct risk profiles and reduce granularity. We prioritized a solution that maximized within-cluster similarity of SHAP profiles, because our goal is to define groups for which very specific,

tailored interventions can be designed. After clustering, we characterized each cluster both externally by patient attributes and in terms of their respective risk factors.

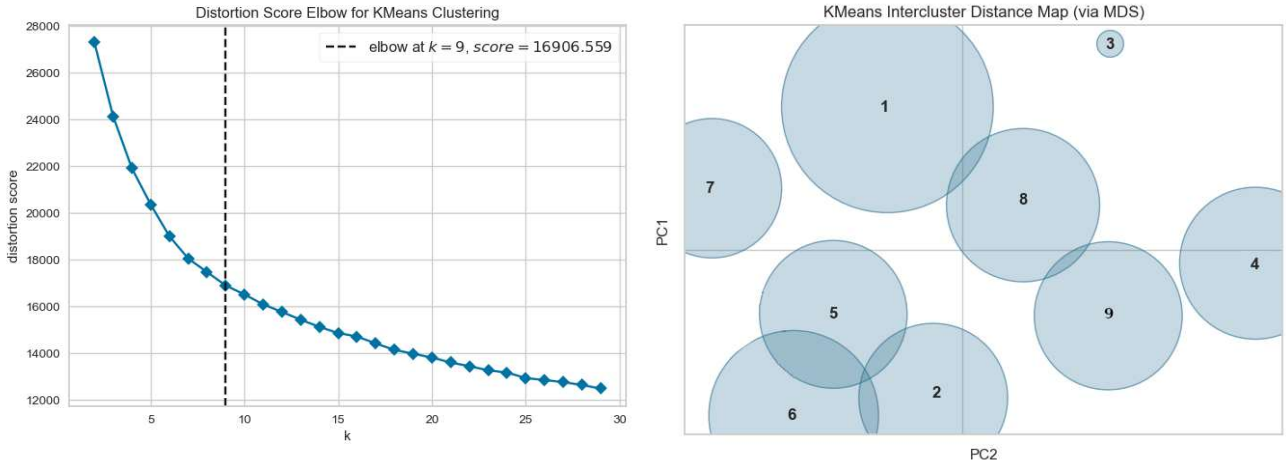



**Fig. 3 (1)** K-elbow plot to determine the number of clusters with an optimum of k=9 and **(2)** intercluster distance map with all features reduced to a two-dimensional vector space using multidimensional scaling (MDS)—a distance-preserving technique that places high-dimensional points in 2D while approximately maintaining their pairwise dissimilarities [47].

We evaluated whether explanation-space clustering yields risk profiles that are (i) coherent in terms of risk drivers and (ii) clinically meaningful in terms of fall incidence separation among patients who would trigger an alert. We benchmarked SHAP-space K-means against representative baselines operating on (a) raw feature space (K-means, Gaussian mixture model, Ward agglomerative, spectral clustering), (b) reduced feature representations (PCA(95%) + K-means), and (c) score-only stratification (predicted-risk quantile bins).

To quantify clinical stratification among alerted patients, we defined a held-out high-risk evaluation cohort as test-set admissions that would trigger an operational alert at a prespecified probability threshold ($p \geq 0.08$). For each method, we computed cluster-level observed fall incidence with 95% Wilson confidence intervals and summarized outcome separation by (i) incidence range (max–min), (ii) incidence ratio (max/min), and (iii) IQR of incidence across clusters. We refer to a method as having stronger outcome separation when it yields larger values on these stratification summaries, reflecting greater dispersion of observed incidence across clusters in the alerted cohort. We additionally assessed risk–outcome alignment using Spearman correlation between cluster mean predicted risk and observed incidence, and cluster-level calibration error (MAE). For method-level summaries, clusters contributing <1% of the evaluation cohort were collapsed.

In a prospective deployment, new inpatients (not seen during training) predicted as high risk must be assignable to one of the discovered risk clusters to surface the most relevant, subgroup-specific risk drivers. This enables tailored

interventions aligned with guideline-recommended multifactorial care. We propose a simple distance-based method to assign new patients to clusters in the SHAP space.

To support prospective use, we map newly alerted patients to the discovered SHAP-based risk profiles by nearest-centroid assignment in SHAP space via k-means prediction. To avoid unstable labels, we apply two confidence checks: (i) the patient must lie within a distance threshold defined by the empirical distribution of within-cluster distances (e.g., 90th percentile), and (ii) the closest cluster must be sufficiently closer than the second-closest. Patients failing either check are marked unassigned. Distance mapping are reported in full Appendix C and the overall modeling pipeline of section 3 is sketched in Fig. 4.

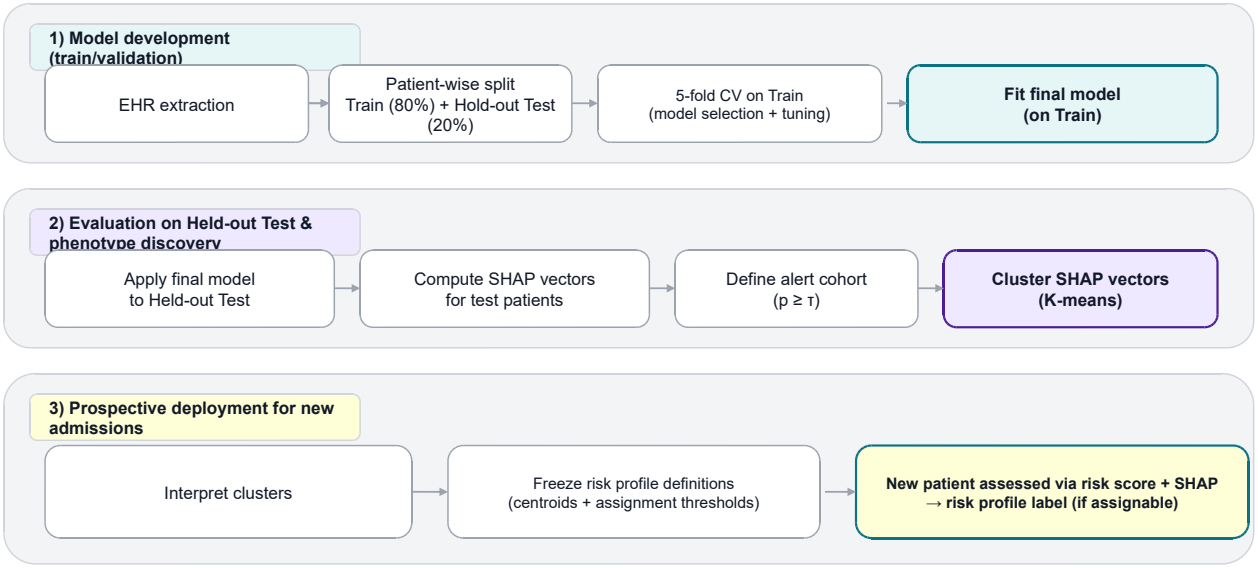



**Fig. 4** Workflow for fall-risk profiling and placement of SHAP-space clustering. Patients with predicted fall-risk probability p above the alert threshold τ form the alert cohort used for risk-profile discovery. Risk profiles are defined by cluster centroids and then used prospectively to assign newly alerted patients by nearest-centroid mapping. Note: $p$ denotes the model-predicted fall-risk probability; in the main analyses $\tau = 0.08$.

# 4 Results

## 4.1 Predictive Modeling

Table 2 presents selected evaluation metrics for representative models. A weighted ensemble model L3 achieved the highest AUPRC (0.357) on the test set, an AUROC of ~0.946, and among the highest recall, while maintaining good precision (Table 2). It significantly outperformed the hospital's existing fall risk score baseline odds ratios (Appendix A). The precision-recall performance, while modest in absolute terms, was far above the no-skill baseline (prevalence ~1.2% so baseline AUPRC ~0.012). By contrast, the simplified model that only used guideline-specified risk factors (26 features corresponding to factors like history of fractures, mobility impairment, cognitive impairment, depression,

etc.) performed much worse (AUPRC ~0.05, AUROC ~0.82). This highlights that additional EHR variables carry important predictive signal that the traditional risk factors alone do not capture. For example, our full model found that certain procedure categories (like hemodialysis, or physical therapy orders) were among the top predictors of falls, whereas a guideline-based model focusing on diagnoses might disregard those entirely.

**Table 2** Model performance metrics on the test set. AUPRC, AUROC and F1 (macro) metrics for various prediction models to predict falls based on patient history. The XGBoost: Fall Guidelines model has been trained only on features that are considered in the World Falls Guidelines and a common German fall standard [6, 24].

| ML Metric | Logistic Regression | Neural Net for Classification | RandomForest Gini BAG L2 | Weighted Ensemble_L3 | LightGBMXT BAG L2 | XGBoost: Fall Guidelines |
|---|---|---|---|---|---|---|
| **F1 (macro)** | 0.5053 | 0.6084 | **0.6345** | 0.6313 | 0.6318 | 0.5057 |
| **AUPRC** | 0.1079 | 0.1639 | 0.3343 | **0.3568** | 0.3449 | 0.0486 |
| **AUROC** | 0.9044 | 0.8802 | 0.9373 | 0.9463 | **0.9472** | 0.8225 |

Using the threshold analysis (Fig. 5), we could consider a threshold around 0.08, which roughly corresponds to a sensitivity of ~0.55 as a reasonable setting, close to the hospitals' baseline sensitivity of the current fall risk assessment. In practice, that means for every ~5 patients flagged, ~1 would be a true faller. Hospital units can likely accommodate this ratio, deploying preventive rounds or physical assistance to five patients to potentially prevent one actual fall. From an implementation standpoint, the final model's Negative Predictive Value was extremely high (>99%) due to the rarity of falls. On the other hand, the Positive Predictive Value in the chosen operating range is more modest (on the order of 20%), but importantly, substantially higher than the ~1% pre-test probability of falls and higher than the precision of simpler screening tools.

