IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

A Mixed Integer Programming Formulation for Risk Stratification

Rachda Naila Mekhaldi, Julia L. Fleck, Raksmey Phan, and Xiaolan Xie, *Fellow, IEEE*

***Abstract*—Risk stratification is the process of segmenting patients into distinct groups of similar complexity and care needs in order to improve resource allocation. Patients are typically risk stratified using statistical or machine learning methods that generate an individual risk score for some measure of resource use. One of the main limitations of existing methods is reduced interpretability, which is often inherent to artificial intelligence techniques. In this work, we propose a novel risk stratification approach that optimizes the representation of different patient groups and generates interpretable risk profiles. We associate risk scores to patient profiles and determine the optimal combination of representative profiles for each patient group using a Mixed Integer Programming (MIP) formulation. We generate continuous ratings for patient risk scores ranging from 0 to 1 that allow for dynamic thresholding. Our method stratifies patients into several risk groups (e.g., low, medium, high risk), which is frequently more clinically significant than binary classification. We apply our approach to both public and proprietary real data in the context of accidental fall risk assessment and show that the generated risk profiles provide clinical insights that can be used for the design of targeted interventions.**

***Index Terms*—Risk stratification, MIP, fall risk assessment.**

I. Introduction

RISK stratification is the process of segmenting patients into distinct groups of similar complexity and care needs in order to improve resource allocation. For the purposes of this work, *risk* refers to clinical risk, or the likelihood of an adverse clinical outcome. The main goals of risk stratification are (i) to identify individuals who are at high risk of an undesirable clinical event and to offer these patients a proactive intervention designed to reduce their risk of experiencing the event, and (ii) to optimize the level of resources offered to each patient and improve health outcomes at the population level. To date, most of the existing risk stratification developments address the issue of unplanned hospital admissions over a 12-month period. However, risk stratification is applicable to any event that is low quality, gives a poor patient experience and is high cost, and can be derived over different time periods (e.g., 30 days, 6 months, 2 years, etc.). Other examples of feasible events include accidental falls, loss of independence, or readmissions to a hospital.

Traditionally, healthcare professionals have risk-stratified patients intuitively using clinical judgement or simple threshold-based rules. The intuitive approach implies the use of clinical knowledge and training, combined with knowledge of the patients, to identify individuals at high risk of requiring unplanned healthcare. Threshold-based rules are a "catch-all" method that identifies any individual who meets a defined high-risk threshold (e.g., "any patient over the age of 65 who has had two or more hospital admissions in the previous 12 months"). This approach has been shown to be of limited effectiveness because threshold-based rules are highly susceptible to selection bias and regression to the mean [1].

More recently, statistical or machine learning methods have been proposed that generate an individual risk score for some measure of resource use [2], [3], [4], [5], [6], [7], [8]. Although such approaches show strong prognostic performance, there is increasing evidence that these models generate representations of the clinician's reasoning manifested through their clinical actions [9]. This means that model performance may only represent the diagnostic value of the clinician-initiated data on which the models rely. Moreover, a critical challenge in providing clinical recommendations is identifying an explanation for each recommendation. However, machine-learning methods typically offer limited interpretability of the models and their predictions.

To address these limitations, in this paper we propose a structured, data-driven algorithmic approach to determine patient risk levels. To our knowledge, our work is the first to frame patient risk stratification as a Mixed Integer Programming (MIP) problem. Our contributions are threefold: (i) we solve the risk stratification problem efficiently by generating patient profiles with associated risk scores and then selecting the optimal combination of representative profiles for each patient group; (ii) we generate interpretable risk profiles that provide clinical insights and can help the design of targeted interventions; (iii) our method is well adapted to lower resolution datasets that do not inherently capture physician behavior. In our experiments, we use only self-reported patient data and minimally invasive patient data that do not include clinician-initiated data.

II. Problem Setting

This paper considers the interpretable stratification of a set of subjects or patients $X = \{x_1, x_2, \ldots, x_J\}$. Each subject x_i is characterized by a vector of binary features $\mathbf{k}_i = [k_{i1}, k_{i2},$

Received 13 August 2025; accepted 12 December 2025. Date of publication 12 January 2026; date of current version 26 January 2026. This article was recommended for publication by Associate Editor N. Kong and Editor C.-B. Yan upon evaluation of the reviewers' comments. This work was supported in part by Fondation MAIF and in part by Saint-Etienne Métropole. *(Corresponding author: Rachda Naila Mekhaldi.)*

The authors are with Mines Saint-Etienne, Univ Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158 LIMOS Saint-Etienne, France (e-mail: naila.mekhaldi@emse.fr).

Digital Object Identifier 10.1109/LRA.2026.3653307

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

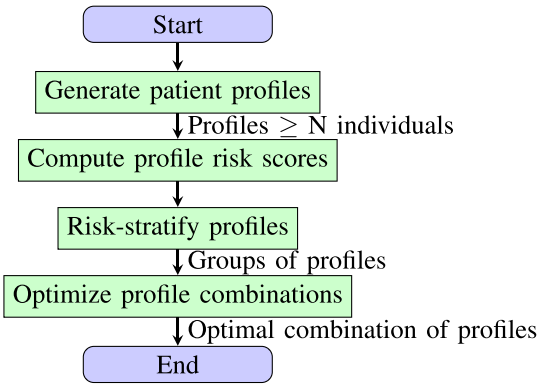

Fig. 1. Methodology to solve the patient risk stratification problem.

