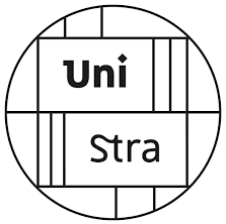

University of Strasbourg

Master 1 STER – Mathematics and Applications
Track : Scientific Computing and Information Mathematics

FRAKITEST Project

Master 1 Internship Report – Academic Year 2024–2025

Prepared by :
Narimane Zaouache

Supervisors :
Laure Combourieu
Pierre-Yves Rohan (IBHGC – ENSAM Paris)

Internship location :
Institut de Biomécanique Humaine Georges Charpak (IBHGC),
Arts et Métiers Institute of Technology (ENSAM ParisTech), Paris

Date : August 2025

Table des matières

Abstract

The aging population represents a major public health challenge, particularly in the early detection of motor frailty. In this context, the Functional Screening Score for Motor Frailty (SFDFM) was designed to provide physiotherapists with a quick, reliable, and practice-ready tool.

This report builds on work initiated in 2022 around the SFDFM. It presents the key steps of its 2025 enhancement : a complete redesign of the digital interface (ergonomics, data security, automation), and an in-depth statistical validation on a sample of 89 patients.

The study assessed the internal robustness of the score, identified the most influential variables, and proposed clinically interpretable thresholds to classify levels of frailty. Results confirm the relevance of the SFDFM, while highlighting areas for improvement, especially when compared to the Fried score, considered the international gold standard.

This work illustrates the value of an integrated approach, combining clinical analysis with digital innovation, to strengthen the use of screening tools in primary care settings.

1 Introduction

1.1 Ageing : a major demographic transition

Population ageing is now a central societal and healthcare issue, both globally and in France. Demographic data show a steady increase in the proportion of older adults. According to the World Health Organization (WHO), by 2050, more than 22% of the world's population will be over the age of 60. In France, INSEE reports that in 2023, 21% of the population was aged 65 or over, compared to 14% in 1991. This trend is driven by three main factors : a long-term decline in birth rates, increased life expectancy, and the ageing of the baby boom generation.

However, ageing does not follow a single path. In fact, several forms exist, influenced by living conditions, personal habits, social environment, and health status. Generally, three major categories are distinguished. **Successful ageing** is characterized by the preservation of functional capacities despite advancing age, reflecting good physiological adaptation and maintained quality of life. **Usual ageing** corresponds to a gradual decline in physiological reserves, without necessarily leading to major disability or loss of autonomy in the short term. Finally, **pathological ageing** refers to an evolution marked by the onset of chronic or neurodegenerative diseases, often associated with loss of independence and increased medical needs.

This ageing process can thus evolve favourably or unfavourably, and should not be considered a uniform fatality. Moreover, it presents new challenges to our healthcare systems, particularly in terms of prevention, autonomy, and quality of life.

1.2 Frailty : a transitional and reversible state of ageing

Between successful ageing and marked dependency lies an intermediate clinical state : **frailty**. This relatively recent concept in geriatrics describes a reduction in an individual's physiological reserves, making them more vulnerable to common health events (infection, falls, treatment changes, hospitalisation, etc.). Unlike dependency, frailty remains potentially reversible if detected early.

Frailty is not a disease, but rather a **multifactorial syndrome**, reflecting an imbalance between stressors experienced by the individual and their ability to cope. It develops progressively and can either lead to recovery or worsen if not properly managed.

Several studies have attempted to model ageing trajectories by including frailty as a key stage. *Jean-Pierre Bouchon* (1984) introduced the notion of a "critical phase of ageing", later illustrated by *Axel Guilbaud* (2020), who highlighted the concept of fluctuating frailty. The main advantage of this idea is that it opens a window of opportunity for intervention, at a time when the person is still autonomous but starting to lose resilience.

The consequences of frailty are well known : it significantly increases the risk of **falls**, **hospitalisations**, **loss of autonomy**, and even **mortality**. Yet frailty often goes underdiagnosed, particularly among people living at home, where it tends to progress silently.

Epidemiological data highlight its growing prevalence. In France, the prevalence of frailty among older adults living at home is estimated to be between 10% and 18%, and that of pre-frailty reaches 43% according to Santé Publique France [1]. These figures underline the urgent need to implement **early screening strategies**, particularly in primary care settings.

As illustrated in Figure 1, frailty lies on a continuum between autonomy and dependency, and may evolve in either direction depending on the interventions implemented.

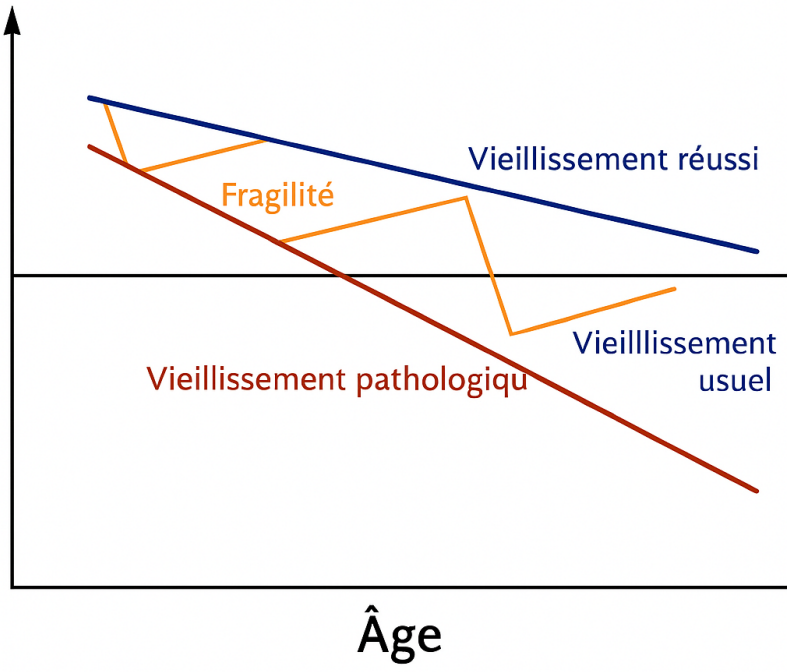

FIGURE 1 – Ageing and frailty trajectories

1.3 Existing frailty assessment models : strengths and limitations

Since frailty is a **multifactorial syndrome**, several assessment models have been developed to quantify and clinically detect it. Among them, two main approaches stand out in the geriatric literature.

Fried's Phenotype (2001)

This model is based on a physiological and unidimensional vision of frailty. It identifies five clinical criteria [3] : unintentional weight loss, self-reported exhaustion, slow walking speed, decreased grip strength, and reduced physical activity.

An individual is considered : **frail** if they meet 3 or more criteria, **pre-frail** if they meet 1 or 2, and **robust** if they meet none.

This model has the advantage of being simple, reproducible, and well-documented. It is often used as a reference in scientific studies. However, its main limitation is that it does not include cognitive, social, or psychological dimensions, which nonetheless play a major role in the overall frailty of older adults.

Rockwood's Frailty Index (2005)

Developed from a multidimensional perspective, this model is based on the principle of deficit accumulation. It includes a large number of variables (symptoms, chronic diseases, cognitive disorders, functional limitations, etc.) to compute a continuous frailty score. The greater the number of deficits, the more frail the person is considered to be.

The strength of this model lies in its comprehensiveness and its ability to reflect the clinical complexity of older adults. However, its routine use in practice is limited : the score is time-consuming to calculate, difficult to standardize, and not well suited for contexts such as private physiotherapy.

To better visualize the differences between these two approaches, Table 1 provides a comparative summary.

Criteria	Fried (2001)	Rockwood (2005)
Approach	Unidimensional	Multidimensional
Number of items	5	$>$ 30 (deficit accumulation)
Ease of use	++	-
Administration time	Short	Long
Clinical integration	Easy	Complex
Cognitive aspects considered	No	Yes

TABLE 1 – Summary comparison of Fried's and Rockwood's frailty models

Despite their usefulness, these tools have several limitations in a field-based screening context : administration time is often too long for systematic use ; their interpretative complexity can hinder routine application ; they are not always validated for certain health professions, such as physiotherapists, or for the French population ; and they offer little direct connection to actionable prevention strategies, which complicates their integration into care pathways.

1.4 Origin and objectives of the SFDFM

The limitations of existing models — including their complexity, administration time, and poor suitability for use in private physiotherapy — highlighted the need to develop a more appropriate screening tool. The goal was to design a score that would be simple, quick, reproducible, and specifically aligned with the role of physiotherapists, particularly for elderly patients living at home.

It was in this context that the **Functional Screening Score for Motor Frailty (SFDFM)** was created in 2022. The project was initiated by the *Public Health and Demographics Commission of the French Order of Physiotherapists* [2], in collaboration with two domain experts : **France Mourey** (university professor specializing in motor function and aging) and **Alexandre Kubicki** (geriatrician specializing in public health).

The SFDFM was developed with several functional objectives in mind. It had to be quick to administer, based on simple clinical tests already familiar to physiotherapists (such as the TUG, single-leg stance, or grip strength test), and include a structured interview addressing falls, pain, and cognitive disorders. Additionally, the tool incorporates relevant sociodemographic data such as the patient's age and living conditions. Together, these components aim to produce a clear, immediately usable total score that allows straightforward classification of motor frailty into three levels : mild, moderate, or severe.

To achieve these goals, the SFDFM is structured into three complementary sections. The first section (A) collects general data such as age, sex, and current treatments. The second section (B) is a targeted interview focusing on clinical signs associated with frailty. The third section (C) consists of a series of standardized functional assessment tests, covering mobility, balance, strength, and transfers.

Designed to be pragmatic and suitable for use in private practice, the SFDFM adheres to methodological standards while remaining compatible with real-world clinical constraints. It is intended to become a reference professional tool, particularly in the context of the upcoming implementation of the **AMK10** billing code (July 2025), which will financially support motor frailty screening in older adults.

1.5 Previous validation work on the SFDFM

To ensure its clinical relevance and methodological robustness, the SFDFM required scientific validation. In 2024, several studies were conducted following the recommendations of the *COSMIN consortium* to assess its psychometric properties : validity, reliability, and discriminant ability.

Clinical validation studies (2024)

Study of the SFDFM's validity The first study, conducted by **Marie Texier**, aimed to evaluate two fundamental aspects of the SFDFM's validity : face validity and concurrent validity. The former corresponds to how healthcare professionals perceive the relevance and coherence of the tool. The latter refers to the ability of the SFDFM to produce results consistent with a recognized reference tool — in this case, the Fried criteria.