## 4.2 Threshold analysis for operational deployment

We performed a post-hoc threshold analysis to quantify this trade-off evaluating model performances at various probability cut-offs to observe how sensitivity and precision change, as shown in Fig. 5. At the default threshold (0.50), weighted ensemble model achieves very high specificity (~99.96%) and precision ~81%, but sensitivity is only ~16%: Only a small fraction of fallers is being identified under this strict cut-off, though those flagged are highly likely to fall. This threshold might be too conservative for clinical use, as it would miss roughly 84% of true fall incidents. If we lower the threshold to capture ~55–60% of fallers, sensitivity improves to ~57–67%, with precision dropping to ~16–21%.

The appropriate threshold thus depends on the costs of false negatives vs. false positives. In inpatient falls, a false negative—failing to identify a patient who then falls—can lead to injury, extended hospitalization, and associated

costs. False positives—flagging someone who wouldn't have fallen—may lead to precautionary interventions like closer monitoring, bed alarms, assistance with toileting, etc. These interventions are generally low-risk and low-cost relative to the consequences of a fall. Therefore, many hospitals choose a threshold that errs on the side of sensitivity, accepting more false-positive alerts to prevent more falls. In our context, moving from 16% sensitivity to ~67% sensitivity would require about a six-fold increase in interventions per actual fall prevented, with precision dropping from 81% to 16%. Whether this is acceptable depends on resource availability. In the remainder of this paper, threshold adjustments used in our capacity analyses were not applied, e.g. when computing SHAP explanations.

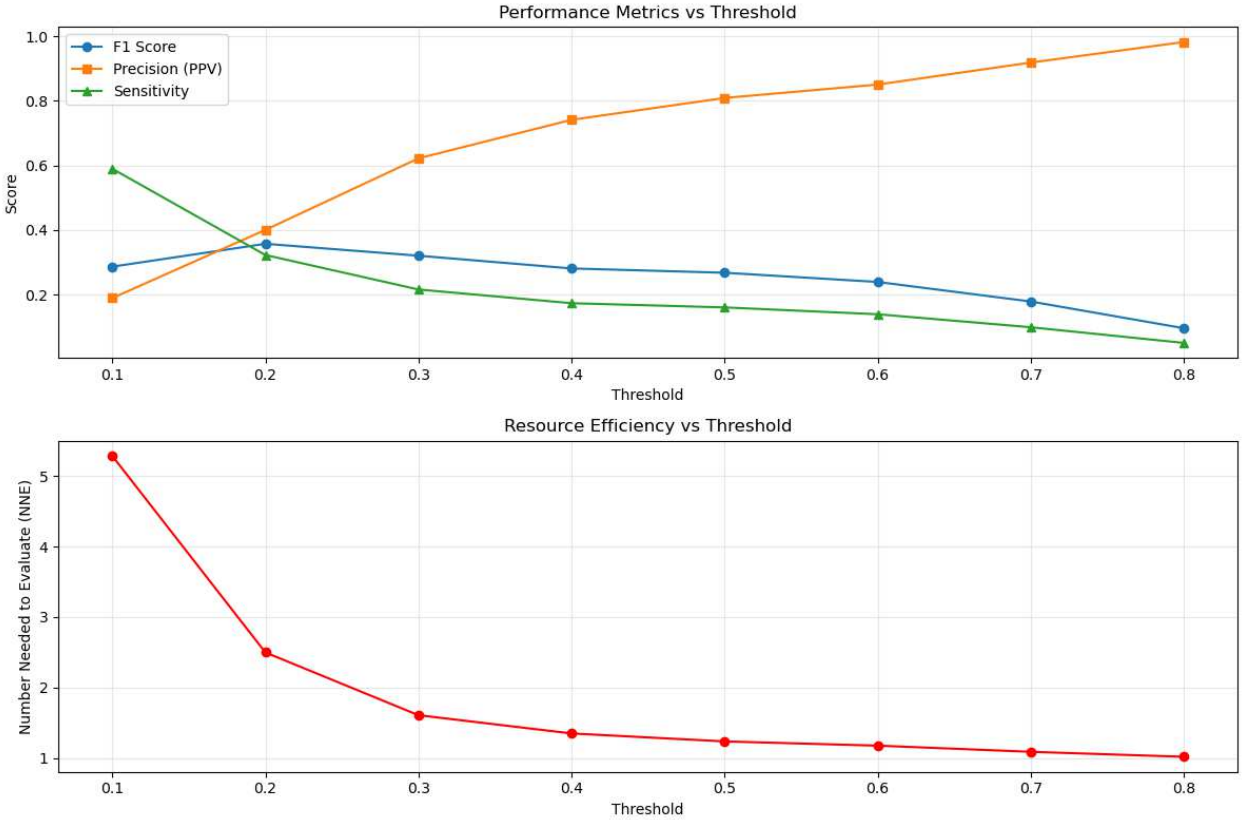



**Fig. 5** The Fig. at the top shows performance metrics at different prediction thresholds. Higher thresholds (right side) yield fewer alerts with high precision but miss more falls (low sensitivity). Lowering the threshold (left side) increases sensitivity (catching more true fall cases) at the expense of more false positives (lower precision). The chosen operating point should balance the cost of missed falls vs. the effort of unnecessary interventions. The Fig. at the bottom shows the *Number Needed to Evaluate NNE* = 1/PPV—that is the number of patients who need to be evaluated to find one true positive.

## 4.3 Risk Profiles by Patient Cluster

For the L3 model chosen, we examined the aggregated local feature importances using SHAP (Appendix D). Based the calculated values we did a SHAP-based K-means clustering as described in the Methods section. We identified nine distinct risk clusters among the patients the model flagged as high risk. The right panel of Fig. 3 shows

an inter-cluster distance map: cluster centroids were computed in SHAP space; for visualization only, centroids were embedded into a 2D space via MDS on the SHAP matrix. All inter-centroid distances reported refer to SHAP L2-space. Clusters 2, 5, and 6 lie adjacent in SHAP-value space, indicating partially overlapping risk profiles driven by complex treatment courses; by contrast, Cluster 3 is well separated. Cluster sizes ranged from ~177 patients (Cluster 3, ~6% of true positives) to ~685 patients (Cluster 1, ~23% of true positives). The clustering approach reveals various risk pathways: complex care needs (Clusters 5, 6), rehabilitation-focused care (Clusters 4, 9), and high-monitoring medical interventions (Clusters 7, 8), with Clusters 1 and 2 representing general care-dependent populations and Cluster 3 standing alone in its mental health focus. Table 3 summarizes the main characteristics of the clusters. For each cluster, interventions suggested by the World Fall Guidelines were added where appropriate. As Cluster 1 and 2 represent general care-dependent populations, general risk-preventing measures were listed. For the rest, tailored measurements were added. Appendix E provides a more detailed cluster description.

**Table 3** Cluster descriptions, including mean principal diagnoses and procedures based on SHAP values. Relative cluster size is calculated as the number of patients assigned to the cluster divided by the number of all patients correctly classified as fallers (true positives). The final column suggests interventions targeted to the cluster's risk profile based upon the World Falls Guidelines by Montero-Odasso et al. [6].

