Original citation:

Permanent WRAP URL:
http://wrap.warwick.ac.uk/97889

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 3.0 International license (CC BY 3.0) and may be reused according to the conditions of the license. For more details see: http://creativecommons.org/licenses/by/3.0/

A note on versions:
The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk
Wearable Inertial Sensors for Fall Risk Assessment and Prediction in Older Adults: A Systematic Review and Meta-Analysis

Luis Montesinos, Student Member, IEEE, Rossana Castaldo, Student Member, IEEE, and Leandro Pecchia, Member, IEEE

Abstract—Wearable inertial sensors have been widely investigated for fall risk assessment and prediction in older adults. However, heterogeneity in published studies in terms of sensor location, task assessed and features extracted is high, making challenging evidence-based design of new studies and/or real-life applications. We conducted a systematic review and meta-analysis to appraise the best available evidence in the field. Namely, we applied established statistical methods for the analysis of categorical data to identify optimal combinations of sensor locations, tasks and feature categories. We also conducted a meta-analysis on sensor-based features to identify a set of significant features and their pivot values. The results demonstrated that with a walking test, the most effective feature to assess the risk of falling was the velocity with the sensor placed on the shins. Conversely, during quite standing, linear acceleration measured at the lower back was the most effective combination of feature-placement. Similarly, during the sit-to-stand and/or the stand-to-sit tests, linear acceleration measured at the lower back seems to be the most effective feature-placement combination. The meta-analysis demonstrated that four features resulted significantly higher in fallers: the root-mean-square acceleration in the mediolateral direction during standing with eyes closed (Mean Difference (MD): 0.01 g; 95% Confidence Interval (CI95%): 0.006 to 0.014); the number of steps (MD: 1.638 steps; CI95%: 0.384 to 2.892) and total time (MD: 2.274 seconds; CI95%: 0.531 to 4.017) to complete the Timed Up and Go test; and the step time (MD: 0.053; CI95%: 0.012 to 0.095; p=0.01) during walking.

Index Terms—Inertial sensors, accidental falls, fall prediction, fall risk assessment, systematic literature review, meta-analysis

I. INTRODUCTION

The incidence of accidental falls among older adults, along with their impact in terms of morbidity and mortality, have turned them into a public health concern worldwide. It has been estimated that 28% to 45% of people aged 65 and over fall each year [1]. These events represent 18% to 40% of emergency department attendances and over 80% of all injury admissions to hospitals among the same age group. Among the most serious injuries resulting from falls are hip fracture and traumatic brain injury; the latter accounts for 46% of fatal falls among older adults [2].

Accidental falls have also a great impact in terms of costs for healthcare systems and for the society. Only in the United Kingdom, their annual cost to the National Health System has been estimated in £2.3 billion per year [3]. Moreover, falls lead to indirect costs, such as the loss of productivity of family members and other caregivers. The average lost earnings due to falls has been estimated in US$40,000 per year for the UK [1].

Nowadays, clinical fall risk assessment relies mostly on moderately to highly comprehensive medical, fall-risk specific and functional mobility assessment tools in the form of questionnaires, physical tests, gait analysis, and physical activity measurements [4]. Among the most popular assessment tests and tools are the Timed Up and Go (TUG) test [5], the Tinetti Assessment Tool [6], the STRATIFY score [7] and the Five-Times-Sit-to-Stand (FTSS) test [8].

More recently, researchers have investigated the potential use of instrumented fall-risk assessment and prediction tools based on features extracted from inertial sensors (i.e. accelerometers and gyroscopes) attached to the subject’s body during specific assessment tasks (e.g. walking, quiet standing, sit-to-stand transitions) [9]–[11]. In those studies, machine learning methods were used to automatically identify fallers (F) and non-fallers (NF). Subjects were labelled as F/NF using at least one of the following methods: a fall-risk assessment test conducted in the clinical setting (e.g. TUG test), self-reported fall occurrence within a follow-up period from the assessment or fall history.

Howcroft et al. [9] and Shany et al. [11] have presented insightful accounts of features, classification models and validation strategies related to sensor-based fall-risk testing (SFRT). In their investigations, these authors found large heterogeneity in terms of sensor placement, tasks assessed, and sensor-based features. Not surprisingly, they also found disparate levels of reported sensitivity (55-100%), specificity (15-100%) and accuracy (62-100%).