For face validity, a questionnaire was sent to 14 physiotherapists. The result : 78.6% of them considered the SFDFM to be relevant for detecting motor frailty, confirming good acceptability within the profession.

Concurrent validity was assessed on a sample of 81 patients over 65 years old living at home. These patients were evaluated using both the SFDFM and Fried's criteria. Statistical analysis revealed a significant correlation, with a Spearman coefficient of 0.582 (ordinal variables) and a Pearson coefficient of 0.70 (continuous variables), indicating good agreement between the two scores.

This study therefore provides solid preliminary evidence of validity, confirming that the SFDFM effectively measures a construct similar to that described in the literature, while incorporating more functional and contextualized dimensions [13].

Study of inter-rater reliability In a second study, **Nicolas Leroy** focused specifically on the **inter-rater reliability** of the SFDFM — that is, its ability to yield consistent results when administered by different professionals.

The study involved 58 patients evaluated by two independent physiotherapists using a standardized protocol. The analysis showed a very high level of agreement, with an intra-class correlation coefficient (ICC) of 0.921, indicating excellent reliability. Cohen's Kappa for categorical data was 0.801, suggesting near-perfect agreement between raters. Finally, Kendall's Tau reached 0.828, confirming a strong correlation for ordered classifications.

These results demonstrate that the SFDFM can be used confidently by different physiotherapists, without major risk of subjectivity or excessive variability in interpretation [7].

Analysis of prevalence and discriminant items Lastly, **Aude Le Ménez** explored the practical use of the SFDFM in actual patient populations monitored by private physiotherapists. Her study focused on two complementary aspects : first, the distribution of motor frailty levels in an elderly population, and second, the identification of the most sensitive or discriminant components of the score.

Among the 85 analyzed scores, results showed a predominance of **moderate frailty** (44.7%), followed by **mild frailty** (29.4%) and **severe frailty** (25.9%). These findings, consistent with European literature, confirm the SFDFM's ability to capture the diversity of geriatric profiles in outpatient care.

Regarding discriminant items, tests such as the *grip strength*, *single-leg stance*, *TUG*, and *chair rise*, as well as interview elements like *pain*, *falls*, or *fear of falling*, were found to be strongly associated with the overall score. These functional indicators are particularly useful for guiding clinical decisions in both screening and follow-up contexts [6].

Initial digital implementation

In parallel with clinical validations, a digitization effort for the SFDFM was initiated by **Dorian Geraldes Pereira**. In 2024, he developed a first interactive prototype of the questionnaire, aiming to facilitate data entry during consultations and initiate a standardized digital collection process.

This prototype enabled direct response entry via a dedicated web interface. Upon submission, the data were automatically transmitted to a secure platform (*Girder*) through an authentication key system. An individual Excel file was generated for each patient, containing all the collected results.

Although this version validated the project's technical foundations, it still had several limitations : the user interface needed improvement, the storage system was not optimized for large-scale use, and the tool lacked the modularity required to adapt to the diverse needs of healthcare professionals.

1.6 Internship Objectives

As a continuation of the validation work conducted in 2024, this internship focused on further developing and optimizing the SFDFM, through two complementary approaches : statistical analysis of the existing score and consolidation of its digital tool.

The main goal was to strengthen the clinical relevance of the SFDFM using already available data, while laying the groundwork for its broader digital deployment. Two main directions guided this work :

- **Statistical analysis of the score** : the aim was to identify the most influential items in the overall score calculation, assess their discriminative power, and, based on the findings, propose an optimized version of the questionnaire — more concise, yet equally effective.
- **Reevaluation of classification thresholds** : the current thresholds defining frailty levels (mild, moderate, severe) were set empirically. The objective was to refine them using statistical methods, notably by comparing them with levels derived from the Fried criteria.

To conduct this analysis, a dataset from a 2024 study led by **Aude Le Ménez** was used. It included 89 patients aged over 65, monitored in private physiotherapy practices. The data had been collected using the paper version of the SFDFM, and then manually entered into an Excel spreadsheet. Although the web application was improved

during the internship, no new digital data collection was performed due to the absence of an active screening campaign during the internship period. Nevertheless, the tool is ready for use in future experiments.

This work was based on three main hypotheses :

1. **Not all items have equal weight** : certain items — particularly functional tests such as TUG, single-leg stance, or grip strength — are likely more discriminative than others. Identifying these items may justify weighting or simplifying the score.
2. **The current classification thresholds can be improved** : the existing thresholds may not fully reflect clinical reality. A statistically redefined version, aligned with Fried's levels, could enhance the score's consistency.
3. **The tool can be optimized without compromising clinical value** : a shorter version of the SFDFM, focused on the most relevant items, could retain its sensitivity while being easier to use in clinical settings.

1.7 Justification of the Methods Used

To meet the objectives of this internship — namely the identification of the most relevant items in the SFDFM and the re-evaluation of its classification thresholds — it was necessary to use robust statistical methods, tailored to the nature of the available data (heterogeneous variables, small sample size, ordinal classification).

Choice of XGBoost for Classification Among the models tested, XGBoost (*Extreme Gradient Boosting*) stood out for its performance, particularly in classifying frailty levels based on the SFDFM. It offers several advantages that are well documented in the literature : strong performance on small datasets, the ability to handle heterogeneous variables (both categorical and continuous), and enhanced interpretability through SHAP values. This type of model has already been applied in various clinical contexts to extract relevant decision thresholds — notably in endocrinology [8], infectious diseases [14], and cardiovascular surgery [9] — reinforcing its relevance in our setting.

Leave-One-Out Cross-Validation (LOO-CV) Given the relatively small sample size (n = 89), we opted for Leave-One-Out Cross-Validation (LOO-CV) to estimate the predictive performance of our models. In this method, each observation is used once as the test set, while the model is trained on the remaining $n - 1$ observations. This ensures maximal use of the available data and minimizes the risk of overfitting. LOO-CV is particularly recommended in clinical research with limited data or rare events [4]. It is also widely cited in the scientific literature as a preferred approach for obtaining robust performance estimates — such as the Brier score [4] or the c-statistic [4] — especially in logistic models or clinical biometrics, even though it may introduce slight negative bias for the c-statistic [4]. In this context, LOO-CV offers a relevant compromise between minimal bias and full data utilization.

2 Materials and Methods

This section outlines the four methodological components that structured the work carried out during this internship.

The first component focuses on the redesign and deployment of a robust digital interface enabling standardized data collection through the SFDFM questionnaire.

The second component involves the statistical analysis of the most influential questionnaire variables, with the goal of exploring possible simplification or prioritization of items.

The third component aims to identify classification thresholds within the SFDFM score that can reproduce the clinical categorization used in the Fried score.

Finally, the fourth component examines the comparability between patient groups based on their evaluation context (private practice vs hospital) in order to assess the external robustness of the SFDFM.

2.1 Improvement of the SFDFM Digital Interface

To enable smooth, secure, and large-scale use of the SFDFM questionnaire by healthcare professionals, a complete modernization of the digital interface was deemed essential.

Building on the digitization work initiated by Dorian Geraldes Pereira in 2024, a full redesign of the SFDFM application was carried out during this internship. The goal was to deliver a more robust, user-friendly, and practical tool for real-world clinical use.

Technologies used The interface was developed using `Dash` [12] (a Python framework based on `Flask` and `Plotly`) and deployed on the cloud platform `Heroku`. Secure data storage was managed through the `Girder` API with key-based authentication.

Implemented ergonomic improvements Several improvements were made to enhance the overall user experience : the questionnaire was structured into successive steps (*step-form*) to guide data entry, explicit navigation buttons (*Next*, *Submit*) with confirmation messages were added, automatic form reset was implemented after each submission, and help features such as tooltips and visual animations were integrated. A graphic redesign, including a custom background, also contributed to a more professional look and feel.

Data storage optimization The previous version generated one Excel file per submission, resulting in a fragmented database. A new **automatic merge** function was implemented : each new submission is dynamically appended to a centralized file named `DATA.xlsx`, securely stored on the server.

Each row in the file corresponds to a unique patient (identified via a pseudonymized ID), and each column represents one item from the SFDFM questionnaire.

Addition of a secure consultation interface A consultation interface was created to allow dynamic viewing of available files, direct download of data through the application (without manual server access), and a clear functional separation between the questionnaire page and the administrative access page.

Final deployment of the application The entire platform (questionnaire + consultation) was deployed on a remote server using the `Heroku` hosting service [5]. This web version allows real-world use, especially in physiotherapy clinics or healthcare centers, and will be further detailed in the *Results* section.

2.2 Analysis of Influential Variables in the SFDFM Questionnaire

As part of the optimization process of the SFDFM questionnaire, it was necessary to analyze the most influential variables on the overall score, in order to identify the most informative items and consider simplifying the tool.

Objective of the analysis One of the main goals of this work was to assess the determinants of the SFDFM functional score, with the aim of identifying the most explanatory variables among the questionnaire items. This step sought to better understand each item's specific contribution to the total score while detecting potential redundancies. Ultimately, this could support the development of a more concise version of the questionnaire without compromising its clinical value.

Data used The analysis was based on a validated dataset titled `données vérifiées 300625.xlsx`, containing responses from 89 patients over the age of 65 followed in private physiotherapy. Each patient was identified by a unique anonymous code, and each column corresponded to an SFDFM item covering clinical, functional, and social domains. To ensure the quality of the statistical analysis, only complete rows without missing values were retained, yielding a final sample of 65 patients.

Data preprocessing Preprocessing was carried out in a Jupyter Notebook file (`influential_variables.ipynb`). Numeric fields were standardized by replacing commas with periods and converted to float type. Binary variables (e.g., Yes/No, Male/Female) were encoded as 0/1. Missing responses were kept as `NaN` but excluded from the final analysis. Ordinal variables, such as orthopedic exams, were encoded according to their scale. Physical test values retained their original units : seconds for time, kilograms for grip strength, and meters per second for gait speed.

The final table included 14 raw explanatory variables, along with the total SFDFM score (/32) used as the target variable.

Statistical methodology A **multivariate linear regression** was used, a method suited for predicting a continuous variable (the SFDFM score) from multiple predictors. This approach estimates each variable's specific effect while controlling for potential confounding. The model takes the following form :

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

where $\hat{Y}$ is the predicted functional score, X_i the explanatory variables (tests, interview data, social characteristics), and β_i the estimated coefficients.

Leave-One-Out Cross-Validation (LOO) Given the small sample size (65 observations), a rigorous cross-validation strategy was applied : **Leave-One-Out (LOO)**. This method involves removing one patient at a time, training the model on the remaining data, and predicting the removed patient's score. The process is repeated for each individual. This approach avoids overfitting and ensures maximum use of the available data. Each observation is used once for testing and 64 times for training.