| ***Cluster*** | ***Size*** | ***Age*** | ***Sex (m)*** | ***Salient features based on SHAP values*** | ***Tailored Interventions based on World Fall Guidelines [6]*** |
|---|---|---|---|---|---|
| **C1:** ***General Care*** | 685 (.23) | 79.0 | 0.52 | 9-98 to 9-99: Supplementary measures and information (2.63)<br>Falls in the last 12 months (1.61)<br>I10 to I15: Hypertension (0.50)<br>Z80 to Z99: Health risks based on family or self-history (0.36) | Multidomain intervention based on general *patient education and awareness* (e.g., fear or concerns of falling, fall history etc. yoga, Pilates—p.23 / p.24 ) Consider individual *patient preferences and values. Measures of physicality;* Considerations on activities in daily life (ADL); *consideration on environmental context* (e.g., getting familiar with surroundings; adequate lightning; low bed); *general health state* of patient (hearing aid; walking impairment, Vitamin D supplements—p. 23 etc.)*;Medication review* (especially psychoactives, antihypertensives); *Continence care* (to avoid urgent unsafe toileting); *Cognitive support* and general mental state (e.g., depression, morbidity). |
| **C2:** ***General Complex Care*** | 297 (.10) | 61.2 | 0.54 | Falls in the last 12 months (2.43)<br>8-97 to 8-98: Complex care (2.15)<br>8-80 to 8-85: Circulatory measures (0.59)<br>3-20 to 3-26: Computed tomography CT (0.48) | |
| **C3:** ***Behavioral abnormality*** | 177 (.06) | 64.2 | 0.31 | 9-60 to 9-64: Mental, psychosomatic and behavioral disorders (2.99)<br>9-98 to 9-99: Other complementary measures (1.06)<br>Z00 to Z13: Seeking health care for examination and clarification (0.55)<br>I10 to I15: Hypertension (0.32) | *Nutritional assessment* (adequate vitamin D intake; serum 25 OH vitamin D levels), *assessment of substance abuse and excessive alcohol intake*—p. 18; *Physical exercise* (in community settings; Pilates, Yoga) p.23.; *Behavioural therapy* (for mild cognitive impairment/ dementia: group exercise. cognitive behavioural therapy and Occupational therapy)—p.12 |
| **C4:** ***Frail & Rehab-Focused*** | 307 (.10) | 87.3 | 0.46 | Falls in the last 12 months (2.18)<br>8-55 to 8-60: Early rehabilitation and physical therapy (1.93)<br>Walking and transfer impairments (1.84)<br>I10 to I15: Hypertension (0.50) | *Balance and strength training* (for stroke: balance, strength, walking) p.12; *Multimodal patient education* (e.g., face-to-face, handout, video) p.24. *Physical training* (for balance to people with complex phase PD) p.13. *Medication optimization* (to maximize motor function and minimize side effects—dyskinesia and hypotension) p. 25. *Nutrition & vitamin D* and serum 25 OH supplementation (address malnutrition, bone health).—p.18; *Backward Chaining:* Re-learning the skill to rise from the floor through practice of each of the specific movements required where the last step in the chain is taught first.—p.19; *Fracture Management* |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | incl. Osteoporosis treatment—p.18 |
| **C5: *Complex Treatments* (ICU/Neuro)** | 293 (.10) | 77.8 | 0.57 | 8-97 to 8-98: Complex treatment (2.21)<br>Falls in the last 12 months (1.50)<br>I10 to I15: Hypertension (0.54)<br>Z80 to Z99: Health risks based on family or self-history (0.39) | *Considerations on activities in daily life (ADL);"*cognitive stimulation, daily orientation, early mobilisation, vision and hearing, fluid management, constipation management, feeding assistance, sleep and family involvement."*—p.16; Physical Measures* (progressive resistance strength training e.g., TaiChi); *Medication optimization* (avoid weak opioids, "as the adverse events risk may outweigh the benefit. For neuropathic pain, first line treatment includes serotonin norepinephrine reuptake inhibitors, gabapentinoids and Transdermallidocaineorcapsaicin"—p. 22); *Modification* of orthostatic blood pressure—p.21; *Nutrition & vitamin D* and serum 25 OH supplementation (address malnutrition, bone health)—p.18; *involvement* and training of informal (unpaid) and/or formal (paid) caregivers.—p.19 |
| **C6: *Complex Treatments + Metabolic*** | 383 (.13) | 81.8 | 0.64 | 8-97 to 8-98: Complex treatment (2.99)<br>Walking and transfer impairments (1.52)<br>I10 to I15: Hypertension (0.60)<br>E70 to E90: Metabolic disorders (0.43) | |
| **C7: *Radiation and Chemo-therapy & Circulatory*** | 268 (.09) | 64.9 | 0.50 | 8-80 to 8-85: Measures for blood circulation (1.40)<br>8-52 to 8-54: Radiation therapy and chemotherapy (0.85)<br>6-00 to 6-00: Application of medications (0.72)<br>Z00 to Z13: Seeking health care for examination and clarification (0.71) | *Medication optimization* (avoid weak opioids, "as the adverse events risk may outweigh the benefit. For neuropathic pain, first line treatment includes serotonin norepinephrine reuptake inhibitors, gabapentinoids and Transdermallidocaineorcapsaicin"—p. 22); *Pain Management* (defining cause, type (nociceptive, neuropathic) and intensity with pain rating scale for elderly patients)—p.17. *Non-pharmacological approaches* (physiotherapy and cognitive behavioural therapy)—p.22 |
| **C8: *Patient Monitoring (Cardiac/ Neuro)*** | 312 (.10) | 84.5 | 0.60 | Falls in the last 12 months (1.93)<br>8-92 to 8-93: Patient monitoring (0.70)<br>I10 to I15: Hypertension (0.43)<br>Z80 to Z99: Health risks based on family or self-history (0.31) | *Balance and strength training* (for stroke: balance, strength, walking) p.12; *Multimodal patient education* (e.g., face-to-face, handout, video) p.24*.; Hypotension* (management of orthostatic hypotension incl. modification of possible culprit medications, rehydration, compression garments and medications (e.g., fludrocortisone and midodrine)—p.12; p.21; *ADL* (general physical activity; promotion healthy lifestyle)—p.19; *Medication Review "*for bradycardic disorders (sinus node disease, atrioventricular conduction disorders, vasovagal syndrome and carotid sinus syndrome) and *tachyarrhythmias* (atrial fibrillation, supraventricular and ventricular tachycardia) include modification of culprit medications, specific anti-arrhythmic medication and, in some cases, implantable devices (such as pacemakers and implantable cardioverter-defibrillators); *Syncope Management* (as per local syncope guidelines)."—p.21 |
| **C9: *Rehab + Speech/ Hearing*** | 294 (.10) | 74.7 | 0.64 | 8-55 to 8-60: Early rehabilitation and physical therapy (1.70)<br>Walking and transfer impairments (1.39)<br>9-31 to 9-32: Phoniatric and pediatric audiology therapy (0.88)<br>8-97 to 8-98: Complex care (0.61) | *Positioning "*Particle repositioning (Epley) manoeuvres are an effective treatment for benign paroxysmal position and vertigo and may reduce falls rates, but evidence is limited. Vestibular rehabilitation therapy improves postural and gait stability in cases of vestibular dysfunction although the optimal approach is still unclear as is the effect on falls rates"—p.22 |

With regard to the cluster assignment of new patient, as described in section 3.5, our empirical analysis suggests setting the Primary Distance Threshold $\tau_1$ to the 90th percentile of the intra-cluster distances from our training data, and the predefined ratio threshold $\tau_2$ to 1.5 (see Appendix C). These parameters were were set based on the empirical distance distribution in the cluster-discovery (held-out test) cohort and achieved successful cluster assignment for 93% of patients in the held-out test set. For unassigned patients that do not meet the assignment criteria, the top three candidate clusters can be displayed by the system and considered based on clinical judgment.

## 4.4 Outcome Separation Benchmarking of Clustering Approaches

Across clustering methods, SHAP-space K-means showed the largest separation in observed fall incidence across clusters, according to our prespecified summary statistics—incidence range (max–min), incidence ratio (max/min), and IQR across clusters. Specifically, incidence varied widely across its risk profiles (range 0.424, ratio 5.12, IQR 0.155), indicating both substantially higher- and lower-risk subgroups within the alerted population. In contrast, feature-space baselines—raw K-means/GMM/PCA-K-means/Ward/Spectral—yielded markedly narrower risk gradients (range 0.089–0.207, IQR 0.016–0.050), indicating more weakly differentiated outcome strata. Moreover, SHAP-space clusters showed the best alignment between clusters' mean predicted risk and observed incidence (Spearman 0.943, cluster-level calibration MAE 0.014), whereas alignment was substantially weaker for most feature-space baselines (Spearman 0.257–0.714, MAE 0.019–0.023). Taken together, these findings demonstrate that clustering in explanation space not only yields interpretable risk profiles but also produces more outcome-relevant stratification among patients who would be classified as high risk.

**Table 4** Outcome separation benchmarking of clustering approaches in a high-risk evaluation cohort ($p \geq 0.08$).

| Method | $k_{effective}$ | Observed incidence range across clusters (*max–min*) | Observed incidence ratio (*max*/*min*) | IQR of observed incidence | Spearman Correlation ($\bar{risk}$_*cluster*, *obs_rate_cluster*) | Cluster calibration MAE |
|---|---|---|---|---|---|---|
| **SHAP-space K-means** | 6 | **0.4239** | **5.1183** | **0.1550** | **0.9429** | **0.0141** |
| Raw K-means | 8 | 0.2050 | 2.5983 | 0.0175 | 0.3810 | 0.0231 |
| Raw GMM | 8 | 0.2069 | 2.6179 | 0.0163 | 0.3333 | 0.0233 |
| PCA(95%) K-means | 6 | 0.2045 | 2.6170 | 0.0263 | 0.2571 | 0.0229 |
| Raw Ward | 6 | 0.0888 | 1.9060 | 0.0496 | 0.7143 | 0.0188 |
| Raw Spectral | 8 | 0.1462 | 2.6187 | 0.0364 | 0.3095 | 0.0230 |
| Risk-score bins | 2* | 0.6705 | 6.2659 | 0.3353 | NA* | 0.0182 |

Note: Metrics summarize the distribution of observed fall incidence across clusters for each method. Clusters contributing <1% of the evaluation cohort were collapsed for reporting; thus $k_{effective}$ may be < 9. Full cluster-level outcomes are provided in Appendix Table F1. Lower is better for calibration MAE; higher is better for Spearman correlation between cluster mean predicted risk and cluster observed fall incidence.
*Collapsed to 2 bins due to tied/low-variance predicted probabilities; to be treated as a stratification comparator rather than a risk-profile discovery method

# 5 Discussion and Further Research

Using a large-scale EHR from 2016–2022 with 932,102 patients, our weighted ensemble model achieved AUROC ≈0.95 and AUPRC ≈0.36 on a held-out test set, substantially outperforming a guideline-features-only model (AUROC ≈0.82; AUPRC ≈0.05). Outcome-guided clustering of SHAP values on the test set yielded nine interpretable risk profiles. Benchmarking against feature-space and score-based comparators showed that SHAP-space clustering yields the largest between-cluster differences in observed fall incidence among alerted patients, while also maintaining strong alignment between cluster-level predicted risk and observed outcomes. A simple SHAP-space distance-mapping assigned 93% of test patients confidently to one of these clusters. Threshold analysis indicated feasible operating points. e.g., threshold ~0.08 and a sensitivity of ~0.55, PPV ~0.20, NPV >99%) that would allow units to tune alerts to resources. Together, the pipeline supports guideline-concordant, multifactorial, and subgroup-tailored prevention in routine care.