To design effective interventions, it is crucial to identify the optimal combination of three factors: where to place the sensor, which task to be performed and which features should be extracted and analyzed. The latter is particularly relevant to
This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/.
E. Meta-Analysis of Inertial Sensor-Based Features

A meta-analysis of the features extracted from the shortlisted studies was conducted to identify significant individual features and their pivot values. Features were pooled for meta-analysis if: [feature was reported in at least two studies] AND [feature was computed for the same task/subtask] AND [sensor placement and type was the same across studies OR feature was independent of sensor placement and type (e.g. number of steps or stride time)]. Standard methods for combining and reporting continuous outcomes were employed to pool the features [16]: pooled sample size, mean difference (MD) with 95% confidence intervals (95% CI) and statistical significance level (p-value). MDs and 95% CIs were considered significant if the p-value was found to be smaller than 0.05.

Random or fixed effect models were selected based on heterogeneity across studies, assessed using the Q-statistic (computed via a Chi-squared test) and the I² statistic. A significant Q-statistic is indicative of dissimilar effect sizes across studies; a threshold significance level of 0.1 was selected as statistically significant value as suggested in [16]. The I² statistic indicates the percentage of the variability in effect sizes due to heterogeneity across studies, and not due to sampling error within studies. An I² from 30% to 60%, 50% to 90% and 75% to 100% represent moderate, substantial and considerable error within studies. An I² of 0% or larger indicates no/unclear. These questions are organized in 3 dimensions:

- Reporting (11 items) – which assessed whether the information provided in the paper was clear and sufficient to replicate the study and appraise its validity.
- External validity (2 items) – which addressed the extent to which the findings of the study could be generalized to a wider population and context.
- Internal validity (2 items) – which assessed whether the evidence at hand suggests that the study was designed and conducted to minimize bias and confounding.

A summary of the main findings is provided in this paper in an attempt to reveal the methodological issues that future studies in the field should address in order to produce more valid scientific evidence.

III. RESULTS

According to the search strategy described above, 481 records were identified through database search and 18 through linear search. After removing 51 duplicates, 448 titles were screened by title and 257 were excluded as they did not meet the inclusion/exclusion. From the remaining 191 titles, 127 were removed after screening the abstract against inclusions and exclusions.

F. Quality Appraisal of Shortlisted Studies

The methodological quality of the studies was assessed using the checklist provided in the supplementary material (Document S1). This checklist was adapted from Downs and Black [20]. It contains 15 questions that are scored “yes” or “no/unclear”. These questions are organized in 3 dimensions:

- Reporting (11 items) – which assessed whether the information provided in the paper was clear and sufficient to replicate the study and appraise its validity.
- External validity (2 items) – which addressed the extent to which the findings of the study could be generalized to a wider population and context.
- Internal validity (2 items) – which assessed whether the evidence at hand suggests that the study was designed and conducted to minimize bias and confounding.

A summary of the main findings is provided in this paper in an attempt to reveal the methodological issues that future studies in the field should address in order to produce more valid scientific evidence.

Fig. 1. Flowchart indicating the results of the systematic review with inclusions and exclusions.
TABLE I
DESCRIPTION OF SHORTLISTED STUDIES