Evaluation metrics Model performance was assessed using two standard indicators. The coefficient of determination (R^2) measures the proportion of variance in the SFDFM score explained by all variables : the closer to 1, the better the fit. The mean squared error (MSE) quantifies the average difference between actual and predicted scores : lower values indicate better predictive accuracy.

Exclusion-based importance analysis An exclusion-based importance analysis was conducted for each variable : the performance of the full model was compared to that of a model with one variable omitted. The increase in predictive error (ΔMSE) indicated the contribution of each item to the model. This method allows for an objective and reproducible ranking of the most influential variables in building the score.

Group-specific importance analysis by Fried classification In addition to the global model, a secondary analysis was performed by splitting the data according to Fried's clinical classification : *Robust*, *Pre-frail*, and *Frail*. The objective was to examine whether the explanatory variables for the SFDFM score remained consistent across different levels of patient frailty.

After cleaning, the available subset included 61 patients distributed as follows : 36 *Robust*, 20 *Pre-frail*, and 5 *Frail*. For each group, an independent linear regression was performed, always using the SFDFM score as the target variable, but only within that specific subgroup.

Additionally, a second round of modeling was conducted using only six variables identified as the most influential in the global model. The goal was to assess whether these variables maintained sufficient explanatory power when applied independently to each Fried subgroup, particularly in small-sample classes.

Variable importance was assessed using the same exclusion method as for the global model : the impact on MSE was measured for each removed variable. This approach helps compare the predictor hierarchy across clinical profiles and is particularly useful for exploring subgroup-specific effects, such as among highly frail patients where certain tests might carry different weights.

2.3 Determining Classification Thresholds for the SFDFM Score

To ensure the SFDFM score could be used as a true decision-making tool, it was essential to define interpretable thresholds that would link it to an established clinical classification. This step aimed to ensure functional alignment with the frailty categories defined by the Fried score.

Objective The goal of this analysis was to determine the optimal thresholds for the SFDFM score (out of 32 points) to classify patients into three groups : *robust*, *pre-frail*, and *frail*. The aim was to establish a robust correspondence between SFDFM scores and the clinical categories of the Fried scale, facilitating its integration into routine practice.

Data used The analysis was based on the dataset `FRAKITEST VALIDITE_Marie Teixier.xlsx`, containing cross-assessments of SFDFM and Fried scores for 136 older patients. Each patient had an anonymous identifier, a total SFDFM score (/32), and a Fried score (0 to 5). The data were cleaned, missing values removed, and Fried scores grouped into three clinical categories :

— Fried = 0 ⇒ Robust
— Fried = 1–2 ⇒ Pre-frail
— Fried = 3–5 ⇒ Frail

All analyses described below were performed in the Jupyter file methodes_for_finding_thersolds.ipynb, ensuring reproducible processing.

Main methods compared Four approaches were tested to identify two thresholds t_1 and t_2 on the SFDFM score to classify patients into the three Fried categories. Each method relied on a distinct algorithmic strategy :

Method 1 – Exhaustive Grid Search

This method involved systematically exploring all unique SFDFM scores present in the sample to find optimal thresholds t_1 and t_2. First, all possible thresholds were tested to separate **robust patients** from the rest (pre-frail and frail). Then, the same procedure was applied to distinguish **pre-frail** from **frail** patients, this time excluding the robust group. For each tested threshold, **binary accuracy** was calculated by comparing SFDFM-based predictions to actual Fried classes. Finally, the two thresholds that yielded the highest accuracy were selected as optimal classification cut-offs.

Method 2 – Multinomial Logistic Regression

This method fitted a multinomial logistic regression model to predict the probabilities of belonging to each Fried category (Robust, Pre-frail, Frail) from the SFDFM score. The target variable was encoded numerically (0 = Robust, 1 = Pre-frail, 2 = Frail). For each possible SFDFM score, the model computed the probabilities P_k of belonging to each class k. The intersection points between the probability curves defined the two thresholds : t_1 between Robust and Pre-frail, and t_2 between Pre-frail and Frail. Each patient was then classified according to the class with the highest predicted probability.

Method 3 – Decision Tree

This method used a decision tree (`DecisionTreeClassifier`) with the SFDFM score as the only explanatory variable. The tree depth was limited to 2 for readability. The model :

— Automatically identified two optimal splits in the SFDFM score to separate the three classes.

— Generated a direct classification rule, for example :

$$\text{SFDFM} \leq t1 \Rightarrow \text{Robust}, \quad t1 < \text{SFDFM} \leq t2 \Rightarrow \text{Pre-frail}, \quad \text{SFDFM} > t2 \Rightarrow \text{Frail}$$

— These rules were then applied to the full dataset.

Method 4 – Boosted Trees (XGBoost)

XGBoost uses a series of decision trees, where each new tree is trained to correct the errors of previous ones. This boosting technique builds a strong model from many weak classifiers. In this analysis, XGBoost was trained using the SFDFM score as the only explanatory variable.

To extract thresholds from the model, all internal split points from the 50 generated trees were extracted using the `get_dump()` function. All possible threshold pairs (t_1, t_2) were tested to classify patients into three groups. For each pair, the global classification accuracy against Fried classes was calculated. The pair maximizing this performance was selected as the optimal threshold set.

Exploratory methods not retained Two additional approaches were explored but excluded from the final analysis as they do not provide explicit thresholds :

— **SVM (Support Vector Machines)** : the class probabilities are derived from implicit decision functions, not simple threshold values.

— **GAM (Generalized Additive Models)** : although informative, GAMs produce smooth probability curves without clearly defined cut-off points or optimization metrics.

Evaluation criteria To objectively compare the performance of the four main methods, several standard metrics were computed :

— **Overall accuracy** across the three classes (Robust, Pre-frail, Frail),

— **Binary accuracy** for each threshold (Robust vs others, Pre-frail vs Frail),

— **Confusion matrix** comparing predicted and actual Fried classes,

— **Sensitivity (Se)**, **Specificity (Sp)**, and **Youden's index** for each group.

2.4 Comparability Analysis Between Patient Groups

Since the study involved data collected from two different contexts—outpatient physiotherapy clinics on one hand, and geriatric hospital settings on the other—it was essential to ensure that these two populations were sufficiently homogeneous to be analyzed together. This verification is crucial to validate the global analyses based on the SFDFM score.

Objective The aim of this analysis was to assess the statistical comparability between two distinct patient groups who completed the SFDFM questionnaire : those from earlier data (collected in private physiotherapy practices) and those recruited in hospital settings. The goal was to determine whether pooling the two samples was justified in the overall evaluation of the score.

Group definition Two sub-populations were defined based on patient identifiers :
— **Group 1 – Former patients** : assessed in outpatient clinics or at home (n = 89) ;
— **Group 2 – New patients** : assessed in a geriatric hospital (n = 47).
The identifiers of the 47 new hospital patients were explicitly listed, and a new `group` column was added to the main dataset to allow automatic separation.

Compared variables Two scores were compared between the groups :
— The **SFDFM score (/32)**, representing the overall functional score ;
— The **Fried score (/5)**, used as an external clinical reference.

Preprocessing and tools used The analyses were conducted in a Python notebook (`comparability_study.ipynb`) using the source file `FRAKITEST_SEUIL.xlsx`. Preprocessing steps included cleaning and renaming columns based on a standardized naming scheme, automatically assigning group labels based on patient IDs, filtering out rows with missing values, and converting score variables into a usable numeric format.

Statistical methods Each score was analyzed in two steps : a normality test (Shapiro-Wilk), followed—if normality was rejected—by a non-parametric comparison between groups (Mann-Whitney U test).

Significance thresholds The decision thresholds used for statistical interpretation were :
— $\alpha = 0.05$: standard significance threshold ;
— p-values below this threshold were considered indicative of significant differences.

Note on distributions The distributions of the SFDFM and Fried scores were examined separately within each group. When normality was rejected, the Mann-Whitney U test was applied as a robust alternative to the Student's t-test.

3 Results

3.1 Improvement of the SFDFM Digital Interface

The final version of the SFDFM interface developed during this internship consists of two complementary modules : (1) an interactive questionnaire for healthcare professionals, and (2) a secure interface for data consultation.

1. Enhanced Interactive Questionnaire The input interface has been completely redesigned in the form of a multi-step form (*step-form*), with progressive navigation (buttons *Next*, *Finish*) and automatic reset after submission. Transition animations, a custom visual background, and interactive help buttons have been integrated to guide users during data entry.

Additionally, several questions in the SFDFM questionnaire were reformulated to improve clarity and clinical relevance. This helps healthcare professionals better understand the items and reduces entry errors.

Finally, a system for generating a unique identifier was implemented for each patient upon submission, ensuring complete data anonymity.

Figure 2 shows the new ergonomic version of the questionnaire, accessible via a web interface.

FIGURE 2 – New SFDFM questionnaire interface : step-by-step navigation and improved ergonomics

2. Real-Time Data Centralization Submitted data are now automatically merged into a single file named `DATA.xlsx`, updated in real-time on the server. Each row represents a uniquely pseudonymized patient, and each column corresponds to one item of the SFDFM questionnaire (see Figure 3).

