|  |  |  |
| --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | FGAI4H-I-012-A01 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | Plen | E-meeting, 7-8 May 2020 |
| **DOCUMENT** |
| **Source:** | TG-Falls Topic Driver |
| **Title:** | Att.1- TDD update (TG-Falls) [same as Meeting H] |
| **Purpose:** | Discussion |
| **Contact:** | Inês SousaFraunhofer Portugal | Tel: +351 220 430 326Email: ines.sousa@fraunhofer.pt |
| **Contact:** | Maria VasconcelosAssociação Fraunhofer Portugal Research – Fraunhofer AICOSPortugal | Tel: +351 220 430 353Email: maria.vasconcelos@fraunhofer.pt |
| **Contact:** | Pierpaolo PalumboUniversity of BolognaItaly | Tel: +39 3402378412Email: pierpaolo.palumbo@unibo.it |

|  |  |
| --- | --- |
| **Abstract:** | The document presents the status report on the topic Standardized benchmarking of AI to prevent falls among the elderly. A database of 403 subjects who were evaluated for multiple fall risk factors and 12-months prospective falls is available for benchmarking. Contacts have been taken with two epidemiological studies about ageing for additional data. The database aims at contributing to standardize fall risk assessment and creating tools to easily implement it in the clinical practice.This version of the TDD is the same as seen in Meeting H (FGAI4H-H-012-A01), reproduced for easier reference as a Meeting I document.  |

[1 Introduction 3](#_Toc30437333)

[1.1 Topic Description 3](#_Toc30437334)

[1.1.1 Relevance 3](#_Toc30437335)

[1.1.2 Impact 3](#_Toc30437336)

[1.2 Ethical Considerations 3](#_Toc30437337)

[1.3 Existing AI Solutions 4](#_Toc30437338)

[1.3.1 Existing work 4](#_Toc30437339)

[1.4 Existing work on benchmarking 5](#_Toc30437340)

[2 AI4H Topic Group 5](#_Toc30437341)

[2.1 Updates 5](#_Toc30437342)

[2.1.1 Status Update for Meeting B (Lausanne) 5](#_Toc30437343)

[2.1.2 Status Update for Meeting C (New York) 5](#_Toc30437344)

[2.1.3 Status Update for Meeting D (Shanghai) 5](#_Toc30437345)

[2.1.4 Status Update for Meeting E (Geneva) 5](#_Toc30437346)

[2.1.5 Status Update for Meeting F (Zanzibar) 6](#_Toc30437347)

[2.1.6 Status Update for Meeting G (New Dehli) 6](#_Toc30437348)

[2.1.7 Status Update for Meeting H (Brasilia) 6](#_Toc30437349)

[2.2 Next Meetings 7](#_Toc30437350)

[3 Method 7](#_Toc30437351)

[3.1 AI Inputs Data Structure 7](#_Toc30437352)

[3.1.1 Data Availability 7](#_Toc30437353)

[3.1.2 Data Quality 8](#_Toc30437354)

[3.2 AI Output Data Structure 8](#_Toc30437355)

[3.3 Test Data Labels 8](#_Toc30437356)

[3.4 Score and Metrics 8](#_Toc30437357)

[3.5 Undisclosed Test Data Set Collection 8](#_Toc30437358)

[3.6 Benchmarking Methodology and Architecture 9](#_Toc30437359)

[3.7 Reporting Methodology 9](#_Toc30437360)

[4 Results 9](#_Toc30437361)

[5 Discussion 9](#_Toc30437362)

[6 Declaration of Conflict of Interest 9](#_Toc30437363)

[7 References 9](#_Toc30437364)

# Introduction

Former documents:

|  |  |  |
| --- | --- | --- |
| FGAI4H-C-014 | Lausanne, 22-25 January 2019 | Status Report of: Reducing risk of falling among elderly |

This submission was provided in response to the ITU-T FG-AI4H's call for proposals on use cases and data A‑102. The document was presented remotely. The project aims at contributing to standardize fall risk assessment and creating tools based on Artificial Intelligence to easily implement it in the clinical practice.