Our stacked ensemble model achieved is competitive with or exceeds prior falls prediction models in the literature. For instance, one study by Hsu et al. used an XGBoost classifier on electronic health record (EHR) data to predict falls during hospital admission and found an Area Under the Curve (AUC) of 0.700 (95% CI: 0.614, 0.777) in a cohort of 639 adult patients (297 fall patients and 342 controls) [7]. Mishra et al. developed a ML model using gait and geriatric assessment data from 92 geriatric patients to predict 6-month fall risk and achieved an AUC of 0.80 (95% CI: 0.76, 0.85) [13]. Ye et al. used statewide health records from Maine (n = 265,225) to develop an XGBoost model predicting one-year fall risk, achieving a C-statistic (equivalent to the ROC-AUC) of 0.851 in the training set and 0.803 in the validation set [8]. Thapa et al. [36] achieved AUROC ~0.84 in long-term care settings. Complementing these "fall occurrence" models, Wang et al. developed an integrative multi-view ensemble model to predict injurious/severe inpatient falls among fall incidents, achieving AUC 0.808 (95% CI: 0.740–0.876) on a prospective test set, and AUC 0.86 in a high-risk subset—illustrating how machine learning can augment bedside screening by prioritizing patients most likely to experience harmful outcomes [9].

Fall prevention has a low tolerance for false negatives, i.e. missed fall-risk patients, which is why existing tools cast a wide net. Our threshold analysis explicitly allows tuning this balance. For example, if a unit is overwhelmed with alerts, they can raise the threshold modestly. Furthermore, by examining trade-offs, we inform discussions with hospital leadership. This kind of data-driven planning is not possible with legacy checklist scores (e.g., Morse, STRATIFY, Hendrich II) that typically yield point totals/categories rather than calibrated probabilities and do not support unit-specific probability cut-offs [4,29–31]. As a result, units cannot explicitly tune alert thresholds to local capacity, e.g., how many alerts must be evaluated to prevent one fall. Probabilistic ML models can provide such risk estimates, but to be actionable they must also expose why an alert fired and what factors to target. Accordingly, explainable ML methods such as SHAP can help translate model outputs into clinician-interpretable drivers of risk.

The black-box nature of ML models often complicates their interpretation in clinical settings. Therefore, the role of XAI methods, such as SHAP, is critical to build trust among medical professionals [48]. Detailed exploration of different patient risk profiles provides invaluable insights for individualized and subgroup-specific interventions [7, 13]. The emphasis on personalized fall risk assessment is consistent with several scientific findings [49−51] and the World Falls Guidelines assertion of the efficacy of tailored interventions [6]. Instead of a black-box risk score, the SHAP value approach allows for listing patient-specific factors. We found that some factors are consistently important, such as age and fall history, while others vary widely between individuals, e.g., mental health with younger female patients affected. This highlights the need for subgroup-specific strategies. Additionally, the model trained using only guideline factors performed poorly. Although the SHAP plot showed that well-known fall risk factors were among the most important, one may conclude that many factors should be considered when evaluating heterogeneous risk profiles, such as those associated with fall events.

A major impetus for this work was to operationalize the recommendations of the World Falls Guidelines in a hospital setting. Montero-Odasso et al. [6] call for multifactorial falls risk assessments and personalized interventions for those at high risk. In current practice, when a patient is labeled high-risk, generic precautions are applied, but personalized assessments are often not done due to time constraints. In intended use, the proposed model is a decision-support component that (i) flags elevated fall risk and (ii) assigns patients to an interpretable risk profile to guide selection of guideline-consistent prevention measures. Thus, a first-pass multifactorial assessment can prospectively be performed automatically: dozens of factors are being analyzed, and it is indicated which ones are driving a patient's risk via SHAP

analysis. Then, by categorizing the patient's risk profile via clustering, clinicians can match the patient to an "intervention template" aligned with guideline-recommended domains.

### 5.1 Limitations and Future Work

Our study was retrospective and conducted at a single large hospital with several campuses and departments; fall reporting and documentation practices differ elsewhere, potentially affecting model performance. We did not cover all data that can influence falls, such as medication administration records—integrating pharmacy data and real-time vitals could further improve the model. Regarding cluster creation using SHAP values, it is important to note that SHAP only explains how a trained model arrives at its predictions. This means that SHAP is only as good as the underlying model, which was tried to be optimized for the case at hand in this study. Second, our risk profiles are derived via clustering in SHAP space on a held-out cohort after training the prediction model. In a real-world deployment, the prediction model can provide risk scores (and patient-level SHAP explanations) immediately, but risk profile labels require an initial "burn-in" period or a separate profile-discovery cohort with sufficient alerted patients and outcome variation to learn stable clusters. Once learned, risk profiles can be frozen for prospective nearest-centroid assignment and periodically refreshed as case mix or documentation practices change.

Moreover, the physical performance measures, such as the patient's walking and transfer abilities, were not available in all cases since they are not routinely included in the EHR for general inpatients. Future prospective studies might integrate a quick mobility assessment to see if the model improves. Another limitation is that we evaluated interventions theoretically via clusters but did not implement a cluster-driven intervention program and measure fall reduction. That is an important next step: taking this model into a prospective trial where units apply tailored interventions guided by our system, versus units using usual care, to see if fall rates decrease or resources are saved.

## 6 Conclusion

We have demonstrated a methodology to not only predict which patients are at high risk of falls, but to also elucidate why and to guide what to do about it. This represents a step toward both *precision* and *prevention* in healthcare, using machine learning to inform individualized care plans. The approach aligns with contemporary falls prevention guidelines that call for moving beyond broad risk categories to nuanced, patient-specific strategies. By integrating our model into clinical workflow, hospitals can potentially improve patient safety by reducing falls, allocating preventive resources more efficiently, and involving patients in understanding their own risk and prevention plan. The insights

about the importance of nursing care data also encourage health systems to leverage the rich information nurses collect as part of advanced clinical decision support. Overall, our study supports the premise that data-driven, explainable models can enhance patient-centered care in complex domains like fall prevention, fulfilling the promise of healthcare analytics to improve outcomes in a way that is transparent and trustable for hospital staff.

## Statements and Declarations

**Ethics approval.** This retrospective study was approved by the Charité Ethics Commission (Approval No. EA2/184/21). Original approval dated 20 August 2021; amendment approved on 2 December 2024 (amendment submitted on 20 November 2024). For this retrospective study, informed consent was waived by the ethics board. All procedures were performed in accordance with the Declaration of Helsinki and its later amendments (or comparable ethical standards). For retrospective analyses of routinely collected data, informed consent was waived by the ethics committee where applicable.

**Data availability**. The data that support the findings of this study contain sensitive patient information and are not publicly available. De-identified data may be made available from the corresponding author upon reasonable request and subject to institutional approvals and data-protection regulations.

**Code availability**. The codes presented in this study are available on request from the corresponding author.

**Competing interests.** The authors declare that they have no competing interests.

**Author contributions.** Matthias Schulte-Althoff: Writing—original draft; conceptualization; data curation; formal analysis; investigation; methodology; project administration; visualization. Peter Krappen: Writing—original draft; review & editing; conceptualization; data curation; formal analysis; methodology. Felix Biessmann: Review & editing; conceptualization; supervision; project administration. Sebastian Jäger: Review & editing; supervision. Rahel Gubser: Writing; review & editing; conceptualization. Armin Hauß: Investigation; validation; resources. Felix Balzer: Supervision; review & editing. Tim Kilgus: formal analysis, review. Daniel Fürstenau: Review & editing; conceptualization; funding acquisition; data curation; methodology; project administration. All authors read and approved the final manuscript.