Age	Poids	ds il y a 6 n	Genre	ation famil	mbre de ch	ur de chute	te dans les	emoire ou	pour la dé	que des m	mbre inferi	p and go T	le chaise 5	che sur 4	relevé du sc	Grip test	le chiffre 1	de chiffre 1	de
66	84	81	Femme	En couple,	0	0	Non	Oui	Non	Normal	30	34	32	0	61	26	Réussi	Réussi	Ré
66	84	81	Femme	En couple,	0	0	Non	Oui	Non	Normal	42	23	34	0,7	45	27	Réussi	Réussi	Ré
68	84	81	Femme	En couple,	0	0	Non	Oui	Non	Normal	42	23	34	0,7	45	27	Réussi	Réussi	Ré
88	79	79	Homme	Seul(e)	1	4	Oui	Oui	Non	Une limita	15	45	55	0	0	26	Réussi	Réussi	Ré
88	79	79	Homme	Seul(e)	1	4	Oui	Oui	Non	Une limita	15	45	55	0	0	26	Réussi	Réussi	Ré
90	75	73	Homme	En couple,	1	4	Oui	Oui	Non	Une limita	15	45	55	0	0	26	Réussi	Réussi	Ré
79	83	83	Femme	En couple,	0	5	Oui	Oui	Non	Une limita	34	56	52	0,9	85	28	Réussi	Réussi	Ré
87	67	66	Femme	Seul(e)	1	6	Oui	Oui	Oui	Normal	11	45	63	0,12	0	27	Réussi	Réussi	Ré
99	76	77	Femme	En couple,	2	8	Oui	Oui	Non	Une limita	76	87	77	0,98	0	22	Réussi	Réussi	Ra
99	76	75	Femme	En couple,	2	8	Oui	Oui	Non	Plusieurs a	13	76	78	0,45	0	23	Réussi	Réussi	Ra
100	76	75	Homme	Seul(e)	2	8	Oui	Oui	Non	Plusieurs a	13	76	78	0,9	0	23	Réussi	Réussi	Ra
74	68	70	Femme	Seul(e)	2	6	Oui	Oui	Oui	Une limita	4	22	19	0,72	0	16	Réussi	Réussi	Ré
68	75	74	Homme	En couple,	0	2	Non	Non	Non	Normal	8	11	11	0,95	1	38	Réussi	Réussi	Ré
84	60	64	Femme	Seul(e)	3	8	Oui	Oui	Oui	Plusieurs a	2	26	30	0,55	0	12	Réussi	Réussi	Ré

FIGURE 3 – Excerpt from centralized file `DATA.xlsx` : one row per patient, one column per item

This storage approach avoids the creation of multiple isolated files, as seen in the initial prototype, and ensures better traceability and simplified exploitation of clinical data.

3. Secure Data Consultation Interface An additional interface allows authorized users to access the collected data. This page (Figure 4) requires entering a secure API key and provides access to view and download stored files from the cloud platform via the `Girder` API.

FIGURE 4 – Secure data consultation interface with API key access

4. Web Deployment The entire application (questionnaire form + data consultation) was deployed on a remote server using the **Heroku** cloud hosting platform. The website is publicly accessible at the following address : `https://depistage-fragilite-motrice-1363f3377112.herokuapp.com/`

Two modules are available from the homepage. The first is the SFDFM input form, structured into successive steps, with submission requiring a valid **authentication key**. This ensures that only authorized users (e.g., supervising clinicians or researchers) can upload data to the secure `Girder` platform. The second module is the data consultation interface, also protected by a key, enabling real-time file listing and access to the `DATA.xlsx` master file containing all submissions to date.

Beyond technical improvements, these changes have a direct impact on clinical practice. They make the SFDFM a more operational tool in real-life settings. The restructured step-by-step interface and interactive elements facilitate completion by healthcare professionals, even in time-constrained environments. Tooltips and visual aids provide useful contextual help, supporting user comprehension and execution of the tests. The help button also enhances autonomy and reduces the likelihood of input errors.

Furthermore, the automatic form reset after submission simplifies the logistics of repeated assessments : a new evaluation can begin immediately without reloading the page or restarting the application, saving time in successive consultations.

Lastly, the creation of the centralized and automatically updated `DATA.xlsx` file provides a unified and directly usable database. This avoids file fragmentation and enables more efficient data analysis. The systematic use of pseudonymized patient identifiers also opens the door to longitudinal monitoring, supporting future studies on the evolution of frailty over time.

The secure remote deployment and dual-interface design (input + consultation) make the tool easily shareable while meeting confidentiality requirements. Altogether, these improvements reinforce the usability and clinical relevance of SFDFM as a large-scale digital screening tool.

3.2 Analysis of Influential Variables in the SFDFM Questionnaire

Final Dataset Used

After all the data preparation steps described previously (cleaning, numerical format conversion, binary encoding, handling missing values), a consistent subset of **65 patients** was retained. Each had complete data for **14 raw explanatory variables** from the SFDFM questionnaire.

These variables cover clinical, functional, and social aspects. The full dataset served as the input for the regression model.

Figure 5 illustrates the structure of the cleaned dataset used for the analysis : each row represents a patient, and each column corresponds to a variable from the SFDFM questionnaire.

	Genre	Situation familiale	Chutes	Peur de chuter	Douleur chronique	Trouble mémoire	Médicaments	Examen orthopédique	Appui unipodal (s)	TUG (s)	Lever de chaise (s)	Vitesse marche (m/s)	Relevé sol	Grip test (kg)	Score fonctionnel
0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	33.00	8.86	8.37	1.300000	0.0	39.6	1.0
1	1.0	0.0	0.0	0.0	1.0	0.0	0.0	2.0	24.96	12.54	13.20	0.790000	0.0	14.9	9.0
2	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	10.32	6.27	6.60	1.170000	0.0	21.0	3.0
3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	22.07	8.57	9.69	1.200000	0.0	19.4	3.0
4	1.0	1.0	0.0	0.0	1.0	0.0	1.0	2.0	3.04	12.03	12.03	0.830000	0.0	22.3	10.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
60	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	8.30	9.48	11.55	1.470000	0.0	31.2	4.0
61	1.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	30.00	8.06	10.23	1.403509	0.0	31.2	5.0
62	0.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	2.00	14.87	26.62	0.940000	0.0	32.3	13.0
63	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	9.29	9.81	10.50	1.150000	0.0	50.4	4.0
64	1.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	3.30	10.38	11.23	1.210000	0.0	20.9	10.0

65 rows × 15 columns

FIGURE 5 – Final dataset after cleaning and filtering (65 patients, 14 variables + functional score)

Overall Model Performance

The chosen model for analyzing variable influence was a **multivariate linear regression**, applied to a subset of **65 patients** with complete data for the **14 SFDFM items**.

Validation using the **Leave-One-Out (LOO)** method yielded remarkable predictive performance, with a **coefficient of determination** $R^2 = 0.8185$. This means that over **81% of the variability in the SFDFM score** is explained by these 14 variables. Additionally, the **mean squared error (MSE)** was relatively low (5.09), corresponding to an average error of approximately **2.26 points out of 32**, indicating **satisfactory clinical accuracy**.

Variable Importance Analysis

An **exclusion-based analysis** was conducted to evaluate the individual contribution of each variable. At each iteration, one variable was removed from the model, and the resulting MSE was compared to that of the full model.

Table 2 presents the MSE obtained after removing each variable, the corresponding increase in prediction error (ΔMSE), and a qualitative ranking of importance.

Removed Variable	**MSE without variable**	**Impact** (ΔMSE)	**Importance level**
Orthopedic exam	10.95	+5.86	High
Chair rise (s)	8.39	+3.30	High
Ground rise	7.71	+2.61	High
Family situation	7.36	+2.26	High
Unipedal stance (s)	6.98	+1.89	High
Memory issue	6.88	+1.79	High
Fear of falling	6.57	+1.47	High
Falls	6.43	+1.34	High
Grip test (kg)	6.23	+1.14	Moderate
Walking speed (m/s)	5.76	+0.66	Moderate
Medications	5.13	+0.03	Low
Sex	5.02	–0.07	Low
Chronic pain	4.55	–0.55	Low
TUG (s)	2.75	–2.34	Low

TABLE 2 – Impact of removing each variable on model MSE (raw values)

To complement this tabular view, Figure 6 shows a visual representation of the ΔMSE for each variable, helping identify the most influential predictors.

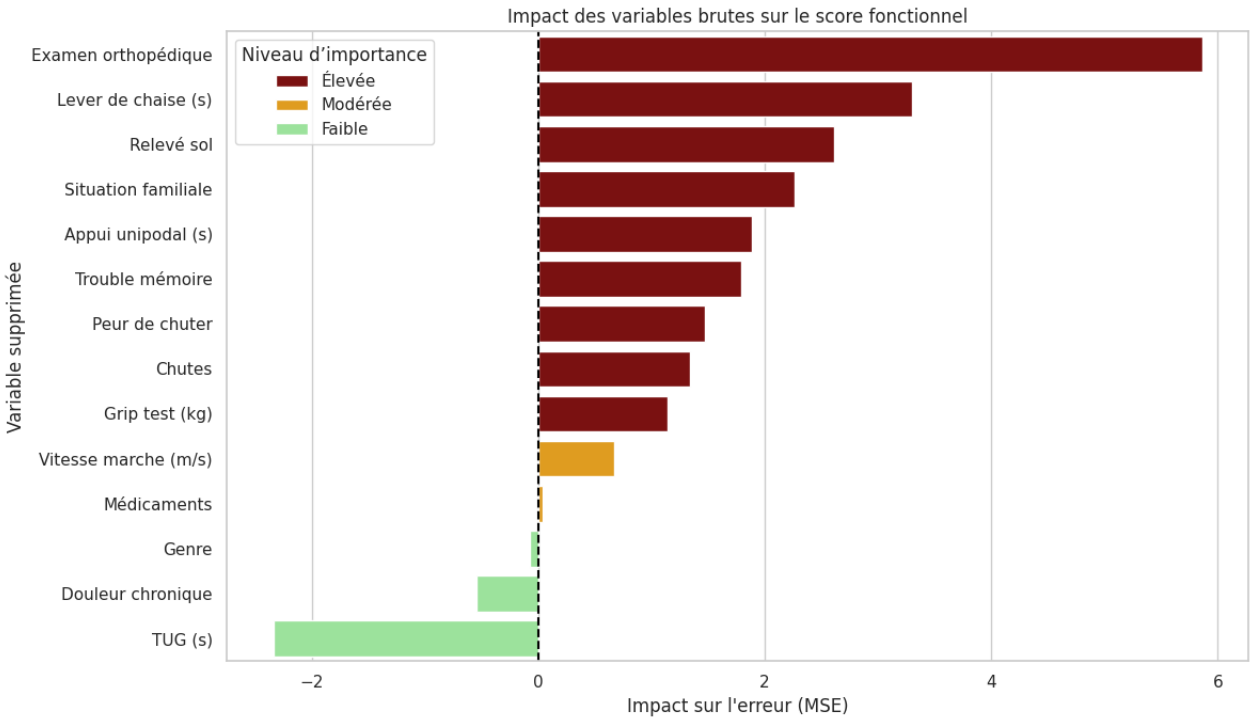

FIGURE 6 – Effect of removing each variable on prediction error (MSE)

Interpretation

Variables whose removal causes a strong increase in error (high ΔMSE) are considered highly influential. This is the case for the orthopedic exam, which emerged as the top predictor of the score. Other key contributors include functional tests such as chair rise, ground rise, and unipedal stance, all of which significantly affect prediction accuracy.

Additional influential factors include social and subjective aspects like family situation, memory troubles, fear of falling, and fall history, confirming the multidimensional nature of the SFDFM score.

Conversely, some variables were found to have only moderate impact (e.g., walking speed, grip strength), while others had minimal or even negative effects (e.g., sex, chronic pain, medication, TUG), suggesting they may be less relevant in predicting the total score.