## Topic Description

### Relevance

Falls are one of the most common health problems in the elderly population, about a third of community-dwelling adults aged 65 years or older fall each year [1], and these events represent more than 50% of the hospitalizations due to lesions in this age group. Falls are also considered one of the main causes for loss of independence and institutionalization. In 10% of cases falls result in fractures, thus contributing to significant increases in morbidity and mortality. Direct health care costs associated with this phenomenon are high, reaching yearly costs of 25 billion euros in the European Union [2] and 31 billion dollars in the United States of America [3]. Falls have a multifactorial origin, however most of the fall risk factors are amendable by implementing falls prevention programs based on improving strength and balance and modifying behaviours. Even though, fall risk screenings and the implementation of such falls’ prevention programs are rarely part of the community-dwelling elder’s routine. The current state-of-the-art assessment of fall risk factors is based mainly in applying clinical scales, such as the Morse Fall Scale [4], the Berg Balance Scale [5], and the Performance Oriented Assessment of Mobility Problems in Elderly Patients [6]. Despite being recommended by international health bodies, such as the National Institute for Health and Care excellence (NICE)[[1]](#footnote-1), multifactorial fall risk screening is still not widespread in the clinical practice. One of the reasons for this shortcoming is the difficulty in combining the multiple parameters evaluated in a meaningful scale that is able to discriminate those who are more likely to fall in a period of time following the assessment.

### Impact

Raise awareness for multifactorial assessment of fall risk factors, contribute to standardize fall risk assessment and create tools to easily implement it in the clinical practice.

Artificial Intelligence (AI) techniques can be of great value in generating models that combine multiple sources of data and enable the implementation and standardization of a multifactorial assessment of the risk of falling. This would enable the creation of a meaningful scale that is able to differentiate those who are more likely to fall in one year after the assessment.

## Ethical Considerations

Specific ethical considerations should include the fact that fear of falling is also a risk factor for falls, thus an indication of the presence of high fall risk should be accompanied by a plan for mitigation of this risk and comprehensive explanation of preventive measures.

## Existing AI Solutions

TBC

Existing studies focused on clinical, self-reported, or variables resulting from the application of clinical tests (e.g., [7], [8]), report sensitivities varying from 43% to 100% (median = 80%), whereas the specificity ranged from38% to 96% (median = 75%).

Howcroft et al. [9] reviewed previous studies focusing on the fall risk assessment with inertial sensors. The authors concluded that future research should i) consider investigating the relationship between the models’ predictive variables and specific fall risk factors and ii) focus on groups with an increased fall risk due to some diseases. A weak point of most studies is not having used separate datasets for model training and validation, which could have impacted the models’ applicability beyond the training set population. Another aspect to be considered is that clinical assessment thresholds were not used consistently across the research studies included in the review. The prospective fall occurrence rate is considered to be the most reliable criterion for dividing subjects into non-fallers and fallers [9]; however, this criterion was only used in 15% of the studies. Regarding the retrospective fall assessment, the most relevant limitations are the inaccurate recording of fall histories most commonly assessed by self-reported questionnaires and the fact that balance, strength, and gait parameters can change due to past falls.

### Existing work

There is previous work developed under the project FallSensing where Fraunhofer AICOS, Coimbra Health School and Sensing Future Technologies, have collected a dataset of 537 test subjects, to whom a multifactorial assessment of fall risk factors was applied following the protocol described in [10].

The screening includes questions about demographic and anthropometric data, health and lifestyle behaviours, a detailed explanation about procedures to accomplish 6 functional tests (grip strength, Timed Up and Go, 30 seconds sit to stand, step test, 4-Stage Balance test “modified,” and 10-meter walking speed), 3 questionnaires concerning environmental home hazards, and an activity and participation profile related to mobility and self-efficacy for exercise.

In addition to these state-of-the-art fall risk evaluation methods, some of the functional tests have been instrumented, i.e. have been carried out while using a pressure platform and two inertial sensors. This has the potential to make these functional tests more valuable, since the sensors provide additional quantitative information and eliminate the bias introduced by observation [11]. However, since the use of sensors may not be widespread or feasible for most organizations, the inclusion of these data is still under evaluation.

A similar protocol has been employed in the last two waves of the InCHIANTI[[2]](#footnote-2), an epidemiological longitudinal study about mobility in older adults. Prospective falls have been recorded with monthly phone interviews and functional tests have been instrumented with inertial sensors embedded in smartphones.

In a recent study [12], the FallSensing multifactorial screening protocol was applied to 281 community-dwelling adults aged over 65, and their 12-month prospective falls were annotated. Clinical and self-reported data, along with data from instrumented functional tests, involving inertial sensors and a pressure platform, were fused using early, late, and slow fusion approaches. For the early and late fusion, a classification pipeline was designed employing stratified sampling for the generation of the training and test sets. Grid search with cross-validation was used to optimize a set of feature selectors and classifiers. According to the slow fusion approach, each data source was mixed in the middle layers of a multilayer perceptron. The three studied fusion approaches yielded similar results for the majority of the metrics. However, if recall is considered to be more important than specificity, then the result of the late fusion approach providing a recall of 78.6% is better compared with the results achieved by the other two approaches.