# 7 REFERENCES


1. Chen Y, Dai F., Huang S, Qi D, Peng C, Zhang A, Wang Y, Gu Y, Guo J (2025) Global, regional, and national burden of falls among older adults: findings from the Global Burden of Disease Study 2021 and Projections to 2040. npj Aging 11, 85. https://doi.org/10.1038/s41514-025-00275-4

2. Heinrich S, Rapp K, Rissmann U, Becker C, König HH (2010) Cost of falls in old age: A systematic review. Osteoporosis International 21:891–902. https://doi.org/10.1007/S00198-009-1100-1

3. Hshieh TT, Yang T, Gartaganis SL, Yue J, Inouye SK (2018) Hospital Elder Life Program: Systematic Review and Meta-analysis of Effectiveness. The American Journal of Geriatric Psychiatry 26:1015–1033. https://doi.org/10.1016/j.jagp.2018.06.007

4. Park SH (2018) Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. Aging Clinical and Experimental Research 30:1–16. https://doi.org/10.1007/s40520-017-0749-0

5. Morris ME, Webster K, Jones C, Hill AM, Haines T, McPhail S (2022) Interventions to reduce falls in hospitals: a systematic review and meta-analysis. Age and Ageing 51(5):afac077. https://doi.org/10.1093/ageing/afac077

6. Montero-Odasso M, van der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J (2022) World guidelines for falls prevention and management for older adults: a global initiative. Age and Ageing 51:afac205. https://doi.org/10.1093/ageing/afac205

7. Hsu Y-C, Weng H-H, Kuo C-Y, Chu T-P, Tsai Y-H (2020) Prediction of fall events during admission using eXtreme Gradient Boosting: a comparative validation study. Scientific Reports 10:16777–16777. https://doi.org/10.1038/s41598-020-73776-9

8. Ye C, Li J, Hao S, Liu M, Jin H, Zheng L (2020) Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm. International Journal of Medical Informatics 137:104105–104105. https://doi.org/10.1016/j.ijmedinf.2020.104105

9. Wang L, Xue Z, Ezeana CF, et al. (2019) Preventing inpatient falls with injuries using integrative machine learning prediction: a cohort study. npj Digital Medicine 2:127. https://doi.org/10.1038/s41746-019-0200-3

10. Garriga R, Mas J, Abraha S, Nolan J, Harrison O, Tadros G, Matic A (2022) Machine learning model to predict mental health crises from electronic health records. Nature Medicine 28:1240–1248. https://doi.org/10.1038/s41591-022-01811-5

11. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G (2014) Big data in healthcare: using analytics to identify and manage high-risk and high-cost patients. Health Affairs 33:1123–1131. https://doi.org/10.1377/hlthaff.2014.0041

12. Shmueli G, Koppius OR (2011) Predictive analytics in information systems research. MIS Quarterly 35:553–572. https://doi.org/10.2307/23042796

13. Mishra AK, Skubic M, Despins LA, Popescu M, Keller J, Rantz M, Abbott C, Enayati M, Shalini S, Miller S (2022) Explainable fall risk prediction in older adults using gait and geriatric assessments. Frontiers in Digital Health 4:869812–869812. https://doi.org/10.3389/fdgth.2022.869812

14. Hur S, Lee Y, Park J (2025) Comparison of SHAP and clinician friendly explanations reveals effects on clinical decision behaviour. npj Digital Medicine 8:578–578. https://doi.org/10.1038/s41746-025-01958-8

15. Ivziku D, Matarese M, Pedone C (2011) Predictive validity of the Hendrich Fall Risk Model II in an acute geriatric unit. International Journal of Nursing Studies 48:468–474. https://doi.org/10.1016/j.ijnurstu.2010.09.002

16. Cho EH, Woo YJ, Han A, Chung YC, Kim YH, Park H-A (2020) Comparison of the predictive validity of three fall risk assessment tools and analysis of fall-risk factors at a tertiary teaching hospital. Journal of Clinical Nursing 29:3482–3493. https://doi.org/10.1111/jocn.15387

17. Matarese M., Ivziku D., Bartolozzi F., Piredda M., De Marinis M.G. (2014) Systematic review of fall risk screening tools for older patients in acute hospitals. Journal of Advanced Nursing 71(6), 1198–1209. https://doi.org/10.1111/jan.12542

18. Turner K, McNett M, Potter C, Cramer E, Al Taweel M, Shorr RI, Mion LC (2023) Alarm with care-a de-implementation strategy to reduce fall prevention alarm use in US hospitals: a study protocol for a hybrid 2 effectiveness-implementation trial. Implementation Science 18(1):70. https://doi.org/10.1186/s13012-023-01325-9

19. Miotto R, Li L, Kidd B, Dudley JT (2016) Deep Patient: An unsupervised representation to predict the future of patients from the electronic health records. Scientific Reports 6:26094–26094. https://doi.org/10.1038/srep26094

20. Loftus TJ, Shickel B, Balch JA, Tighe PJ, Abbott KL, Fazzone B (2022) Phenotype clustering in health care: a narrative review for clinicians. Frontiers in Artificial Intelligence 5:842306–842306. https://doi.org/10.3389/frobt.2022.842306

21. Lundberg S, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), pp. 4768–4777

22. Li Y, Zhang X, Qian B, Gao Z, Guan C, Zheng Y, Zheng H, Wu F, Li C (2021) Towards Interpretability and Personalization: A Predictive Framework for Clinical Time-series Analysis. In: 2021 IEEE International Conference on Data Mining (ICDM). IEEE, Auckland, New Zealand, pp 340–349

23. Carr O, Javer A, Rockenschaub P, Parsons O, Durichen R (2021) Longitudinal patient stratification of electronic health records with flexible adjustment for clinical outcomes. In: Roy S, Pfohl S, Rocheteau E, Tadesse GA, Oala L, Falck F, Zhou Y, Shen L, Zamzmi G, Mugambi P, Zirikly A, McDermott MBA, Alsentzer E (eds) Proceedings of Machine Learning for Health. PMLR, pp 220–238

24. (DNQP) (2022) Expertenstandard Sturzprophylaxe in der Pflege—2. Aktualisierung 2022. DNQP an der Hochschule Osnabrück. https://www.dnqp.de/fileadmin/HSOS/Homepages/DNQP/Dateien/Expertenstandards/Sturzprophylaxe/Sturz_2Akt_Auszug.pdf

25. Yang Y, Hu X, Zhang Q, Zou R (2016) Diabetes mellitus and risk of falls in older adults: a systematic review and meta-analysis. Age and Ageing 45:761–767. https://doi.org/10.1093/ageing/afw140

26. Kojima G (2015) Frailty as a predictor of future falls among community-dwelling older people: A systematic review and meta-analysis. Journal of the American Medical Directors Association 16:1027–1033. https://doi.org/10.1016/j.jamda.2015.06.018

27. Lan X, Li H, Wang Z, Chen Y (2020) Frailty as a predictor of future falls in hospitalized patients: A systematic review and meta-analysis. Geriatric Nursing 41:69–74. https://doi.org/10.1016/j.gerinurse.2019.01.004

28. Trevisan C, Crippa A, Ek S, Welmer AK, Sergi G, Maggi S, Manzato E, Bea JW, Cauley JA, Decullier E, Hirani V, LaMonte MJ, Lewis CE, Schott AM, Orsini N, Rizzuto D (2019) Nutritional Status, Body Mass Index, and the Risk of Falls in Community-Dwelling Older Adults: A Systematic Review and Meta-Analysis. Journal of the American Medical Directors Association 20:569-582.e7. https://doi.org/10.1016/j.jamda.2018.10.027

29. Morse JM, Morse RM, Tylko SJ (1989) Development of a scale to identify the fall-prone patient. Canadian Journal on Aging 8:366–377. https://doi.org/10.1017/S0714980800008576

30. Oliver D, Britton M, Seed P, Martin FC, Hopper AH (1997) Development and evaluation of evidence-based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: case-control and cohort studies. BMJ 315:1049–1053. https://doi.org/10.1136/bmj.315.7115.1049

31. Gray-Miceli D, Greenberg S (2007) Fall risk assessment for older adults: the Hendrich II fall risk model. American Journal of Nursing 107:50–58.

32. Podsiadlo D, Richardson S (1991) The Timed "Up & Go": A Test of Basic Functional Mobility for Frail Elderly Persons. Journal of the American Geriatrics Society 39:142–148. https://doi.org/10.1111/j.1532-5415.1991.tb01616.x

33. Tinetti ME (1986) Performance-oriented assessment of mobility problems in elderly patients. Journal of the American Geriatrics Society 34:119–126. https://doi.org/10.1111/j.1532-5415.1986.tb05480.x

34. Ossai CI, O'Connor L, Wickramasinghe N (2020) Real-time Inpatients Risk Profiling in Acute Care: A Comparative Study of Falls and Pressure Injuries Vulnerabilities. In: 33rd Bled eConference – Enabling Technology for a Sustainable Society.