Analysis by Fried Subgroups

To gain deeper insights into the mechanisms behind the SFDFM score, an additional analysis was conducted by dividing the population into three clinical groups according to Fried's classification : **Non-frail**, **Pre-frail**, and **Frail**.

Separate linear regression models were trained for each subgroup, with the same exclusion method applied to assess the relative importance of variables within each level of frailty.

Table 3 summarizes the top influential variables and model performance for each subgroup.

Fried Class	Most Influential Variables	R^2	MSE
Non-frail	Orthopedic exam, Unipedal stance, Falls	0.6811	1.92
Pre-frail	Fear of falling, Falls, Memory issue	0.1668	10.71
Frail	Sex, Chronic pain, Ground rise	–19.37	1178.31

TABLE 3 – Most influential variables per Fried subgroup (14-variable models)

Some variables such as **orthopedic exam** and **memory issues** remained influential across multiple groups,

suggesting cross-group robustness. Others were highly predictive only within specific profiles — for instance, **fear of falling** among pre-frail patients, or **sex** and **chronic pain** in the frail group.

However, caution is warranted in interpreting these results, especially in the **frail** group, where model performance was highly unstable (strongly negative R^2, very high MSE). This reflects the small sample size for this group ($n = 5$), leading to large variability and low statistical reliability.

To test whether a simplified model would still perform adequately, new models were trained using only the 6 most influential variables from the global model : *Orthopedic exam*, *Chair rise*, *Ground rise*, *Family situation*, *Unipedal stance*, and *Memory issue*.

Table 4 presents a comparison of full (14-variable) versus reduced (6-variable) models for each Fried subgroup.

Fried Class	**Model**	**Variables**	**R^2**	**MSE**	**Improvement**
Non-frail	Full	14	0.6811	1.92	–
	Reduced	6	0.4343	3.41	↓ performance
Pre-frail	Full	14	0.1668	10.71	–
	Reduced	6	0.2711	9.36	↑ performance
Frail	Full	14	–19.37	1178.31	–
	Reduced	6	0.3961	34.93	↑ performance

TABLE 4 – Performance comparison by Fried subgroup – full vs reduced (6-variable) models

These results show that the reduced model significantly improved predictive quality for the more unstable subgroups, particularly for **frail** patients, where R^2 became positive and MSE dropped sharply (from 1178 to 34.9). An improvement was also observed in the **pre-frail** group.

However, among **non-frail** patients, the full model still performed better, indicating that the original model's complexity is more useful for stable clinical profiles.

In summary, these results confirm the SFDFM's relevance as an integrative functional score, capable of capturing much of the clinical variability observed in older adults.

From a statistical standpoint, the multivariate regression achieved a high explanatory power ($R^2 = 0.8185$), showing that the 14 variables effectively capture individual functional differences — a strong indicator of the questionnaire's internal validity.

Clinically, the importance analysis highlighted key items (e.g., orthopedic exam, ground rise, fear of falling) whose exclusion significantly reduces model accuracy. In contrast, some items had minimal or no impact, suggesting potential for targeted simplification without performance loss.

The Fried subgroup analysis revealed that predictor importance shifts with frailty level. Variables such as sex and chronic pain became more relevant for frail individuals, unlike in the overall model.

This divergence suggests the SFDFM's explanatory structure is not stable across Fried categories — a noteworthy concern since Fried's classification is a well-established clinical reference. It raises questions about SFDFM's sensitivity to certain aspects of frailty.

This led to testing a reduced model using only the six most influential global variables. This simplification improved predictions substantially in the most unstable groups — especially the **frail** (R^2 improved from –19.37 to 0.3961 ; MSE dropped by 97%). A performance gain was also observed for **pre-frail** patients.

These findings suggest that a targeted variable subset may improve model robustness in vulnerable profiles. The relevance of adopting a reduced or adaptive model will be further discussed in the *Discussion* section.

3.3 Threshold Identification for SFDFM Classification

All threshold analyses presented in this section were conducted using the same cleaned dataset, which includes **136 older patients**. Each patient is associated with a global SFDFM functional score (ranging from 0 to 32) and a reference Fried score (ranging from 0 to 5). For classification purposes, Fried scores were grouped into three clinically coherent categories : *Non-frail* (Fried = 0), *Pre-frail* (Fried = 1–2), and *Frail* (Fried = 3–5).

After reclassification, the dataset distribution was as follows : 58 patients were classified as Non-frail, 63 as Pre-frail, and 15 as Frail. This imbalance in group sizes—especially the low number of Frail patients—should be

considered when interpreting the model performance.

A graphical analysis was conducted to better understand how SFDFM scores are distributed across Fried categories. Two types of plots were generated :

On one hand, a **violin plot** visualizes the density of SFDFM scores within each Fried group, highlighting medians, interquartile ranges, and any asymmetry. On the other hand, a **swarm plot** displays the individual position of each patient according to their SFDFM score, enabling the visualization of overlap or separation zones between groups.

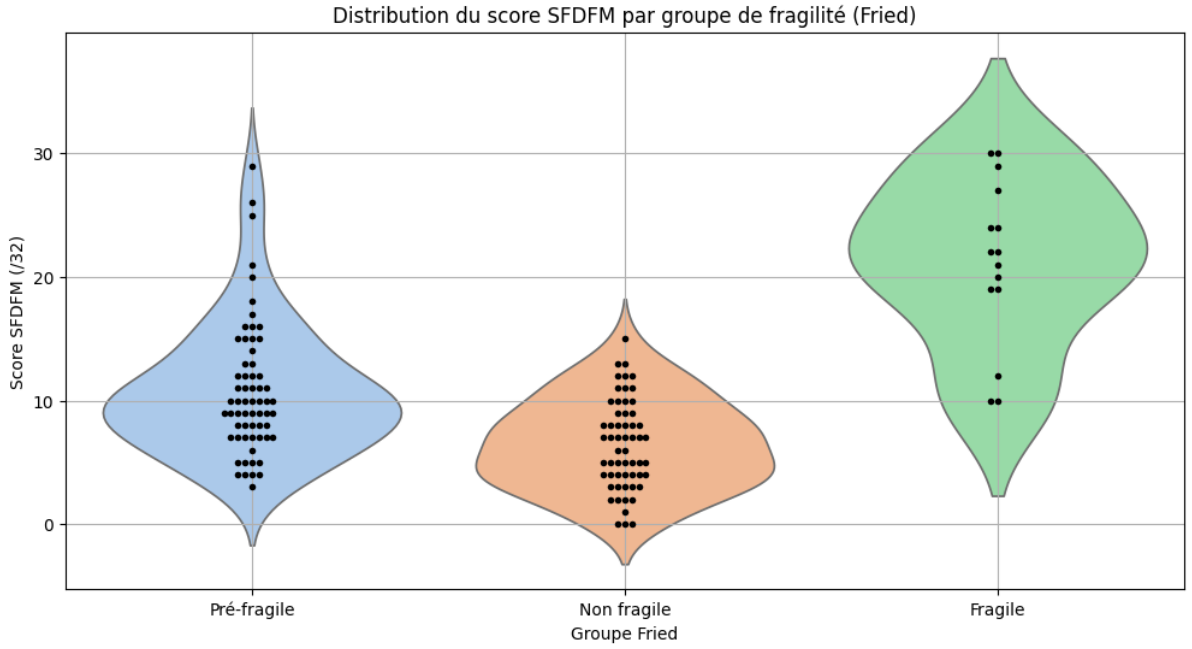

FIGURE 7 – Distribution of SFDFM scores by Fried group — visualized using *violin plot* and *swarm plot*

Figure 7 shows a progressive structure of SFDFM scores according to frailty level. Non-frail patients tend to score low, typically below or equal to 10. Pre-frail profiles show greater variability, generally between 9 and 18. Frail patients have significantly higher scores, often above 20, with a peak density around 25 to 30.

These visual patterns suggest the presence of natural transition zones on the SFDFM scale, supporting the idea of identifying two optimal thresholds (t_1 and t_2) to objectively discriminate the three clinical groups.

Method 1 – Exhaustive Grid Search

This first approach involves exhaustively testing all possible threshold combinations to segment the SFDFM score into three groups. The idea is to evaluate, for each possible score value (integer from 0 to 32), its ability to act as a binary cut-off. Two successive comparisons are conducted : first, between *Non-frail* patients (Fried $= 0$) and the others ; then, between *Frail* patients (Fried ≥ 3) and the others. For each threshold candidate, binary classification is simulated and the accuracy is computed. The best-performing thresholds are then combined to generate a final trichotomous classification.

Two optimal thresholds were identified : the first, $t_1 = 8$, distinguishes Non-frail patients from others with a classification accuracy of 68.60%. The second, $t_2 = 18$, effectively isolates Frail patients, with a binary accuracy of 89.74%. These thresholds define three distinct functional zones on the SFDFM scale.

The corresponding classification rule is :

$$\text{SFDFM} \leq 8 \Rightarrow \textbf{Non-frail}, \quad 8 < \text{SFDFM} \leq 18 \Rightarrow \textbf{Pre-frail}, \quad \text{SFDFM} > 18 \Rightarrow \textbf{Frail}$$

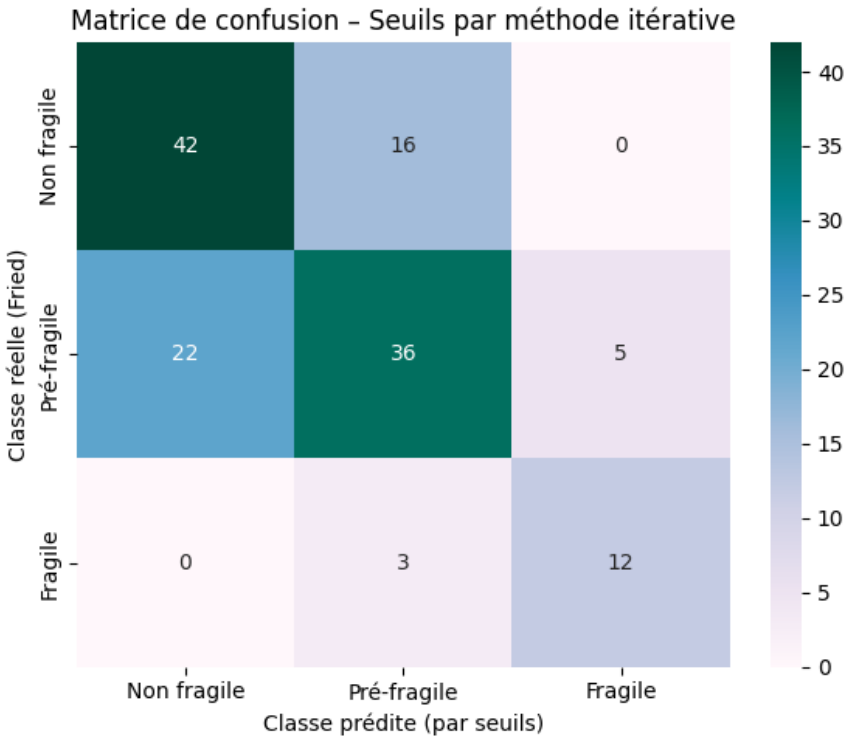

FIGURE 8 – Confusion matrix – Method 1 (thresholds $t_1 = 8$, $t_2 = 18$)

The global accuracy of this rule is **66.18%**, meaning that approximately two out of three patients were correctly classified based on their SFDFM score.