## Existing work on benchmarking

TBC

The Timed Up and Go Test (TUG) is one of the most widespread among the traditional tools for risk screening. Its performance has been evaluated many times over the years in different studies and population. Two systematic reviews report much heterogeneity in its performances across studies and a relatively low average predictive accuracy [13, 14].

Among the new tools for predicting falls in the elderly that consider multiple risk factors, FRAT-up was validated on four European datasets of longitudinal studies about ageing. It showed to be more accurate that simple traditional tools and, similarly to TUG, exhibits much heterogeneity in its performance across different populations [15, 16].

Heterogeneity across datasets and populations was found also on fall incidence and fall risk factors prevalence rates, the reason being yet to be fully uncovered [17]. From this experience we believe that benchmarking fall prediction algorithms on different datasets/populations is necessary to obtain robust estimates of their performance. Furthermore, these datasets should be as much as possible representative of their target populations.

# AI4H Topic Group

## Updates

### Status Update for Meeting B (Lausanne)

First submission was provided in response to the ITU-T FG-AI4H's call for proposals on use cases and data A‑102. The document was presented remotely.

|  |  |  |
| --- | --- | --- |
| FGAI4H-C-014 | Lausanne, 22-25 January 2019 | Status Report of: Reducing risk of falling among elderly |

### Status Update for Meeting C (New York)

The topic group description was refined and presented remotely.

### Status Update for Meeting D (Shanghai)

Inês Sousa participated remotely in the Shangai meeting and provided an update on the progress of the topic “Standardized benchmarking of AI to prevent falls among the elderly”.

Main points:

– There were no contacts or manifestations of interest from other research groups regarding this topic;

– It was suggested that some of the groups that have been actively publishing in this area could be contacted;

– It was mentioned that the possibility of enlarging the scope of the topic to include fall detection datasets could also be considered, despite the unavailability of Fraunhofer AICOS to provide a dataset.

### Status Update for Meeting E (Geneva)

The Personal Health Systems Laboratory from University of Bologna joined the Topic Group following the manifestation of interest sent by Pierpaolo Palumbo, biomedical engineer, working on algorithms for health risk assessment, with a focus on fall risk in community-dwelling older adults and lower-limb amputees. He is a post-doctoral fellow at the Personal Health Systems Laboratory, headed by Prof. Lorenzo Chiari, at the University of Bologna.

### Status Update for Meeting F (Zanzibar)

Following the suggestion from the Personal Health Systems Laboratory, a list of longitudinal studies on ageing with data on falls has been drafted. A draft letter was created inviting these studies to share the data of the new waves with our consortium for benchmarking the algorithms.

### Status Update for Meeting G (New Dehli)

Update on contacts with longitudinal studies on ageing with data on falls and groups that have been actively publishing in this area, results in demonstrations of interest to join the Focus Group from Kim van Schooten, PhD, Human Frontier Science Program Postdoctoral Fellow, Conjoint Senior Lecturer, UNSW Medicine, UNSW Ageing Futures Institute, and, Barry Greene from Kinesis, Ireland.

A paper [12] was published.

### Status Update for Meeting H (Brasilia)

Demonstrations of potential interest from two groups relative to two epidemiological studies about ageing with data on falls: InCHIANTI[[3]](#footnote-3) and TILDA[[4]](#footnote-4).

Questions:

* The InCHIANTI dataset is generally shared with interested researchers under formal agreements with a non-disclosure clause. The InCHIANTI board keep track of the researchers that have accessed the different versions of their dataset. Furthermore, the waves that could be available for the benchmarking activities of the Focus Group have been shared with a relatively small number of persons. Could the benchmarking framework accept these data (excluding from the benchmark the models coming from researchers that have accessed the data)?
* The different datasets are mostly similar but slightly different in terms of available variables. Should we keep all useful variables from both datasets or should we restrict the datasets to the variables that are present in both datasets?
* Because of their design, neither FallSensing nor InCHIANTI can be considered rigorously representative of the Portuguese or Italian older population. How do we take this into account?
* What is the approximate time schedule of our activities?

A conference call to which the subjects expressing interest in the activities of the Topic Group have been invited, was held.