35. Chu W-M, Kristiani E, Wang Y-C, Lin Y-R, Lin S-Y, Chan W-C (2022) A model for predicting fall risks of hospitalized elderly in Taiwan—a machine learning approach based on both electronic health records and comprehensive geriatric assessment. Frontiers in Medicine 9:937216–937216. https://doi.org/10.3389/fmed.2022.937216

36. Thapa R, Garikipati A, Shokouhi S, Hurtado M, Barnes G, Hoffman J (2022) Predicting falls in long-term care facilities: machine learning study. JMIR Aging 5:e35373–e35373. https://doi.org/10.2196/35373

37. Huang X, Latt M, Khushi M, Pelicioni P, Brodie MA, Lord SR, Loy C, Poon SK (2021) A Multi-view Classification Framework for Falls Prediction: Multiple-domain Assessments in Parkinson's Disease. In Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS). 1–9. https://doi.org/10.24251/HICSS.2021.413

38. Yu S, Chai Y, Samtani S, Liu H, Chen H (2024) Motion Sensor–Based Fall Prevention for Senior Care: A Hidden Markov Model with Generative Adversarial Network Approach. Information Systems Research 35:1–15. https://doi.org/10.1287/isre.2023.1203

39. Mithas S, Chen Y, Lim Y, Silveira ADO (2022) On causality and plausibility of treatment effects in operations management research. Production and Operations Management 31:4558–4571. https://doi.org/10.1111/poms.13863

40. Kent DM, Steyerberg E, van Klaveren D (2018) Personalized evidence-based medicine: predictive approaches to heterogeneous treatment effects. BMJ 363:k4245–k4245. https://doi.org/10.1136/bmj.k4245

41. Dahabreh IJ, Hayward R, Kent DM (2016) Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. International Journal of Epidemiology 45:2184–2193. https://doi.org/10.1093/ije/dyw125

42. Holland PW (1986) Statistics and causal inference. Journal of the American Statistical Association 81:945–960. https://doi.org/10.2307/2289064

43. Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66:688–701. https://doi.org/10.1037/h0037350

44. Lee C, Rashbass J, van der Schaar M (2021) Outcome-oriented deep temporal phenotyping of disease progression. IEEE Transactions on Biomedical Engineering 68(8):2423–2434. https://doi.org/10.1109/TBME.2020.3041815

45. Chen H, Covert IC, Lundberg SM, Lee S (2023) Algorithms to estimate Shapley value feature attributions. Nature Machine Intelligence 5:590–601. https://doi.org/10.1038/s42256-023-00657-x

46. Jethani N, Sudarshan M, Covert I, Lee S-I, Ranganath R (2022) FastSHAP: Real-Time Shapley Value Estimation. ICLR. https://doi.org/10.48550/arXiv.2107.07436.

47. Kruskal JB, Wish M (1978) Multidimensional Scaling. Sage Publications

48. Zhang Y, Weng Y, Lund J (2022) Applications of explainable artificial intelligence in diagnosis and surgery. Diagnostics 12(2):237. https://doi.org/10.3390/diagnostics12020237

49. Radecki B, Reynolds S, Kara A (2018) Inpatient fall prevention from the patient's perspective: a qualitative study. Applied Nursing Research 43:114–119. https://doi.org/10.1016/j.apnr.2018.08.001

50. Spano-Szekely L, Winkler A, Waters C, DeAlmeida S, Brandt K, Williamson M (2019) Individualized fall prevention program in an acute care setting: an evidence-based practice improvement. Journal of Nursing Care Quality 34:127–132. https://doi.org/10.1097/NCQ.0000000000000344

51. LeLaurin JH, Shorr RI (2019) Preventing falls in hospitalized patients: state of the science. Clinics in Geriatric Medicine 35:273–283. https://doi.org/10.1016/j.cger.2019.01.002

# Appendix

## Appendix A: Odds Ratios and Associated Metrics for Fall Risk Assessment Categories at the University Hospital

We evaluated the university hospital's fall risk assessment, which categorizes patients as "0" (no increased risk of falling), "+", or "++". Using logistic regression with a 90%/10% train-test split, the regression coefficients and model statistics are shown in Table A1. Table A2 displays odds ratios, 95% confidence intervals, LLR p-value, and pseudo R-squared. Patients with a "+" score are over four times (OR 4.17) and "++" score over six times (OR 6.31) more likely to fall than those scored "0".

**Table A1**. Results of logistic regression with fall incidents as DV and the university hospital's fall risk assessment as IV

| | | | | | | |
|---|---|---|---|---|---|---|
| No. Observations: | | | 511.551 | Pseudo R-squ.: | | 0.07142 |
| LLR p-value: | | | 0.000 | Log-Likelihood: | | -31983. |
| | coef | std err | z | P >\|z\| | [0.025 | 0.975] |
| const | -3.8824 | 0.103 | -37.654 | 0.000 | -4.085 | -3.680 |
| 0 | -1.4846 | 0.106 | -13.978 | 0.000 | -1.693 | -1.276 |
| + | 0.2461 | 0.104 | 2.361 | 0.018 | 0.042 | 0.451 |
| ++ | 1.2958 | 0.113 | 11.449 | 0.000 | 1.074 | 1.518 |

**Table A2.** Odds Ratios and Associated Metrics for Fall Risk Assessment Categories at the University Hospital

| *Feature* | *OR* | *p-value* | *2.5%* | *97.5%* | *LLR p-value* | *Pseudo R-squared* |
|---|---|---|---|---|---|---|
| 0 (no increased risk of falling) | 0.17 | 0.0 | 0.160 | 0.178 | 0.000 | 0.065 |
| **+** (increased risk of falling) | 4.17 | 0.0 | 3.971 | 4.384 | 0.000 | 0.045 |
| **++** (significantly increased risk of falling) | 6.31 | 0.0 | 5.766 | 6.905 | 0.000 | 0.014 |

Table A3 provides an overview when oversampling for the training sets, so there are no longer any imbalances and then computing a logistic regression.

**Table A3.** Results of logistic regression as modeled in Table B1 except for oversampling

| | *Precision* | *Recall* | *F1-Score* | *Support* |
|---|---|---|---|---|
| 0.0 | 0.99 | 0.66 | 0.79 | 56133 |
| 1.0 | 0.03 | 0.73 | 0.05 | 707 |
| Accuracy | | | 0.66 | 56840 |
| Macro avg | 0.51 | 0.69 | 0.42 | 56840 |
| Weighted avg | 0.98 | 0.66 | 0.78 | 56840 |

F1 Score: 0.05; AUC: 0.69%

## Appendix B: Most Common Diagnoses & Procedures

The most common procedures for fallen and non-fallen patients are shown in Table B1, and the most common diagnoses are shown in Table B2.

**Table B1** Most common procedures for fallen and non-fallen patients

| ***Fallen Patients*** | | ***Non-fallen Patients*** | |
|---|---|---|---|
| **Procedure group** | **Quan. (%)** | **Procedure group** | **Quan. (%)** |
| Computer Tomography CT | 17,326 (16.33) | Computer Tomography CT | 607,914 (15.07) |
| Circulatory Measures | 13,227 (12.47) | Circulatory Measures | 413,399 (10.25) |
| Radiotherapy, nuclear medical therapy, and chemotherapy | 6,260 (5.90) | Radiotherapy, nuclear medical therapy, and chemotherapy | 260,039 (6.45) |
| Patient monitoring | 5,610 (5.29) | Examination of individual body systems | 222,475 (5.51) |
| Treatment for mental and psychosomatic, and behavioral disorders in adults | 5,170 (4.87) | Surgeries on the locomotor organs | 188,950 (4.68) |

**Table B2** Most common diagnoses for fallen and non-fallen patients

| ***Fallen Patients*** | | ***Non-fallen Patients*** | |
|---|---|---|---|
| **Diagnosis group** | **Quan. (%)** | **Diagnosis group** | **Quan. (%)** |
| Forms of heart disease | 5,339 (5.35) | Potential health risks based on family/personal history | 413,035 (5.36) |
| Metabolic disorders | 5295 (5.31) | Forms of heart disease | 338,275 (4.39) |
| Potential health risks based on family/personal history | 4,197 (4.21) | Metabolic disorders | 321,760 (4.17) |
| Hypertension | 3,184 (3.19) | Hypertension | 258,783 (3.36) |
| Diseases of the muscles | 2,812 (2.82) | Persons using the health care system for the purpose of specific measures and medical care | 171,177 (2.22) |

## Appendix C: SHAP-based risk profiles by nearest-centroid assignment

**(I) Distance Mapping.** For a new patient $p$, we define their SHAP value vector $Sp$ as:

$$Sp = (s_1, s_2, ..., s_n),$$

where $s_i$ represents the SHAP value for feature $i$, and $n$ is the total number of features used in the fall risk prediction model; thus each $s_i$ represents how much that particular feature contributes to the model's prediction for that patient. Each cluster $C_k$ ($k \in \{1,...,m\}$) with $m$ being the total number of identified clusters is represented by its centroid vector:

$$\mu_k = (\mu_{k1}, \mu_{k2}, ..., \mu_{kn}),$$

where $\mu_{ki}$ is the mean SHAP value for feature $i$ in cluster $k$. The distance $d(p,k)$ between a patient $p$ and a cluster centroid $\mu_k$ is calculated using a Euclidean distance:

$$d(S_p, \mu_k) = \|S_p - \mu_k\|_2 = \sqrt{(\Sigma\ (s_i - \mu_{ki})^2)}.$$

**(II) Assignment Criteria.** We define two metrics for cluster assignment that stabilize decisions under local explanation noise and scale differences compared with plain L2 distance:

1. *Primary Distance Threshold* ($\tau_1$): A patient $p$ is considered assigned to cluster $k$ if: $d(p,k) \leq \tau_1$, where $\tau_1$ is a threshold to be determined as the $n$th (e.g., $90^{th}$) percentile of intra-cluster distances from the cluster-discovery set.