..., k_{iK}] and a binary indicator y_i of whether the subject has experienced an adverse event of interest. Let y_i be a realization of a Bernoulli random variable $Y \sim \text{Ber}(r_i)$ associated with individual i, where $r_i \in [0, 1]$ is the true but unknown risk of the individual. The problem consists in partitioning (stratifying) X into I groups X_1, X_2,..., X_I of subjects in decreasing order of risk of adverse events and explaining the profiles of different groups.

The main challenge here is that only the realization of the adverse event is known but the stratification relies on the risk level of each subject. In reality, the realization of the adverse event for each subject x of feature vector $\mathbf{k} = [k_1, k_2,..., k_K]$ is random depending on the corresponding risk level r or adverse event probability. The true risk r of each subject x is unknown and instead we know only the random realization y_i of each subject x_i. As a result, two subjects x_i and x_j with the same feature vector $\mathbf{k}$ might have different adverse event realizations with $y_i = 1$ and $y_j = 0$.

Another major challenge is achieving a balance between the risk level and the statistical significance of a profile. For instance, a profile defined by many binary features might have a high risk level but concern only fewer subjects, being hence statistically insignificant. On the contrary, a profile defined by one or two binary features may be statistically significant with risk level close to the global average.

In the following, only profiles of at least a pre-defined minimum population size are considered to ensure statistical significance, and a rigorous risk score is defined in order to measure the risk level of subjects and their partition into different risk groups. Finally, an optimization approach is given to discover the most meaningful profiles for each risk group. Importantly, the problem and the method proposed in this paper extend to numerical and multi-valued qualitative features, and multi-valued adverse event realization. This extension clearly leads to a problem of feature space explosion that we do not address in this paper.

III. Methodology

We propose a four-step methodology to solve the problem of patient risk stratification (Fig. 1). We begin by identifying feature subsets that correspond to profiles with different risk scores. In the second step, only profiles containing at least N individuals are retained to ensure statistical robustness (Section III-A). In the third step, a risk score is computed for each retained profile (Section III-B). The fourth step involves stratifying individuals according to their profile-specific risk levels using statistical analysis (Section III-C). Finally, a Mixed Integer Programming (MIP) formulation is applied to identify the optimal combination of representative profiles for each patient group (Section III-D).

A. Generation of Profiles

Understanding risk in healthcare requires analyzing interactions between multiple features. While individual features provide valuable insights, their combination captures more complex patterns that may influence the risk value. By combining features into meaningful profiles, we can better represent the underlying patterns associated with risk factors and enhance the interpretability of the results.

In this context, our first step consists of determining subsets of features that define profiles associated with different risk scores. To increase model interpretability and reduce computation time, the maximum number of features per profile is limited to a constant value M. If we consider a set of K binary features so that each profile can have between 2 and M features, the total number of possible profiles C is given in Eq. (1). Here we exclude single-feature profiles as they do not capture feature interactions.

$$C = \sum_{i=2}^{\text{M}} \frac{K!}{i!(K-i)!} \tag{1}$$

Each resulting profile is a binary feature combination of up to M elements. To ensure that each profile contains a sufficiently large number of patients, we apply a threshold-based policy and retain only those profiles that contain at least N individuals. The value of N is directly proportional to the number of individuals in the dataset. The number of individuals in profile p represents a subset of the dataset for which all the features in p are true.

Although our method can represent OR/NOT relationships, this study focuses on identifying the most representative feature combinations for each group, which are more efficiently captured using AND-type combinations. An individual is said belonging to profile p if its values of features in p are all equal to 1.

B. Risk Score Evaluation

In binary classification problems, the predicted outcome is a discrete label (e.g., "yes/no", "1/0"). For healthcare applications, it is often preferable to generate a continuous output, which can then support healthcare professionals in decision making. Continuous outputs are particularly relevant for patient risk stratification. For our purposes, we define a positive class as the subgroup of patients who have experienced at least one instance of the adverse event of interest.

In this work, we generate a real-valued risk score for each profile retained in the previous step. As defined in Section III-A, a profile is a combination of binary features. For each profile p

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Algorithm 1: Attribute risk scores to patient profiles.

Input: Dataset of D individuals with K binary features
Parameter: Minimum number of individuals N, Maximum number of features M
Output: Set of profiles P and set of risk scores R
$P = \emptyset$
Generate feature combinations S of size [2, M] from K features
for s in S **do**
  if Count(individuals with all features in s) $\geq N$ **then**
    Add profile defined by s to P
  end if
end for
for p in P **do**
  $R(p) = \frac{\text{Number of positive class individuals in } p}{\text{Total number of individuals in } p}$
end for
return P, R

in the set of profiles P, we compute a risk score that takes values ranging from 0 to 1. Let $R(p)$ be the risk score for profile p. The value of $R(p)$ is calculated using Eq. (2), where $N_f(p)$ represents the number of individuals in the positive class belonging to profile p and $N(p)$ represents the total number of individuals belonging to profile p.

$$R(p) = \frac{N_f(p)}{N(p)} \tag{2}$$

The risk score $R(p)$ measures the proportion of individuals who have experienced at least one instance of the undesirable event within profile p, which is defined by a specific combination of features. Hence, $R(p)$ represents the observed risk of a certain event within a population of interest. The pseudo-code for profile generation and risk score evaluation is given in Algorithm 1.