Performance indicators per class were also computed. For Non-frail patients : sensitivity = 0.724, specificity = 0.718, Youden index = 0.442. For the Pre-frail class : sensitivity = 0.571, specificity = 0.740, Youden index = 0.311. For the Frail group : sensitivity = 0.800, specificity = 0.959, Youden index = 0.759.

These results confirm that the exhaustive search method provides simple, effective, and usable thresholds for segmenting the population by functional frailty.

Method 2 – Multinomial Logistic Regression

The second approach uses a multinomial logistic regression model to predict the probability of each Fried class (Non-frail, Pre-frail, Frail) based on the SFDFM score. Unlike a decision tree or explicit threshold search, this method produces continuous probability curves whose intersections suggest transition zones between groups.

The model was trained on all 136 patients using only the SFDFM score as input. Each patient is assigned a triplet of probabilities (P_0, P_1, P_2) for being Non-frail, Pre-frail, or Frail. Figure 9 shows how these probabilities evolve with the SFDFM score.

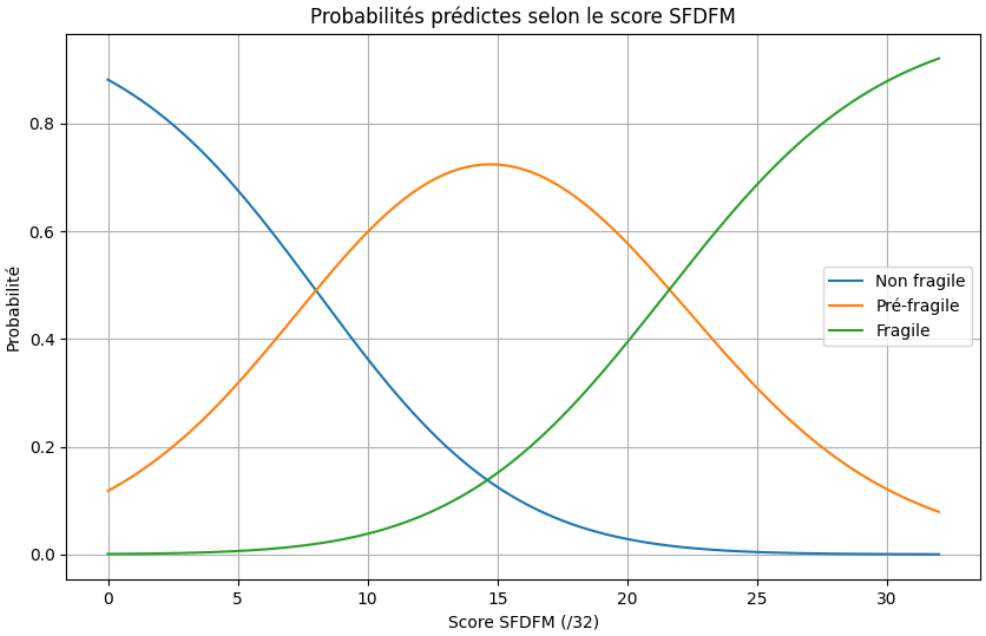

FIGURE 9 – Predicted probabilities from multinomial logistic regression by SFDFM score

The intersection points yield two thresholds : around 7.88 (Non-frail vs. Pre-frail) and 21.55 (Pre-frail vs. Frail). These were used to define three score intervals : ≤ 7.88 for Non-frail, between 7.88 and 21.55 for Pre-frail, and $>$ 21.55 for Frail.

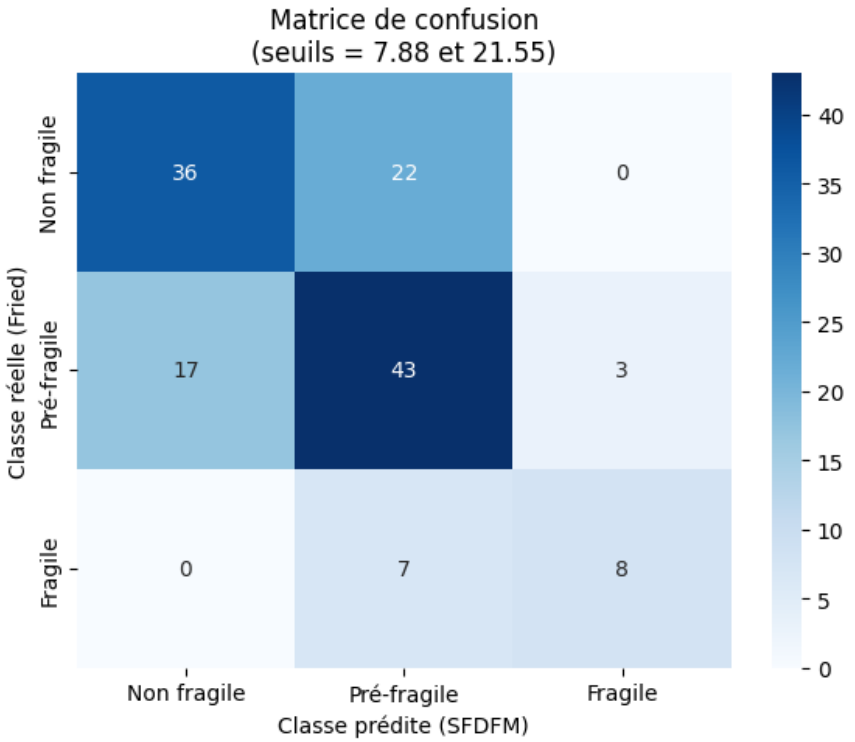

FIGURE 10 – Confusion matrix – Method 2 (thresholds 7.88 and 21.55)

Global accuracy was 63.97%. Class-wise, Non-frail had sensitivity = 0.621, specificity = 0.782, Youden = 0.403. Pre-frail had sensitivity = 0.683, specificity = 0.603, Youden = 0.285. Frail had sensitivity = 0.533, specificity = 0.975, Youden = 0.509.

These results show the method offers a smooth probabilistic interpretation aligned with clinical observations, though slightly less performant than exhaustive grid search, especially for Frail detection.

Method 3 – Decision Tree

This third approach uses a decision tree trained on the SFDFM score to predict the Fried class. The model generates simple conditional rules as explicit thresholds.

Trained on all 136 patients, the model identified two main splits at 6.5 and 18.5 :

$$\text{SFDFM} \leq 6.5 \Rightarrow \text{Non-frail}, \quad 6.5 < \text{SFDFM} \leq 18.5 \Rightarrow \text{Pre-frail}, \quad \text{SFDFM} > 18.5 \Rightarrow \text{Frail}$$

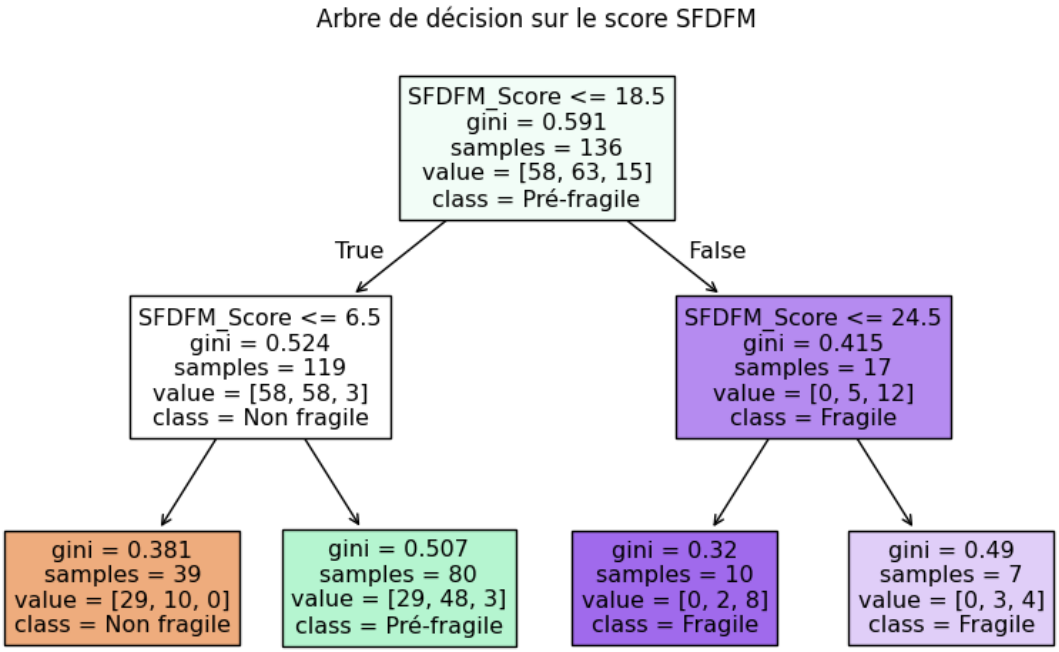

FIGURE 11 – Decision tree based on SFDFM score

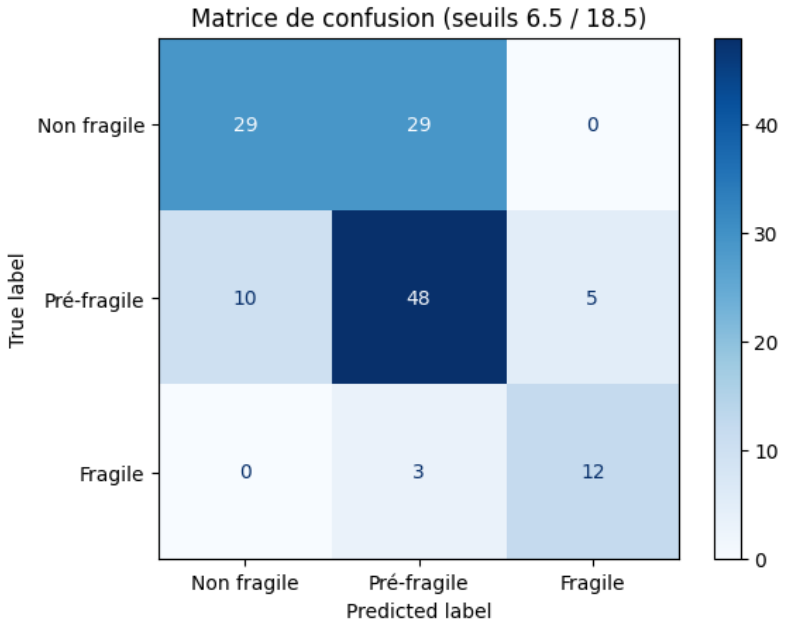

FIGURE 12 – Confusion matrix – Method 3 (Decision Tree)

Global accuracy : 65.44%. Sensitivity and specificity for Frail patients were 0.800 and 0.959 respectively, with a Youden index of 0.759, demonstrating excellent detection of at-risk individuals.