2. *Relative Distance Ratio* ($R$): $R(p,k) = d(p,k) / (\min_{j \in \{1, ..., m\} \setminus \{k\}} d(p,j))$. A patient is uniquely assignable to cluster $k$ if: $R(p,k) \leq \tau_2$, where $\tau_2$ is a predefined ratio threshold. This measures how close the patient $p$ is to cluster $k$ relative to other clusters. The condition $R(p,k) \leq \tau_2$ should ensure that the assignment is sufficiently unique.

The final assignment function $A(p)$ is defined as:

$$A(p) = \arg\min_k d(p,k) \text{ s.t. } d(p,k) \leq \tau_1 \wedge R(p,k) \leq \tau_2, \text{ else “unassigned”.}$$

## Appendix D: Global Risk Factors and SHAP Explanations

For the L3 model chosen, we examined the aggregated local feature importances using SHAP. Fig. D1 is a SHAP summary plot for the top features contributing to fall risk. Each point in Fig. D1 represents a patient; points are plotted horizontally by SHAP value, i.e., the impact on model output, and are colored by the feature value (e.g., for age: red = older, blue = younger). Fall history (number of falls in the 12 months before admission) emerged as the single most influential predictor. Some nursing features and procedural codes were particularly notable. E.g., early rehabilitative and physical therapy procedures (OPS 8-55) showed that patients in rehabilitation therapy often had higher fall risk. This likely reflects confounding by procedure—these patients are in rehabilitation because of mobility impairments, which also put them at risk of falling.

High-risk medications and comorbidities also appeared. For instance, the model included features for psychotropic medication use, captured indirectly via procedure code "application of medications" (OPS 6-00) and diagnoses like depression or dementia. Male sex had a small positive SHAP on average, consistent with the slight male excess in fallers, though sex was far less important than age or medical factors. Hypertension (ICD I10-I15) and metabolic disorders (ICD E70-E90) had nuanced effects—they were common in both fallers and non-fallers, so on average they weren't among the very top predictors, but in combination with other factors they sometimes elevated risk, for example, a patient with hypertension plus balance issues might be at risk due to orthostatic hypotension from antihypertensive medication. Overall, the SHAP summary indicates that our model's predictions are driven by a mix of direct fall risk indicators, such as prior falls, age, mobility impairments, and indirect proxies of vulnerability, such as complex treatment courses and need for monitoring.

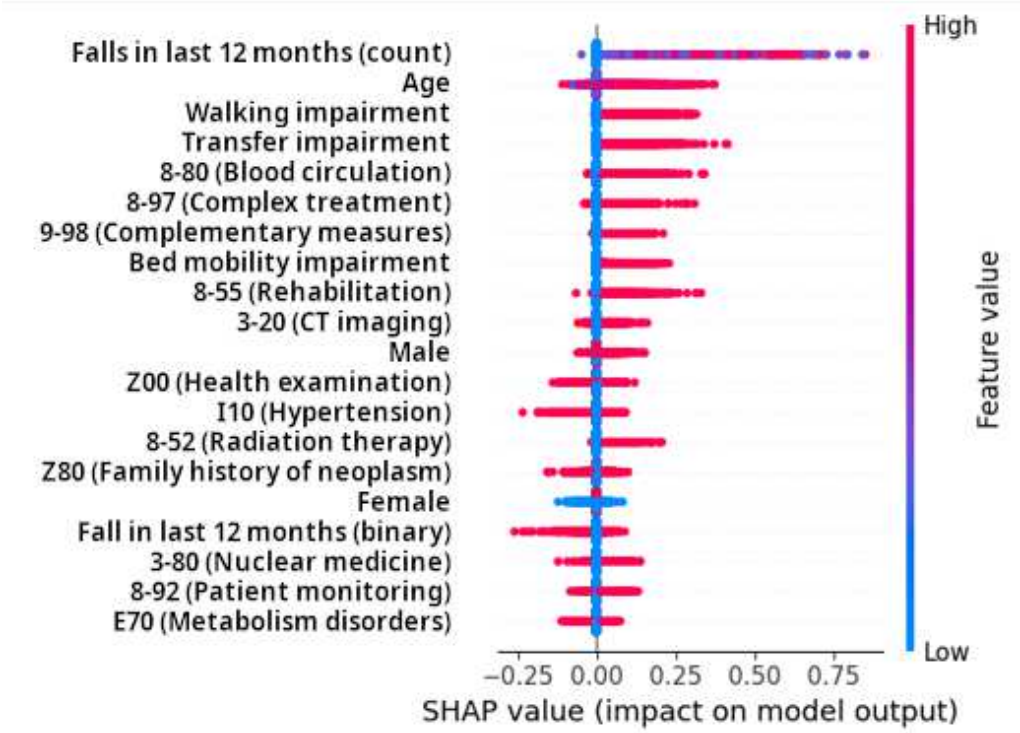



**Fig. D1** SHAP summary plot of top 20 features for fall risk prediction. The features are ranked by mean absolute SHAP value, i.e. the largest average magnitude of contribution to the model output, regardless of direction. Points to the right indicate that the feature increases fall risk, to the left decreases risk. Color indicates the feature value, i.e., red for high, blue for low. For continuous features like age, red means older age. For categorical features (procedures, diagnoses), red indicates presence of that code, blue indicates absence.

## Appendix E: Detailed Cluster Description

**Cluster 1: "Miscellaneous High Care Needs"**—Cluster 1 was the largest cluster (23% of cases) and appears to represent a general high-dependency group without a single defining specialty. This cluster seems to be patients who are generally of high acuity or dependency (perhaps needing help with basic care) but not fitting one of the more specific profiles.

**Cluster 2: "Complex Care—Younger, Multi-system"**—Cluster 2 patients required high levels of care despite being relatively younger with a mean of 61.2 years, a younger cluster. They had frequent recent falls, complex care needs, and often had circulatory system procedures, and frequent CT imaging. This paints a picture of patients with serious illnesses, such as trauma or severe chronic diseases that necessitate intensive interventions.

**Cluster 3: "Behavioral Abnormality"—**This was a relatively small cluster of patients with mental health and psychosomatic disorders as dominant features. They tended to be younger (average ~64 years) and more often female (only 31% male). These patients often had psychiatric conditions or delirium contributing to falls.

**Cluster 4: "Frail & Rehabilitation-Focused"**—Patients in Cluster 4 were generally very elderly and frail. This cluster had the highest average age (87.3 years). Key features included a history of multiple falls (high SHAP for "falls in last 12 months"), early rehabilitation and physical therapy orders, and mobility impairments (difficulty walking/transferring). In essence, Cluster 4 represents geriatric patients with severe mobility limitation and general frailty.

**Cluster 5: "Complex Treatments"**—Cluster 5 patients were those undergoing highly complex medical/surgical treatments, often in intensive or step-down care. Their average age was ~77.8 years, and 57% were male. They frequently had multiple recent procedures; the standout feature was OPS 8-97/8-98 "complex treatment," which is ICU care or complex surgery. They also had elevated fall risk from a recent fall history and were often hypertensive.

**Cluster 6: "Complex Treatments with Metabolic Component"**—Patients in Cluster 6 were similar to Cluster 5 but with an added profile of metabolic disorders, particularly diabetes. Average age was ~81.8 (slightly older than Cluster 5), 64% male. They had the highest frequency of "complex treatment" codes of any cluster, indicating many were receiving invasive procedures or intensive therapies. They also often had gait or transfer impairments noted and hypertension, but a distinguishing factor was the presence of metabolic disease codes in their top features. Many of these patients were diabetics with complications, or other metabolic issues like electrolyte imbalances or obesity.

**Cluster 7: "Oncology & Circulatory Interventions"**—Cluster 7 consisted of patients with active cancer treatments and other major medical interventions. Average age 64.9, 50% male. The salient features include measures for blood circulation (OPS 8-80 to 8-85) and radiation/chemotherapy procedures (OPS 8-52 to 8-54). This indicates many were oncology patients receiving chemo or radiotherapy, often with concurrent cardiovascular procedures. They also frequently had "application of medications" codes (6-00). These patients likely fall because of treatment side effects: chemotherapy can cause neuropathy, fatigue, anemia; radiation can cause weakness; extensive diagnostic tests mean they are in unfamiliar environments often.

——**Cluster 8: "Frequent Monitoring**"—Cluster 8 included patients around 84.5 years old on average. They were characterized by continuous patient monitoring, meaning they likely were on telemetry or similar, and had high frequency of falls in last year, plus hypertension and a history of different health risk factors. So cluster 8 seems to be older patients under a lot of medical monitoring, maybe for cardiac or neurological issues.

—**Cluster 9: "Rehabilitation with Specialized Therapy (Phoniatric)**"—Cluster 9 patients share features with Cluster 4 (early rehab, walking/transfer impairments) but also have phoniatric and audiology therapy in their profile. This suggests patients who are in rehabilitation settings and also receiving speech or hearing therapy—possibly stroke patients or others with neurological deficits affecting speech/swallow. They also have complex care needs.

## Appendix F: Cluster-level Descriptions

For each clustering method and each resulting cluster, we report cluster size ($n$), number of falls, observed fall incidence with 95% Wilson confidence interval, and mean predicted fall risk. Clusters are identified by the method-specific cluster label (C0, C1, …). If a minimum cluster-size rule is applied (e.g., <1%), this is reflected in $k_{effective}$ as reported in Table 4.