The complexity of Algorithm 1 depends mainly on the number of feature combinations generated and the computational effort required to determine how many individuals match each combination. More specifically, for a profile p of i features, the computation of $R(p)$ requires comparing the i feature values for each of the D individuals and has a time complexity of $D \cdot i$. Hence the time complexity of Algorithm 1 can be expressed as $\mathcal{O}(D \cdot \sum_{i=2}^{M} i \cdot \binom{K}{i})$, which simplifies to $\mathcal{O}(D \cdot K \cdot 2^K)$ in the worst case. The case of large number of subjects and large number of features is beyond the scope of this paper, as discussed in Section V.

C. Risk Score Stratification

Up to this point, we have computed a risk score associated with a patient profile. However, any given patient may belong to more than one profile. For the purposes of patient stratification, it is therefore necessary to determine the value of the risk score for each patient. To do so, we first identify all the profiles to which any single patient belongs. Let P the set of profiles defined by a combination of binary features representing individual characteristics. For each individual patient x_i, we identify all profiles $P_i \subseteq P$ such that x_i satisfies every criteria for each profile $p \in P_i$.

Next, we retrieve the risk score values associated with all the profiles in P_i. Let $R(x_i)$ be the risk score of patient x_i. We take the value of $R(x_i)$ to be the maximum of all risk score values among all the profiles to which x_i belongs, as in Eq. (3). This conservative approach assigns individual risk scores based on the most vulnerable profile. While it may overestimate risk in some cases, this choice is intentional, prioritizing the identification of high-risk profiles and aligning with clinical goals of prevention over underestimation.

$$R(x_i) = \max_{p \in P_i} R(p) \tag{3}$$

Once a value of $R(x_i)$ has been assigned to each patient, we compute quartile statistics for individual risk scores considering all the patients in our dataset. We adopt quartiles to define risk groups following standard practice in clinical risk stratification. This ensures statistically grounded thresholds. As a result, we obtain $Q1$ and $Q3$ values of the individual risk scores, which correspond to the first and third quartiles, respectively. We then proceed by stratifying the profiles based on the values of $Q1$ and $Q3$ such that $0 \leq Q1 \leq Q3 \leq 1$. This allows us to partition the set of patients and the set of profiles. The set of patients X is partitioned into three groups: Group 3 (high risk) is defined as $X_3 = \{x_i \in X : R(x_i) > Q3\}$; Group 2 (intermediate risk) is defined as $X_2 = \{x_i \in X : Q1 < R(x_i) \leq Q3\}$; Group 1 (low risk) is defined as $X_1 = \{x_i \in X : R(x_i) \leq Q1\}$. The set of profiles P can be partitioned into three groups correspondingly.

D. Optimal Profiling

To increase interpretability of our results, we seek to identify the profiles that can best represent each of the groups generated in Section III-C. In other words, for each group we wish to select a subset of at most Z profiles which maximize the sum of risk scores for the given group. The value of Z is defined by the user, as detailed in Section IV. Because each profile is a combination of binary features, the number of generated profiles can increase rapidly as the number of initial features grows, leading to a significant expansion in the number of generated profile combinations. This is typical of combinatorial optimization problems. Here we present a Mixed Integer Programming (MIP) formulation to determine the optimal combination of representative profiles for each patient group.

Let I be the set of individuals. Each individual $i \in I$ is represented by x_i. For each profile $p \in P$ and individual $i \in I$, we define binary decision variables $x_p \in \{0, 1\}$ and $y_{ip} \in \{0, 1\}$, where $x_p = 1$ indicates that profile p is selected, and $y_{ip} = 1$ indicates that individual i is assigned to profile p. We also define Z as the maximum number of profiles that can be selected. We then formulate the optimal search for profiles as:

$$\text{maximize} \sum_{p} \sum_{i \in p} R(p) \cdot y_{ip} \tag{4}$$

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

subject to:

$$\sum_p x_p \leq Z \tag{5}$$

$$y_{ip} \leq \mathbf{1}_{\{i \in p\}} \cdot x_p \quad \forall i, p \tag{6}$$

$$\sum_p y_{ip} \leq 1 \quad \forall i \tag{7}$$

The optimal profiling leads to the subject risk estimate in Eq. (8), which is smaller than or equal to the real risk.

$$\hat{R}(x_i) = \max_{p \in P_i \wedge x_p = 1} R(p) \tag{8}$$

The objective function (4) maximizes the total risk of a profile combination by identifying, for each group of individuals, the Z profiles of highest risk. An alternative formulation would be to maximize the number of subjects covered with $R(p) = 1$ in order to identify the Z most representative profiles. Constraint (5) ensures that the number of selected profiles is no greater than Z, (6) ensures the assignment of an individual to selected profiles containing it, (7) assigns an individual to at most one profile, and (8) estimates the risk of patient x_i.