Method 4 – Boosted Trees (XGBoost)

This method uses the XGBoost algorithm with a maximum depth of 3 and 50 trees, trained on the SFDFM score only. The best thresholds found were $t_1 = 8$ and $t_2 = 18$, resulting in :

$$\text{SFDFM} \leq 8 \Rightarrow \text{Non-frail}, \quad 8 < \text{SFDFM} \leq 18 \Rightarrow \text{Pre-frail}, \quad \text{SFDFM} > 18 \Rightarrow \text{Frail}$$

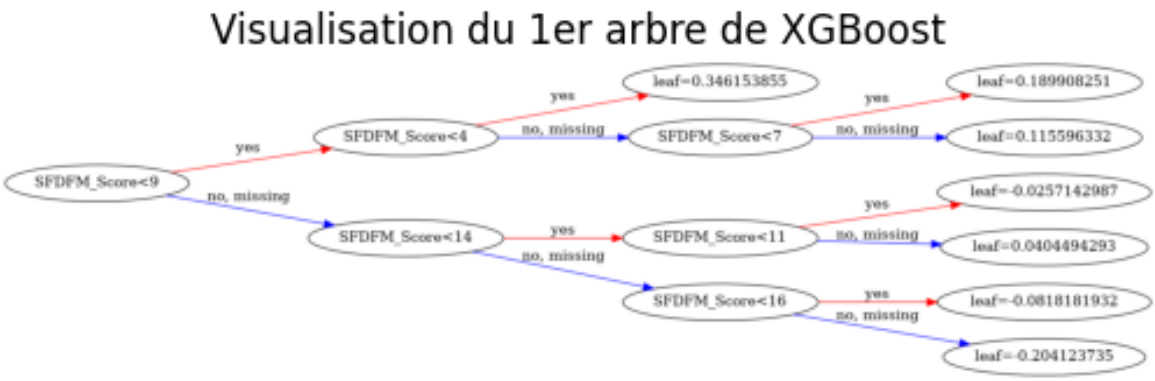

FIGURE 13 – Extract from XGBoost tree (SFDFM → Fried)

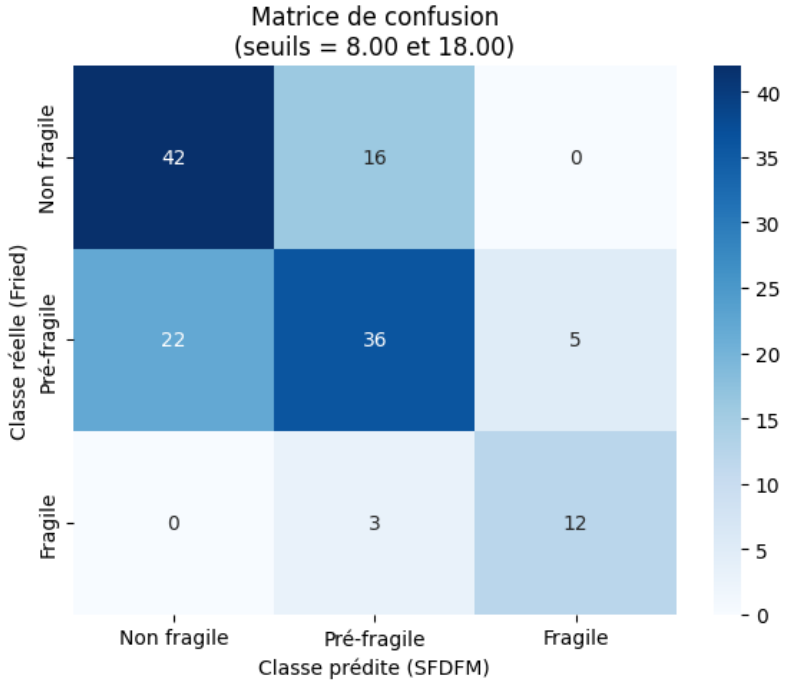

FIGURE 14 – Confusion matrix – Method 4 (XGBoost with thresholds 8 and 18)

Accuracy reached 66.18%, equivalent to Method 1. Sensitivity for Frail patients = 0.800 ; specificity = 0.959 ; Youden index = 0.759.

Summary : All methods—grid search, multinomial regression, decision tree, and XGBoost—yielded interpretable thresholds consistent with Fried categories. The recurrent thresholds of 8 and 18 reinforce their robustness and external validity.

These results confirm the SFDFM's potential as a screening tool. The strong performance in identifying Frail individuals ($Se = 0.800$, $Sp = 0.959$) supports its clinical application for targeted assessment and personalized interventions in geriatric care.

3.4 Comparability Analysis Between Patient Groups

To ensure the validity of analyses based on the SFDFM score, it was necessary to verify that the profiles of patients included in the study were comparable regardless of their assessment context. Two groups were distinguished : former patients assessed in private practices or at home ($n = 89$), and new patients assessed in a hospital setting ($n = 47$). The goal of this comparison was to evaluate whether significant differences existed between these two populations in terms of measured scores, particularly the SFDFM and Fried scores.

For the SFDFM functional score (/32), a Shapiro-Wilk normality test was applied separately to each group. In both cases, the distributions were found to be non-normal ($p < 0.001$), justifying the use of a non-parametric test. The Mann-Whitney U test was subsequently applied and revealed no statistically significant difference between the former and new patients. The obtained U statistic was 2067.0 with a p-value of 0.9123.

Figure 15 illustrates this absence of difference. The distributions appear very similar in terms of both medians and interquartile ranges. This suggests that the SFDFM score remains stable across populations, regardless of data collection setting.

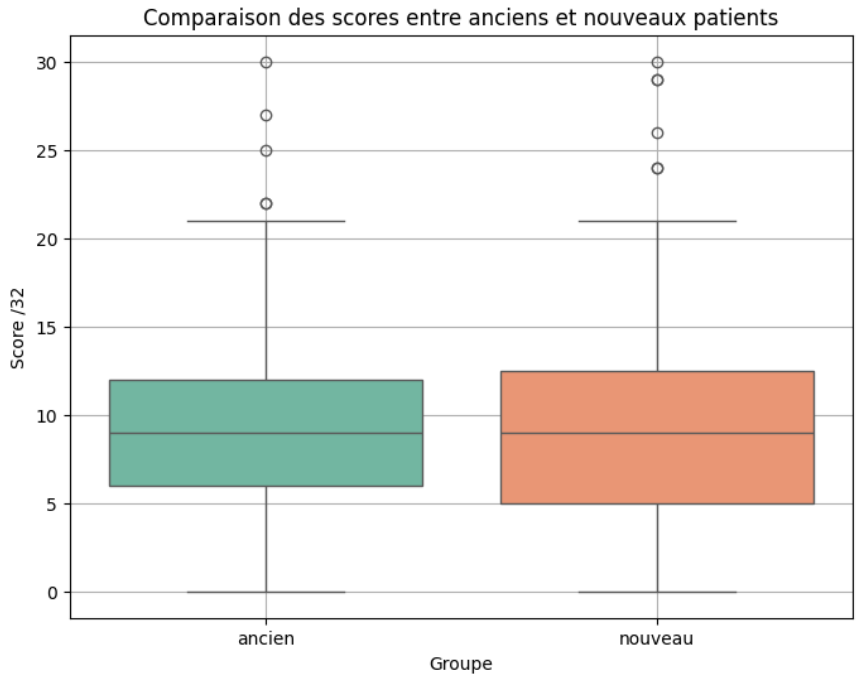

FIGURE 15 – Distribution of SFDFM score (/32) by group (former vs. new patients). No significant difference observed ($p = 0.9123$).

In contrast, analysis of the Fried score (/5) revealed a different outcome. Again, both distributions were non-normal ($p < 0.001$), leading to the use of the Mann-Whitney U test. This time, a statistically significant difference was detected between groups, with a U statistic of 1661.5 and a p-value of 0.0377.

As shown in Figure 16, hospitalized patients exhibit slightly higher Fried scores on average. The median shifts upward, and the score dispersion is greater, indicating a generally higher level of frailty in this group. This result may reflect a selection effect—hospitalized patients may have more vulnerable profiles—or a greater sensitivity of the Fried score to certain clinical aspects (e.g., fatigue, weight loss).

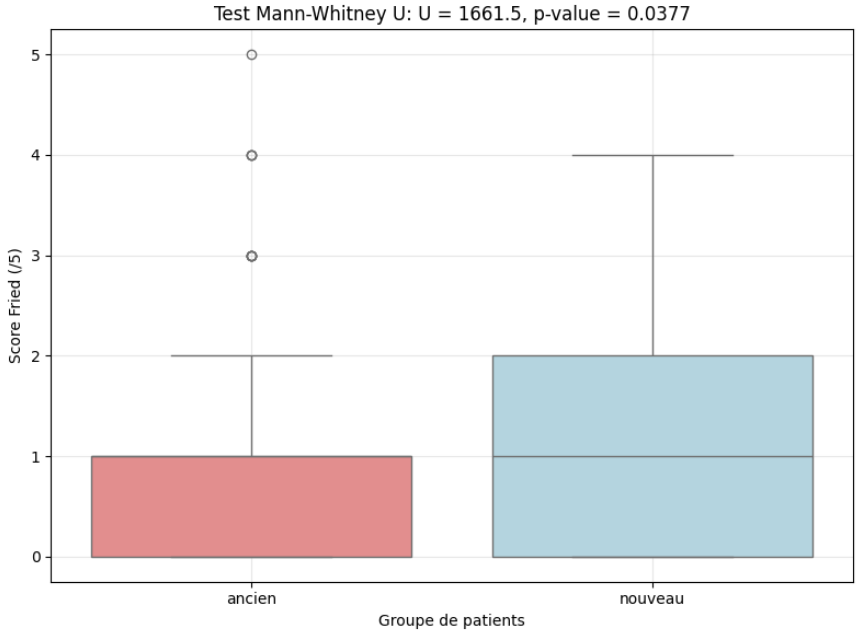

FIGURE 16 – Distribution of Fried score (/5) by group (former vs. new patients). A significant difference is observed ($p = 0.0377$).