**Table F1** Cluster-level outcomes in the high-risk evaluation cohort (alert threshold p≥0.08)

| Method | Cluster | n (patients) | % of cohort | n (Falls) | Observed incidence | 95% Wilson CI | Mean predicted risk | SD predicted risk |
|---|---|---|---|---|---|---|---|---|
| SHAP_KMeans | 0 | 1604 | 17.55 | 212 | 0.13 | [0.12, 0.15] | 0.15 | 0.11 |
| SHAP_KMeans | 2 | 272 | 2.98 | 28 | 0.10 | [0.07, 0.14] | 0.15 | 0.09 |
| SHAP_KMeans | 5 | 93 | 1.02 | 49 | 0.53 | [0.43, 0.63] | 0.44 | 0.28 |
| SHAP_KMeans | 6 | 143 | 1.56 | 43 | 0.30 | [0.23, 0.38] | 0.27 | 0.11 |
| SHAP_KMeans | 7 | 340 | 3.72 | 91 | 0.27 | [0.22, 0.32] | 0.30 | 0.24 |
| SHAP_KMeans | 8 | 6583 | 72.03 | 1010 | 0.15 | [0.14, 0.16] | 0.16 | 0.13 |
| RAW_KMeans | 0 | 1892 | 20.70 | 302 | 0.16 | [0.14, 0.18] | 0.15 | 0.13 |
| RAW_KMeans | 1 | 525 | 5.74 | 84 | 0.16 | [0.13, 0.19] | 0.16 | 0.10 |
| RAW_KMeans | 2 | 1221 | 13.36 | 192 | 0.16 | [0.14, 0.18] | 0.15 | 0.09 |
| RAW_KMeans | 3 | 2089 | 22.86 | 268 | 0.13 | [0.11, 0.14] | 0.17 | 0.15 |
| RAW_KMeans | 4 | 2263 | 24.76 | 305 | 0.13 | [0.12, 0.15] | 0.16 | 0.09 |
| RAW_KMeans | 6 | 672 | 7.35 | 224 | 0.33 | [0.30, 0.37] | 0.29 | 0.26 |
| RAW_KMeans | 7 | 133 | 1.46 | 20 | 0.15 | [0.10, 0.22] | 0.15 | 0.09 |
| RAW_KMeans | 8 | 307 | 3.36 | 54 | 0.18 | [0.14, 0.22] | 0.22 | 0.09 |
| RAW_GMM | 0 | 1896 | 20.75 | 300 | 0.16 | [0.14, 0.18] | 0.15 | 0.12 |
| RAW_GMM | 1 | 518 | 5.67 | 82 | 0.16 | [0.13, 0.19] | 0.16 | 0.10 |
| RAW_GMM | 2 | 1218 | 13.33 | 190 | 0.16 | [0.14, 0.18] | 0.15 | 0.09 |
| RAW_GMM | 3 | 2088 | 22.85 | 267 | 0.13 | [0.11, 0.14] | 0.17 | 0.15 |
| RAW_GMM | 4 | 2243 | 24.54 | 302 | 0.13 | [0.12, 0.15] | 0.16 | 0.09 |
| RAW_GMM | 6 | 699 | 7.65 | 234 | 0.33 | [0.30, 0.37] | 0.29 | 0.26 |
| RAW_GMM | 7 | 133 | 1.46 | 20 | 0.15 | [0.10, 0.22] | 0.15 | 0.09 |
| RAW_GMM | 8 | 307 | 3.36 | 54 | 0.18 | [0.14, 0.22] | 0.22 | 0.09 |
| PCA_KMeans | 0 | 767 | 8.39 | 128 | 0.17 | [0.14, 0.19] | 0.17 | 0.10 |
| PCA_KMeans | 2 | 1570 | 17.18 | 239 | 0.15 | [0.14, 0.17] | 0.16 | 0.09 |
| PCA_KMeans | 3 | 2143 | 23.45 | 271 | 0.13 | [0.11, 0.14] | 0.17 | 0.15 |
| PCA_KMeans | 4 | 1727 | 18.90 | 278 | 0.16 | [0.14, 0.18] | 0.15 | 0.13 |
| PCA_KMeans | 7 | 2145 | 23.47 | 289 | 0.13 | [0.12, 0.15] | 0.16 | 0.09 |
| PCA_KMeans | 8 | 695 | 7.60 | 230 | 0.33 | [0.30, 0.37] | 0.28 | 0.26 |
| RAW_Ward | 0 | 2055 | 22.49 | 384 | 0.19 | [0.17, 0.20] | 0.22 | 0.21 |
| RAW_Ward | 1 | 2348 | 25.69 | 362 | 0.15 | [0.14, 0.17] | 0.16 | 0.09 |
| RAW_Ward | 2 | 149 | 1.63 | 18 | 0.12 | [0.08, 0.18] | 0.15 | 0.09 |
| RAW_Ward | 4 | 1744 | 19.08 | 316 | 0.18 | [0.16, 0.20] | 0.16 | 0.15 |
| RAW_Ward | 5 | 2446 | 26.76 | 335 | 0.14 | [0.12, 0.15] | 0.15 | 0.09 |
| RAW_Ward | 7 | 357 | 3.91 | 35 | 0.10 | [0.07, 0.13] | 0.15 | 0.11 |
| RAW_Spectral | 0 | 149 | 1.63 | 22 | 0.15 | [0.10, 0.21] | 0.17 | 0.13 |
| RAW_Spectral | 1 | 1949 | 21.33 | 461 | 0.24 | [0.22, 0.26] | 0.21 | 0.21 |
| RAW_Spectral | 3 | 106 | 1.16 | 15 | 0.14 | [0.09, 0.22] | 0.16 | 0.09 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RAW_Spectral | 4 | 438 | 4.79 | 82 | 0.19 | [0.15, 0.23] | 0.17 | 0.13 |
| RAW_Spectral | 5 | 154 | 1.69 | 17 | 0.11 | [0.07, 0.17] | 0.18 | 0.15 |
| RAW_Spectral | 6 | 4707 | 51.50 | 687 | 0.15 | [0.14, 0.16] | 0.16 | 0.10 |
| RAW_Spectral | 7 | 1085 | 11.87 | 98 | 0.09 | [0.07, 0.11] | 0.16 | 0.12 |
| RAW_Spectral | 8 | 513 | 5.61 | 64 | 0.12 | [0.10, 0.16] | 0.17 | 0.13 |
| RiskBin | 0 | 8725 | 95.47 | 1111 | 0.13 | [0.12, 0.13] | 0.14 | 0.07 |
| RiskBin | 3 | 94 | 1.03 | 75 | 0.80 | [0.71, 0.87] | 0.68 | 0.02 |

**Notes**: Clusters <1% were collapsed for reporting (excluded from the method-level separation metrics), so the effective number of clusters ($k_{effective}$) may be <9. Observed incidence is computed on the held-out evaluation cohort (alerts at $p \geq 0.08$), and Wilson CIs are reported for binomial proportions. Mean predicted risk is the model's mean predicted probability within each cluster.

Fig. F1 visualizes the same cluster-level outcomes (observed incidence with 95% Wilson CI and mean predicted risk), highlighting the stronger outcome gradient for SHAP-space clustering.

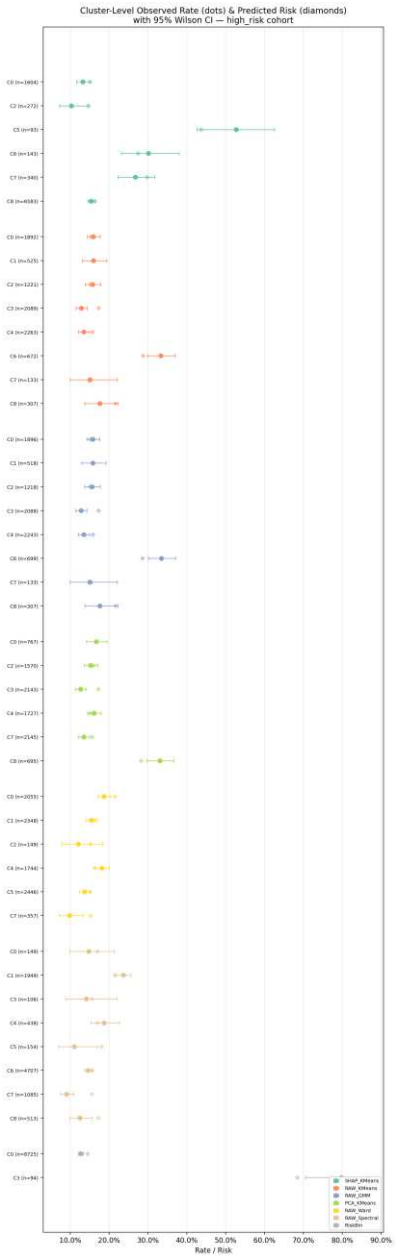



**Fig. F1** Cluster-level observed fall incidence (dots, 95% Wilson CI) and mean predicted risk (diamonds) in the high-risk cohort ($p \geq 0.08$).