Our formulation addresses the problem of managing a large number of stratified profiles. For one, (4) maximizes the sum of individual risks for the chosen profiles, thus guaranteeing that the most relevant profiles are selected, which is essential for insightful analysis. Furthermore, our formulation reduces the dimensionality of the original feature set by restricting the number of selected profiles to Z. Accordingly, the results are easier to interpret and to handle computationally. This is particularly crucial for healthcare applications, where decision-makers wish to understand the rationale behind patient profiling.

We note that model complexity can be further reduced without loss of optimality by relaxing the integrity constraint of y_{ip}. More specifically, for each risk group of patients x_i and the corresponding group P_i of profiles, a standard MIP solver can be used to solve the optimal profiling problem (4)-(9) and the resulting optimal solution x gives the most representative profiles of the group, i.e. the list of profiles with $x_p = 1$. In fact, for each given combination of profiles $\mathbf{x}$, y_{ip} of the optimal solution equals to 1 for $p = \pi(i, \mathbf{x})$ and 0 for all other profiles p where $\pi(i, \mathbf{x})$ is the highest reward profile among all selected profiles p containing individual i, i.e. $i \in p$ and $x_p = 1$. Algorithm 2 summarizes the search process for optimal profile combinations.

Algorithm 2: Search for optimal profile combination.

Input: Dataset of X individuals with K features, Risk scores R, Profiles P
Parameter: Maximum combination size Z
Output: Optimal combinations C for each group
$C = \emptyset$
for x in X **do**
  $P_d =$ profiles containing d
  $R(x) = \max_{p \in P_d} R(p)$
end for
Calculate Q_1 and Q_3
Assign profiles to group based on quartiles
Assign individuals to group based on their profiles
for group g **do**
  $c =$ MIP(Z, maximize sum of risk scores)
  Add c to C
end for
return C

IV. Results

In what follows, we illustrate the use of our approach to stratify patients according to the risk of accidental fall. We conducted experiments using both public and proprietary real data generated in the context of accidental fall risk assessment.

A. Data Description

The proprietary dataset is derived from self-reported questionnaires and includes general housing information, sociodemographic factors, and detailed information about fall incidents [1]. Although the dataset also includes socioeconomic features including income, the goal of this work is to generate proof-of-concept results for the proposed MIP-based risk stratification approach; we do not, therefore, address the issue of resource allocation policies that could be derived from such features.

The dataset is imbalanced, with fallers accounting for only 16.74% of the cohort. The public dataset was previously described in [10], and includes sociodemographics and data gathered using cognitive and psychological evaluations such as the LASA Fall risk profile, Mini-Mental State Examination (MMSE) score, and self-reported questionnaires. In this paper, we choose to avoid the use of intrusive data, such as wearable devices (e.g., accelerometers); we therefore excluded data from [10] that had been collected using an accelerometer. The cohort in the public dataset contains 34.55% of fallers.

B. Experiments

Preprocessing was performed in Python using Scikit-Learn, and the MIP was solved with CPLEX. Code will be released upon acceptance; the dataset cannot be shared due to regulatory constraints.

In all our experiments, we let Z=3. Only profiles that contain a minimum of N=50 individuals (proprietary dataset) or N=30 individuals (public dataset) are retained for optimal profiling. The values of these parameters were selected based on a trade-off between statistical significance, interpretability, and computational feasibility. The parameter M, defining the maximum number of features per combination, should be set case-by-case, potentially guided by domain expertise, to ensure human-understandable combinations. Similarly, Z, the maximum number of profiles retained per combination, is chosen to maintain a manageable number of representative profiles while covering the data adequately. Finally, N, the minimum number of individuals required per profile, is adjusted according

[1] https://www.observatoire-mavie.com/login

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE I
DATASETS DESCRIPTION

Dataset	Individuals	Original Features		Binary Features K	Generated Profiles (2 to 6 features)	Selected Profiles
		Numerical	Categorical			
Proprietary	5,025	6	18	64	83,277,936	4,929,024
Public	301	8	12	45	9,530,994	351,334

TABLE II
PROFILES RETAINED FOR OPTIMAL PROFILING

Dataset	High Risk		Intermediate Risk		Low Risk	
	Profiles	Individuals	Profiles	Individuals	Profiles	Individuals
Proprietary	204	1,338	2,645	2,652	4,926,175	1,035
Public	308	82	9,238	163	341,788	56

TABLE III
OPTIMAL PROFILES TO REPRESENT DIFFERENT FALL RISK LEVELS (RESULTS FROM THE PROPRIETARY DATASET)

Risk Level	Profile 1	Profile 2	Profile 3
High Risk	Age 72–102; socio-professional level middle management; no consumption of 6 glasses of alcohol; monthly income $>$ 2500; no rug in back of bathroom.	Socio-professional level executive management; height 135–165 cm; frequent alcohol consumption; home with outdoor space; home with no water body; retirement income resources.	Age 72–102; monthly income $>$ 2500; alcohol consumption 1–2 glasses; settlement population: 5K–30K; no rug in bathroom exit; house with 7–27 rooms.
Intermediate Risk	Weight 80–168 kg; alcohol consumption every 2–3 weeks; home with no water body; other place of residence; living area 86–130 m^2.	Weight 37–67 kg; height 165–174 cm; no rug in bathroom exit; BMI 12.8–23.73 kg/m^2; home with gas heating; professional income resources.	No pet; age 65–72 years; BMI: 12.8–23.73 kg/m^2; professional income resources; dwelling type house; event type: daily.
Low Risk	Height 135–165 cm; tobacco consumption; no consumption of 6 glasses of alcohol; has pet; home with water body; rug in bathroom exit.	Tobacco consumption; BMI 12.8–23.73 kg/m^2; rug in bathroom exit; event type: daily; home with garage; home with electric heating.	Home with no garage; home with 1–5 rooms; home with gas heating; professional income resources; other dwelling type.