The comparative results are summarized in Table 5.

Score	Group	Test Used	p-value
SFDFM (/32)	Former (n = 89) New (n = 47)	Mann-Whitney U	0.9123
Fried (/5)	Former (n = 89) New (n = 47)	Mann-Whitney U	0.0377

TABLE 5 – Comparison of SFDFM and Fried scores between former and new patient groups

In summary, the SFDFM score exhibits stable distribution across the two patient groups, indicating consistent behavior regardless of the assessment context (clinic or hospital). However, this apparent stability should not be interpreted as a strength in itself. Indeed, the Fried score, considered a clinical gold standard, detects a significant difference between groups, suggesting greater sensitivity to contextual or clinical variations.

This discrepancy between the two scores raises questions about the SFDFM's ability to capture subtle aspects of frailty. The fact that it fails to reflect a difference clearly identified by Fried limits its use as a fully equivalent alternative at this stage. These findings call for a critical analysis of its actual sensitivity and for methodological adjustments to strengthen its value as a clinical screening tool.

4 Discussion

Objectives and Hypotheses Revisited The primary objective of this study was to assess the clinical validity and statistical robustness of the SFDFM score, with the aim of developing a reliable, simple screening tool tailored to physiotherapists' practices. The initial hypothesis was that a score integrating functional and contextual dimensions could match, or even rival, the discriminative power of the Fried score, which is considered the international gold standard for frailty assessment.

Main Findings and Interpretations The analyses confirmed **strong internal validity** of the SFDFM, with an R^2 of 0.8185, reflecting solid statistical coherence. However, intergroup comparisons highlighted a **clinical inferiority** of the SFDFM compared to the Fried score : the latter successfully discriminated patients by care setting (hospital vs. home), a distinction not captured by the SFDFM. This discrepancy suggests that key dimensions of frailty—such as fatigue or unintentional weight loss—are underrepresented in the current version of the questionnaire.

Additionally, classification analyses enabled the definition of **robust numeric thresholds** (8 and 18), derived from both grid-based and XGBoost approaches. These thresholds provide a clear basis for clinical interpretation of the score, with moderate overall accuracy (around 66%) but satisfactory performance for identifying the most frail patients ($Se = 0.800$, $Sp = 0.959$).

Explanatory Heterogeneity Across Subgroups Analysis by Fried subgroups revealed a **marked instability** in explanatory variables depending on clinical profile. Among frail patients, the model became non-predictive (strongly negative R^2), due in part to a very small sample size (n = 5) and a high number of variables (14). Moreover, influential variables varied considerably between groups : while motor tests dominated in the global model (orthopedic exam, chair rise...), unexpected variables such as gender or chronic pain emerged among the frail group.

This instability raises concerns about the **internal coherence** of the SFDFM. The fact that predictive parameters are not reproducible across subgroups, while Fried classification remains clinically stable, suggests insufficient sensitivity of the score to certain specific frailty dimensions. This points to the potential relevance of a more adaptable model tailored to individual patient profiles.

The complementary analysis using a reduced model, based on the six most influential variables from the global model, yielded promising results. For the most frail patients, switching to a reduced model significantly improved statistical performance, with a positive R^2 and a drastic drop in MSE. A similar improvement was observed for pre-frail patients.

Conversely, in **non-frail** patients, the complete model maintained better performance, suggesting that a more detailed approach remains preferable for clinically stable cases.

These findings pave the way for an **adaptive version** of the SFDFM, where the patient's level of frailty could guide the use of either a simplified or complete version of the score. This would optimize both screening feasibility and prediction accuracy, while accommodating real-world constraints.

Original Contribution : Bridging Digital and Clinical Practice One of the distinctive contributions of this work lies in the **advanced digital implementation** of the SFDFM. The redesign of the interface, step-by-step structure, pseudonymized identifiers, and automatic centralization of data into a single file represent major improvements. They reinforce the **practical usability** of the SFDFM : the score is no longer merely theoretical or statistical—it becomes a concrete, accessible, standardized, and practical tool.

This synergy between **clinical refinement** (score improvement) and **digital development** (web application) reflects an integrated approach, essential for the large-scale deployment of screening tools. It enables **longitudinal data collection**, real-time automated analysis, and greater adoption by healthcare professionals.

Limitations and Perspectives Several limitations must be noted. The sample size remains modest, particularly in the subgroups (n = 5 for frail patients), which limits the statistical robustness and generalizability of the models. Furthermore, no comparison was made with other validated tools such as the *Short Physical Performance Battery (SPPB)* or the *Clinical Frailty Scale*. Such comparisons would help better position the SFDFM among current screening instruments. Lastly, some essential clinical dimensions for characterizing frailty—such as fatigue or weight loss—are currently absent from the questionnaire, potentially limiting its sensitivity.

The future directions opened by this work are both clinical and technological. Clinically, it would be relevant to refine the questionnaire content to better capture subtle frailty signals, assess threshold reproducibility in larger samples, and consider developing simplified or modular score versions depending on patient profiles. Technologically, the digital interface offers exciting possibilities : longitudinal patient tracking, automated alerts in response to functional changes, and support for clinical research in primary care via centralized and structured data collection.

Concluding Remarks In conclusion, the SFDFM appears to be a promising score at the intersection of clinical screening and digital innovation. While it does not yet match the Fried score in terms of sensitivity, it provides a solid foundation for structured, reproducible, and user-friendly frailty detection. The integration of both statistical and operational dimensions offers tangible prospects for the early identification of motor frailty in older adults, whether in community or hospital settings.

5 Conclusion

This work is part of a broader effort to optimize the Functional Motor Frailty Screening Score (SFDFM), a tool specifically designed to meet the needs of physiotherapists in screening frailty among older adults. By combining rigorous statistical approaches with advanced digital development, this internship laid a solid foundation for the practical clinical use of the SFDFM.

The results confirmed the good internal validity of the score, with a high explanatory power ($R^2 = 0.8185$), and identified key contributing items (orthopedic assessment, fear of falling, single-leg stance). Robust numerical thresholds (8 and 18) were also defined to classify patients by frailty level in a simple and reproducible manner.

However, several limitations were highlighted, including lower sensitivity of the SFDFM compared to the Fried score, and instability of explanatory variables across clinical subgroups. These findings support the continuation of development work toward an optimized and potentially modular version of the questionnaire.

Moreover, the digital evolution of the SFDFM represents a significant step forward : the developed web interface is ergonomic, secure, and ready for large-scale data collection. This combination of clinical tool and digital solution strengthens the applicability of the score and paves the way for use in primary care, telehealth, or regional prevention programs.

In the short term, it would be relevant to validate these results on a larger and more diverse cohort, and progressively integrate currently missing clinical dimensions. In the medium term, deploying the tool in real-world settings will help refine its use, enhance its clinical relevance, and establish the SFDFM as a reference instrument for identifying motor frailty in frontline healthcare.

Références

[1] Delphine Blanchet, Helene Camerlynck, Carole Duchesne, and Nathalie Brouard. Pre-frailty, frailty and multimorbidity : prevalences and associated characteristics from two french national surveys. `https://www.santepubliquefrance.fr/docs/pre-frailty-frailty-and-multimorbidity-prevalences-and-associated-characteristics-from-two-french-national-surveys`, 2017. Santé Publique France.

[2] Conseil National de l'Ordre des Masseurs-Kinésithérapeutes. Guide prévention primaire – dépistage par les kinésithérapeutes de la fragilité motrice par un score fonctionnel chez les personnes de 65 ans et plus, 2022. Commission santé publique et démographie, en collaboration avec Pr. F. Mourey et Dr. A. Kubicki.

[3] Linda P Fried, Catherine O Tangen, Jeremy Walston, Anne B Newman, Charles Hirsch, John Gottdiener, Teresa Seeman, Rick Tracy, Willem J Kop, Gregory L Burke, and Mary Ann McBurnie. Frailty in older adults : evidence for a phenotype. *The Journals of Gerontology Series A : Biological Sciences and Medical Sciences*, 56(3) :M146–M156, 2001.

[4] Angelika Geroldinger, Lara Lusa, Mariana Nold, and Georg Heinze. Leave-one-out cross-validation, penalization, and differential bias of some prediction model performance measures—a simulation study. *Diagnostic and Prognostic Research*, 7(1) :9, 2023.

[5] Heroku. Heroku : Cloud platform as a service, 2025. Accessed : 2025-04-04.

[6] A. Le Ménez. Analyse des niveaux de fragilité motrice chez les patients âgés, 2025. Encadré par Mme Combourieu.

[7] Nicolas Leroy. Abord clinimétrique de la fiabilité inter-évaluateur du score de dépistage de la fragilité motrice chez les personnes âgées de plus de 65 ans, 2024. Sous la direction de Mme COMBOURIEU, MKDE, PhD, et sous l'expertise de M. GOMES AMARAL, MKDE.

[8] X. Liu, Y. Wang, and M. Chen. Application of xgboost to identify idiopathic central precocious puberty without gnrh stimulation test. *BMC Endocrine Disorders*, 25(1) :112–120, 2025.

[9] R. Martinez, A. Lopez, and J. Dubois. Xgboost-based prediction of 30-day mortality after cardiac surgery : A multicenter study. *Bioengineering*, 11(4) :75–84, 2024.

[10] Pandas Development Team. Pandas documentation. Official documentation for the Pandas library.

[11] Plotly. Dash canvas documentation, 2025. Accessed : 2025-04-04.

[12] Plotly. Dash : Python framework for web applications, 2025. Accessed : 2025-04-04.

[13] M. Texier. Étude de la validité du score fonctionnel de dépistage de la fragilité motrice (sfdfm), 2024. Sous la direction de L. Combourieu.

[14] H. Zhang, L. Zhao, and Q. Sun. Risk prediction of infections in decompensated cirrhotic patients using a single-tree xgboost model. *BMC Gastroenterology*, 23(2) :45–53, 2023.