Participants:

Inês Sousa, Fraunhofer AICOS

Pierpaolo Palumbo, University of Bologna

Stefania Bandinelli, SOC Geriatria -USLToscana Centro, Firenze

Barry Greene, Chief Technology Officer, Kinesis Health Technologies, Ireland

Salman Khan, Assistant Professor in the department of electrical engineering, University of Engineering and Technology Peshawar, Pakistan

Brief summary of the points discussed:

* A systematic assessment of all solutions and studies regarding fall risk assessment is missing;
* Quality levels and standards for algorithm evaluation should be defined;
* Most datasets available are heterogenous and consider different variables and functional tests, may include data from sensors or not.

Action Points:

* Systematize information regarding fall risk assessment;
* Continue the discussion of the variables to be considered, and methods/best practices for algorithm evaluation;
* Discuss with the Working Group how should the Benchmarking Framework deal with heterogenous datasets.

#### Next steps

The topic group would benefit from further expertise of the medical and AI communities and from additional data. Disseminate the Call for Topic Group Participation among groups with similar research interest.

Agreement on methodological aspects after discussion within the Topic Group and interactions with Leadership Board and Working Groups.

Continue the development of the Action Points defined in the Topic Group conference call.

## Next Meetings

The Focus Groups meets about every two months at changing locations. The upcoming meetings are:

● I: Singapore, 16-20 March 2020

● J: Geneva, 4-8 May 2020

An up to date list can be found at the official ITU FG AI4H website.

# Method

## AI Inputs Data Structure

### Data Availability

Regarding data availability, there are two available datasets so far: FallSensing and InCHIANTI.

FallSensing is made of 403 annotated data samples. The data have been kept undisclosed. Only a small part of it can be made publicly available (1 or 2%) for model training, while the rest can be used for model testing. However, since the data acquisition protocol is published in an open access journal, it can be easily replicated by peers.

The protocol of the InCHIANTI is much similar to the one adopted in FallSensing. Data on prospective falls and instrumented functional tests are available for the last two waves (FU4 and FU5). The data are generally shared with other research groups on the basis of formal agreements with a non-disclosure clause.

A third dataset could become available in 2021: TILDA.

### Data Quality

The FallSensing data were acquired by trained health professionals , following a convenience sampling method. All participants completed the whole assessment, unless they were not physically able to do so. The database is a *comma*-*separated values* file.

The InCHIANTI data were collected within a longitudinal study following a population-based sampling method. Because of attritions due to deaths and dropouts since its baseline, the data samples available in waves FU4 e FU5 are relative to subjects who are older than the initial representative sample.

## AI Output Data Structure

1. **Annotation / Label Quality**

The health professionals called each participate every month for one year after the assessment in order to record the rate of falls. This procedure follows similar studies in the literature.

1. **Data Provenance**

Ethical approval for FallSensing was obtained from the Research Ethics Committee of Polytechnic Institute of Coimbra (Nº6/2017). All participants gave written informed consent before data collection begins as per the Declaration of Helsinki.

The InCHIANTI study protocol was approved by the ethical committee of the Italian National Institute of Research and Care of Aging and complies with the Declaration of Helsinki.

## Test Data Labels

After the assessment, 403 of the participants from the FallSensing study received monthly phone calls over a 12-month period to record the rate of falls in this period. The dataset is thus annotated with the rate of reported falls in the period of 12 months to record whether a fall occurred in this period. The dataset is thus annotated with the number and month of reported falls in the period of 12 months following the assessment. This annotation can be converted in a binary outcome, diving the dataset in fallers (subjects who fell at least once in the 12-month follow-up period), and non-fallers (remaining). Within InCHIANTI FU4 and FU5 a similar protocol was followed. Precise determination of the number of available participants is still to be done.

## Score and Metrics

As possible performance metrics, we are currently considering the sensitivity, specificity and area under the receiver operating characteristic curve (ROC AUC) applied to a binary classification problem (occurrence of at least one fall in subsequent year vs. non-occurrence of falls in that period). Other possible outcomes are the predicted time until the first fall (in months), or the probability of suffering a fall in a given period during the year after the assessment. The problem can also be formulated as multiclass classification, allowing to distinguish groups of first-time fallers after the assessment, recurrent fallers and non-fallers, for example.

If we require to have probabilistic predictions (e.g. absolute risk of falling within 12 months), we will also be able to evaluate the models with respect to their calibration.