TABLE IV
OPTIMAL PROFILES TO REPRESENT DIFFERENT FALL RISK LEVELS (RESULTS FROM THE PUBLIC DATASET)

Risk Level	Profile 1	Profile 2	Profile 3
High Risk Level	Age 71–78 years; MMSE score 29–30; experienced fall in past 6 months or year; no pet; LASA score 6–19.	MMSE score 20–27; experienced fall in past 6 months or year; no pet; no frequent dizziness; can ascend/descend stairs.	Age 78–96 years; experienced fall in past year; no frequent alcohol consumption; LASA score 6–19; depression level 6–21; fear of falling 20–52.
Intermediate Risk Level	Female; can ascend/descend stairs; LASA score 6–19; depression level 6–21.	Lives independently; no partner; experienced fall in past 6 months; can use public transportation; depression level 2–6; no walking aid.	Lives independently; MMSE score 27–29; no fall in past year; can clip own toenails; higher education; depression level 0–2.
Low Risk Level	Male; MMSE score 29–30; no frequent dizziness; has a pet; higher education; no walking aid.	Age 71–78 years; male; lives with partner; MMSE score 29–30; no frequent dizziness; depression level 2–6.	Lives with partner; MMSE score 29–30; no fall in past 6 months; no frequent dizziness; LASA score 3–6.

to dataset size, with smaller datasets requiring lower values to retain meaningful profiles. In our experiments, these parameters were selected to optimize interpretability and performance while remaining computationally feasible. We discretized numerical features into categorical ones using the quantile discretization technique [11] followed by one-hot encodin [12].

Table I shows the number of individuals and features contained in each dataset, as well as the total number of generated profiles with at most $M = 6$ features. The size and complexity of the two datasets used in the study vary greatly. The proprietary dataset has a larger number of features (64 binary features vs. 45in the public dataset), which leads to a considerably greater number of generated profiles (nearly 83 million for the proprietary dataset vs. 9.5 million for the public dataset). Only a small portion of the many possible profiles are kept for stratification based on the minimum number of individuals required per profile (N).

In Table II we present the number of profiles retained per risk group for profile optimization. The proposed MIP model aims to identify the optimal combination of up to Z profiles that maximizes the overall risk score within each group. These selected combinations, referred to as the most pertinent profiles, are intended to support clinical decision-making. For example, in the high-risk group of the proprietary dataset, only 3 profiles that best represent the group will be selected from a total of 204 profiles.

The profiles that optimize the risk score per group are shown in Tables III and IV. An individual is assigned to a profile combination if they belong to at least one profile within the combination. Our MIP model assigns each patient to a fall risk level (high, intermediate, or low) to enhance real-world applicability and make it easier for healthcare professionals to interpret results. Three different profiles are used to represent each risk level, and a patient is placed in a risk group if they meet the requirements of at least one profile in that group. In

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

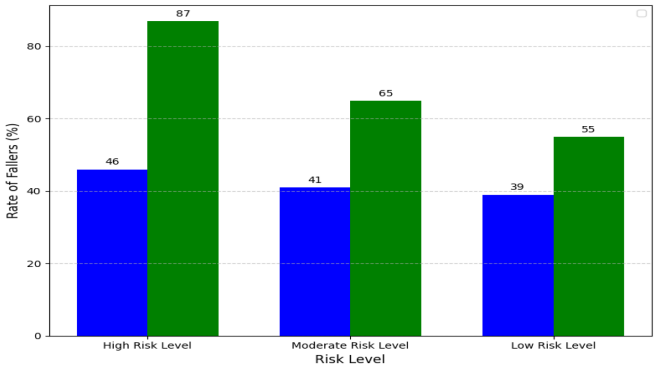

Fig. 2. Relative percentage of patients with a confirmed history of falls by risk level (proprietary dataset in blue and public dataset in green).

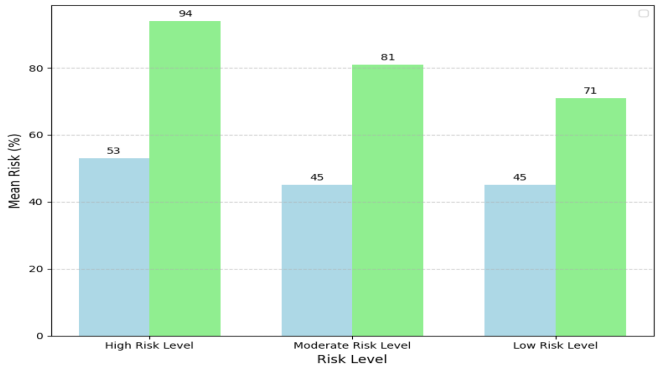

Fig. 3. Average risk per group (proprietary dataset in light blue and public dataset in light green).

addition, a patient is only assigned to a particular profile if they meet all of the binary conditions in that profile simultaneously.

Results for the proprietary dataset indicate that age is a key risk factor for fall risk. This corroborates established clinical knowledge. Our results reveal additional factors, such as the presence of bathroom rugs that are associated with lower risk of fall. Interestingly, our results also suggest that living in households with a water body lowers the risk of falling, while the presence of exterior spaces increases the risk of accidental falls.

Results for the public dataset reveal that previous falls are consistently associated with high risk of subsequent falls across all profile. Higher MMSE scores (29-30, indicating better cognitive function) are generally linked to lower risk. Additionally, living with a partner and the absence of dizziness seem to have a protective effect, being associated with lower risk. Depression levels correlate with risk, with higher depression scores observed in higher risk categories. Common protective factors include being male, having a higher education level, living with a partner, having good cognitive function (MMSE 29-30), not experiencing dizziness, and not requiring walking aids. Interestingly, the absence of a pet appears to increase risk. Furthermore, the ability to perform specific activities, such as climbing stairs or clipping one's toenails, seems to be a risk indicator.

Fig. 2 presents the relative percentage of patients with a confirmed history of falls by risk level while Fig. 3 highlights the average risk score for each group. In Fig. 2, we observe a clear descending trend in fall rates from high to low risk levels in both datasets. The public dataset consistently shows higher fall rates at all risk levels, reflecting a greater proportion of fallers compared to the proprietary dataset. The high-risk group exhibits the highest fall rates in both datasets. The intermediate-risk group shows a distinct separation from the high-risk group, while the low-risk group still experiences relatively high fall rates. In Fig. 3, a descending trend in average risk scores is observed from high to low risk levels, mirroring the trend from Fig. 2. The public dataset consistently shows higher average risk percentages, with the gap most pronounced at the high-risk level. Within the risk groups, the highest scores are found in the high-risk group for both datasets, with a notable drop in the public dataset and a smaller decline in the proprietary dataset for the intermediate risk group. Interestingly, the proprietary dataset's low-risk group exhibits the same risk level as the moderate-risk group.

Taken together, our results indicate the effectiveness of our approach for risk stratification, as fall rates align consistently with the identified risk levels. The significant differences observed between the public and proprietary datasets can be attributed to the distinct characteristics of their respective populations. Specifically, the proprietary dataset contains broader data on life accidents, such as strokes, rather than being confined solely to falls. In contrast, the public dataset was generated during a clinical study focusing on frailty. Furthermore, variations in environmental factors and reporting methods may also have contributed to these discrepancies.

C. Discussion

In healthcare applications, the main consideration in risk score generation is whether a score is needed or can be shown to provide clinical benefit. Risk scores should reduce uncertainty, prompt missed diagnoses, increase efficiency, and improve outcomes. In the context of accidental falls, the most commonly used risk assessment tools in clinical practice are the Hendrich II Fall Risk Model (HFRM II) [13], the Morse Fall Scale (MFS) [14] and the St. Thomas Risk Assessment Tool (STRATIFY) [15]. Previous studies report the use of AI-based tools to predict fall risk for an elderly population. In [16], machine-learning techniques, including Random Forest and Deep Neural Network models, were used to analyze the data, predict patient falls, and identify the most important risk factors for falls. A classification-based approach was proposed in [17] to distinguish fallers from non-fallers using both Ada-boost and Decision Tree algorithms. The XGBoost algorithm was used in [4] to identify older adults at higher risk for fall and analyze the associated risk factors. To date, the majority of studies on fall risk assessment has focused on the ability to discriminate between fallers and non fallers, on determining cutoff values, and on assessing their sensitivity and specificity [20], [21]. Table V provides a comparison between existing fall risk assessment methods and the proposed MIP-based approach based on four key criteria: interpretability (i.e., whether the approach provides

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE V
COMPARISON OF FALL RISK ASSESSMENT METHODS

Method	Interpretability	Outputs	Static Risk Scores	Automatic Stratification
Clinical Risk Assessment Tools				
Hendrich II Fall Risk Model (HFRM II) [13]	High	Numerical	Yes	No
Morse Fall Scale (MFS) [14]	High	Numerical	Yes	No
St. Thomas Risk Assessment Tool (STRATIFY) [15]	High	Numerical	Yes	No
AI-Based Methods				
Logistic Regression [5], [6]	Relatively High	Probability Score	Yes	No
Random Forest [16]	Medium	Binary / Probability	Yes	No
Deep Neural Network [16]	Low	Binary / Probability	Yes	No
AdaBoost and Decision Tree [17]	Medium	Binary / Probability	Yes	No
XGBoost [4]	Low	Binary / Probability	Yes	No
CETAF [18]	High	Numerical	Yes	No
Falling Rule List [19]	High	Continuous	Yes	No
Proposed Method				
MIP-based Risk Stratification	High	Continuous	No	Yes

clear justification for its results), output type, whether the risk score is static or can be updated dynamically based on new inputs or recalculations, and whether automatic stratification is performed. We consider clinical assessment tools to be highly interpretable because they are expert-defined and transparent. Machine learning models are categorized according to their transparency: rule-based and linear models (e.g., FRL, CeTAF Score, Logistic Regression) are highly interpretable, tree-based and ensemble models (e.g., Decision Tree, Random Forest, AdaBoost) provide intermediate interpretability, and complex models (e.g., XGBoost, Deep Neural Networks) have limited explainability and interpretability.