TBC

## Undisclosed Test Data Set Collection

TBD

## Benchmarking Methodology and Architecture

For the benchmarking task, participants should submit AI models to combine multiple fall risk factors assessed in community-dwelling adults aged over 50 years old and distinguish fallers from non-fallers, i.e. the subjects that suffered at least one fall in the year subsequent to the assessment from those who did not fall in that period.

TBC

## Reporting Methodology

TBC

# Results

TBC

# Discussion

TBC

# Declaration of Conflict of Interest

There are no conflicts of interest.

# References

[1] World Health Organization. WHO Global Report on Falls Prevention in Older Age. Ageing and life-course 2007.

[2] ProFound. Active ageing through preventing falls: “Falls prevention is everyone's business. Prev Falls Netw Dissem 2015.

[3] Burns EB, Stevens JA, Lee RL. The direct costs of fatal and non-fatal falls among older adults—United States. J Safety Res 2016:58.

[4] Morse JM, Morse RM, Tylko SJ. Development of a scale to identify the fall-prone patient. Can J Ageing 1989;8(4):366-377.

[5] Berg K, Wood-Dauphinen S, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. Physiother Canada 1989;41(6):304-311.

[6] Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. J Am Geriatr Soc 1986 Feb;34(2):119-126. [Medline: 3944402]

[7] A. Nelson, K. L. Perell, L. Z. Rubenstein, N. Prieto-Lewis, R. L.Goldman, and S. L. Luther, “Fall Risk Assessment Measures: AnAnalytic Review,”The Journals of Gerontology: Series A, vol. 56,pp. M761–M766, 12 2001.

[8] M. Vassallo, L. Poynter, J. C. Sharma, J. Kwan, and S. C. Allen, “Fallrisk-assessment tools compared with clinical judgment: an evaluation ina rehabilitation ward,”Age and Ageing, vol. 37, pp. 277–281, 01 2008.

[9] J. Howcroft, J. Kofman, and E. D. Lemaire, “Review of fall riskassessment in geriatric populations using inertial sensors,”Journal ofNeuroEngineering and Rehabilitation, vol. 10, p. 91, Aug 2013.

[10] Martins, A. C., Moreira, J., Silva, C., Silva, J., Tonelo, C., Baltazar, D., . . . Sousa, I. Multifactorial screening tool for determining fall risk in community-dwelling adults aged 50 years or over (FallSensing): Protocol for a prospective study. *Journal of Medical Internet Research,* 2018, *20*(8) doi:10.2196/10304

[11] Silva J, Sousa I. Instrumented timed up and go: Fall risk assessment based on inertial wearable sensors. 2016 Presented at: 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA); 15-18 May 2016; Benevento, Italy. [doi: 10.1109/MeMeA.2016.7533778]

[12] J. R. Silva, I. Sousa and J. S. Cardoso, "Fusion of Clinical, Self-Reported, and Multisensor Data for Predicting Falls," in IEEE Journal of Biomedical and Health Informatics.
doi: 10.1109/JBHI.2019.2951230

[13] E. Barry, R. Galvin, C. Keogh, F. Horgan, and T. Fahey, “Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta- analysis.,” BMC Geriatr., vol. 14, no. 1, p. 14, Jan. 2014.

[14] D. Schoene et al., “Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis.,” J. Am. Geriatr. Soc., vol. 61, no. 2, pp. 202–8, Feb. 2013.

[15] P. Palumbo, L. Palmerini, S. Bandinelli, and L. Chiari, “Fall Risk Assessment Tools for Elderly Living in the Community: Can We Do Better?,” PLoS One, vol. 10, no. 12, p. e0146247, Dec. 2015.

[16] P. Palumbo et al., “Predictive Performance of a Fall Risk Assessment Tool for Community-Dwelling Older People (FRAT-up) in 4 European Cohorts,” J. Am. Med. Dir. Assoc., vol. 17, pp. 1106–1113, 2016.

[17] K. Rapp et al., “Fall incidence in Germany: results of two population-based studies, and comparison of retrospective and prospective falls data collection methods.,” BMC Geriatr., vol. 14, p. 105, Jan. 2014.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. https://www.nice.org.uk/guidance/cg161/chapter/recommendations#multifactorial-assessment-or-multifactorial-falls-risk-assessment [↑](#footnote-ref-1)
2. http://inchiantistudy.net/wp/ [↑](#footnote-ref-2)
3. http://inchiantistudy.net/wp/ [↑](#footnote-ref-3)
4. https://tilda.tcd.ie/ [↑](#footnote-ref-4)