To our knowledge, existing methods cannot be applied to our datasets due to differences in data characteristics, evaluation protocols, and reported metrics. Hence, a direct numerical comparison is not possible. Since no comparable baselines exist, our method can serve as a foundation for future comparisons.

Existing fall risk assessment tools, such as the Morse Fall Scale (MFS), generate a risk score based on a set of factors, but do not necessarily reveal which factors are most important for predicting falls [22]. In addition, most machine learning methods produce highly complex models which are not designed to provide an ability to reason about each prediction. This leaves a gap, where predictive models are not directly aligned with the clinical decisions that need to be made from them [19]. Moreover, the majority of current models classify individuals into two distinct risk categories: high risk versus low risk. This precludes the identification of intermediate risk patients, with sometimes borderline risk scores, who may benefit from customized interventions since fall risk evolves along a continuum. Also of note is the fact that existing fall risk stratification methods ignore the individual needs and characteristics of particular strata, assuming that features pertinent to the entire dataset are equally relevant across all subpopulations. This, in turn, hinders the generation of customized insights for subgroups with unique risk variables. In sum, existing methods provide either interpretability through clinical tools or predictive power via AI, whereas our MIP-based method integrates both aspects. Moreover, our method accommodates flexibility and automation in defining the risk score and the ability to automatically stratify patients according to their associated risk score, which are lacking in both traditional and AI-based counterparts, while preserving interpretability and rich outputs. As a result, our approach is especially well-suited for medical applications where decision-making must be transparent and flexible.

Moreover, unlike previous approaches, the novel MIP formulation proposed here selects relevant features for each strata of the dataset while considering social determinants of health, demographics, and resource limitations to identify optimal risk profiles. This may enhance equity by ensuring high-risk patients are efficiently identified and appropriately treated.

Of note, our conjunction-based approach provides a straightforward, rule-based classification that is highly interpretable, supporting clinical decision-making by allowing healthcare professionals to tailor interventions to each patient risk category.

For example, consider a patient who experienced a fall in the past 6-12 months, does not own a pet, does not suffer from frequent dizziness, can climb up and down stairs, and has a MMSE score of 22. These characteristics match profile 2 of the high-risk group (Table IV). Therefore, the patient is considered at high risk of falling, and healthcare professionals can try to identify actionable characteristics, which can help determine interventions, by comparing the patient's profile with the profiles of intermediate or low-risk patients. For instance, low-risk patient profile 1 (Table IV) does not suffer from frequent dizziness, similarly to the high-risk patient in question, but has a higher MMSE score and owns a pet. In light of this, the healthcare professional might then recommend pet ownership and/or targeted interventions to increase cognitive function as a means to decrease the patient's risk of fall.

V. CONCLUSION

In this letter, we present a novel Mixed Integer Programming (MIP) formulation for risk stratification aimed at advancing interpretable models for healthcare applications. Our approach relies exclusively on non-invasive and easily collectible data, highlighting its potential for real-world applications in preventive healthcare.

We illustrate our method with an application to accidental fall risk assessment. Our findings consistently indicate that, when the history of falls is available, patients with a confirmed history of accidental falls form the majority of high-risk groups, and

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

the percentage of fallers in each resulting group has a direct relationship with the group's risk score in relation to the total number of patients assigned to that group. The present study focuses on fall risk prediction, but the proposed MIP-based method is applicable to other binary classification problems in the medical domain. Although extending and testing the approach on additional outcomes (e.g., sepsis, readmissions) is beyond the scope of this work, future work will focus on applying the model to larger and more diverse populations, and integrating other prediction tasks in healthcare to offer the ability to identify risks. Additional future work aims at developing heuristic and metaheuristic optimization strategies to enable the scalable analysis of large electronic health records (EHR) datasets while preserving model interpretability. Complexity related to a large number of individuals can be addressed by Monte Carlo sampling as in [23]. Complexity related to the large number of features is more challenging and requires novel optimization techniques that do not rely on the exhaustive enumeration of all profiles. To this end, column generation techniques are a promising direction.

ACKNOWLEDGMENT

The authors would like to thank Calyxis for data curation and acknowledge Tecnalia, Aqui O Thermes, Clinique Marienia, and AESIO Santé as members of the *Previchute* consortium.

REFERENCES

[1] M. Roland and G. Abel, "Reducing emergency admissions: Are we on the right track," *BMJ*, vol. 345, no. 18, Sep. 2012, Art. no. e6017, doi: 10.1136/bmj.e6017.

[2] J. Wiens, J. Guttag, and E. Horvitz, "Patient risk stratification with time-varying parameters: A multitask learning approach," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2797–2819, 2016. [Online]. Available: https://jmlr.org/papers/volume17/15-177/15-177.pdf

[3] Z. Zhang, H. Zhang, and M. K. Khanal, "Development of scoring system for risk stratification in clinical medicine: A step-by-step tutorial," *Ann. Transl. Med.*, vol. 5, no. 21, 2017, Art. no. 436, doi: 10.21037/atm.2017.08.22.

[4] C. Ye et al., "Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm," *Int. J. Med. Informat.*, vol. 137, 2020, Art. no. 104105.

[5] J. Bravo et al., "Development and validation of a continuous fall risk score in community-dwelling older people: An ecological approach," *BMC Public Health*, vol. 21, no. S2, 2021, Art. no. 808.

[6] C. Pereira et al., "Key factor cutoffs and interval reference values for stratified fall risk assessment in community-dwelling older adults: The role of physical fitness, body composition, physical activity, health condition, and environmental hazards," *BMC Public Health*, vol. 21, no. S2, 11 2021, Art. no. 977, doi: 10.1186/s12889-021-10947-x.

[7] M. J. Sillanpää, "Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses," *Heredity*, vol. 106, no. 4, pp. 511–519, 2010, doi: 10.1038/hdy.2010.91.

[8] T. Ching et al., "Opportunities and obstacles for deep learning in biology and medicine," *J. Roy. Soc. Interface*, vol. 15, no. 141, 2018, Art. no. 20170387, doi: 10.1098/rsif.2017.0387.

[9] B. K. Beaulieu-Jones et al., "Machine learning for patient risk stratification: Standing on, or looking over, the shoulders of clinicians?," *NPJ Digit. Med.*, vol. 4, no. 1, 2021, Art. no. 62, doi: 10.1038/s41746-021-00426-3.

[10] K. S. Van Schooten et al., "Daily-life gait quality as predictor of falls in older people: A 1-year prospective cohort study," *PLoS ONE*, vol. 11, no. 7, 2016, Art. no. e0158623.

[11] J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and unsupervised discretization of continuous features," in *Proc. Mach. Learn. Proc.* 1995, A. Prieditis and S. Russell, Eds. San Francisco (CA): Morgan Kaufmann, 1995, pp. 194–202.

[12] T. Al-Shehari and R. A. Alsowail, "An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques," *Entropy*, vol. 23, no. 10, 2021, Art. no. 1258.

[13] A. L. Hendrich, P. S. Bender, and A. Nyhuis, "Validation of the Hendrich II Fall Risk Model: A large concurrent case/control study of hospitalized patients," *Appl. Nurs. Res.*, vol. 16, no. 1, pp. 9–21, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0897189702109025

[14] J. M. Morse, R. M. Morse, and S. J. Tylko, "Development of a scale to identify the fall-prone patient," *Can. J. Aging / La Revue Canadienne Du Vieillissement*, vol. 8, no. 4, pp. 366–377, 1989.

[15] D. Oliver, M. Britton, P. Seed, F. Martin, and A. Hopper, "Development and evaluation of evidence based risk assessment tool (stratify) to predict which elderly inpatients will fall: Case-control and cohort studies," *Brit. Med. J.*, vol. 315, no. 7115, pp. 1049–1053, 1997.

[16] S. Jahandideh et al., "Using machine learning models to predict falls in hospitalised adults," *Int. J. Med. Inform.*, vol. 187, 2024, Art. no. 105436.

[17] P. R. Chakraborty and G. Sorwar, "A machine learning approach to identify fall risk for older adults," *Smart Health*, vol. 26, 2022, Art. no. 100303.

[18] B. Bongue, C. Sass, J. Feroldie, W. Ngan, C. Vannier-Nitenberg, and A. Covez, "Développement d'un programme de prévention du risque de chute adapté à la pratique des professionnels de première ligne development of a falls prevention program adapted to the practice of first-line professionals," *La Revue de gériatrie*, vol. 41, pp. 33–44, 2016.

[19] F. Wang and C. Rudin, "Falling rule lists," in *Proc. 18th Int. Conf. Artif. Intell. Statist.* ser., in Proc. Mach. Learn. Research, G. Lebanon and S. V. N. Vishwanathan, Eds., vol. 38. San Diego, California, USA: PMLR, 09–12 May 2015, pp. 1013–1022.

[20] K. L. Hsieh, L. Chen, and J. Sosnoff, "Mobile technology for falls prevention in older adults," *Journals Gerontol.: Ser. A*, vol. 78, no. 5, pp. 861–868, 2022.

[21] A. Tiedemann, S. R. Lord, and C. Sherrington, "The development and validation of a brief performance-based fall risk assessment tool for use in primary care," *Journals Gerontol.: Ser. A*, vol. 65A, no. 8, pp. 896–903, 2010.

[22] D. S. Lindberg et al., "Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using ehr and electronic administrative data: A machine-learning approach," *Int. J. Med. Informat.*, vol. 143, 2020, Art. no. 104272.

[23] M. Prodel, V. Augusto, B. Jouaneton, L. Lamarsalle, and X. Xie, "Optimal process mining for large and complex event logs," *IEEE Trans. Automat. Sci. Eng.*, vol. 15, no. 3, pp. 1309–1325, Jul. 2018.