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Subjects (Fallers)</th>
<th>Mean age ± SD (years)</th>
<th>(Non-)Faller labelling method</th>
<th>Type of sensor</th>
<th># of sensors</th>
<th>Location</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kojima, 2008</td>
<td>153 (22)</td>
<td>71 ± 7.7</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>1</td>
<td>Lower back</td>
<td>Walking</td>
</tr>
<tr>
<td>O’Sullivan, 2009</td>
<td>17 (12)</td>
<td>77 ± 7.5</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>1</td>
<td>Lower back</td>
<td>Quiet standing</td>
</tr>
<tr>
<td>Greene, 2010</td>
<td>349 (207)</td>
<td>72.4 ± 7.4</td>
<td>Retrospective fall history</td>
<td>Gyroscope</td>
<td>2</td>
<td>Shins</td>
<td>Timed Up and Go test</td>
</tr>
<tr>
<td>Paterson, 2011</td>
<td>97 (54)</td>
<td>68.7 ± 7.1</td>
<td>Prospective fall occurrence</td>
<td>Accelerometer</td>
<td>2</td>
<td>Feet</td>
<td>Walking</td>
</tr>
<tr>
<td>Weiss, 2012</td>
<td>41 (23)</td>
<td>78.2 ± 6.2</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>1</td>
<td>Lower back</td>
<td>Timed Up and Go test</td>
</tr>
<tr>
<td>Doheny, 2012</td>
<td>40 (19)</td>
<td>71.4 ± 7.3</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>2</td>
<td>Shins / and Gyroscope</td>
<td>Walking and Quiet standing</td>
</tr>
<tr>
<td>Greene, 2012</td>
<td>120 (65)</td>
<td>73.7 ± 5.8</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>1</td>
<td>Lower back</td>
<td>Quiet standing</td>
</tr>
<tr>
<td>Itsh, 2012</td>
<td>30 (7)</td>
<td>75 ± 5.7</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>4</td>
<td>Lower back, Knee, Ankle, Big toe</td>
<td>Walking</td>
</tr>
<tr>
<td>Senden, 2012</td>
<td>100 (50)</td>
<td>76.5 ± 5.7</td>
<td>Risk assessment tool</td>
<td>Accelerometer</td>
<td>1</td>
<td>Lower back</td>
<td>Walking</td>
</tr>
<tr>
<td>Doheny, 2013</td>
<td>39 (19)</td>
<td>71.5 ± 6.6</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>2</td>
<td>Thigh, Sternum</td>
<td>Five-Times Sit-to-Stand test</td>
</tr>
<tr>
<td>Doi, 2013</td>
<td>73 (16)</td>
<td>80.7 ± 7.8</td>
<td>Prospective fall occurrence</td>
<td>Accelerometer</td>
<td>2</td>
<td>Lower back, Upper back</td>
<td>10-m Walk test</td>
</tr>
<tr>
<td>Weiss, 2013</td>
<td>71 (32)</td>
<td>78.4 ± 4.7</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>1</td>
<td>Lower back</td>
<td>Walking</td>
</tr>
<tr>
<td>Cui, 2014</td>
<td>81 (39)</td>
<td>78.4 ± 4.8</td>
<td>Retrospective fall history</td>
<td>Accelerometer</td>
<td>1</td>
<td>Lower back</td>
<td>Walking</td>
</tr>
</tbody>
</table>

* Estimated from the data reported in paper
* Two or more falls in the recall period
* One or more falls in the recall period, or one fall resulting in injury or requiring medical attention

Subjects were labelled as (non-)fallers using retrospective fall history in 10 studies, with a recall period of one year for 8 studies and 5 years for 2 studies; prospective fall occurrence through a one-year follow-up period in 2 studies; and a clinical assessment tool (the Tinetti scale) in one study.

Tri-axial accelerometers and gyroscopes were the only type of inertial sensor used in 10 and 1 studies respectively; a combination of sensors were used in 2 studies. In 7 studies, only one sensor was used; in 5 studies two sensors were used; and 1 study used four sensors.

The most common sensor placement was the lower back (i.e. approximately on L3) with 10 studies, followed by shins and feet with 2 studies each. Other placements were knee, ankle, thigh, sternum and upper back (i.e. approximately on C7), with one study each. If placements are grouped in upper body (trunk) and lower body (lower limbs), there were eleven (91.7%) and seven (58.3%) studies, respectively.

Inertial signals were acquired during the following tasks: walking otherwise than a standardized test (7 studies), quiet standing (3 studies), the TUG test (2 studies), the 10-Meters Walking test (10MWT) (1 study), and the Five-Times Sit-to-Stand (FTSS) test (1 study). A brief description of these tasks is presented in Table II; for a more detailed description the reader may refer to the referenced paper.

B. Inertial Sensor-Based Features and Their Trends

The full listing of features extracted from inertial sensors that were reported in the 13 selected papers is provided as supplementary material (Table S1). Green et al. [23] reported features for all the subjects included in their analysis as well as for some subgroups separately (i.e. males, females < 75 year old and females ≥ 75 years old). However, only the results for all the subjects were included in this review. Moreover, Doheny et al. [26] performed an instrumented gait assessment four
Covariate | $\chi^2$ | $p$-value | $C$ | $V$ | Association level a
---|---|---|---|---|---
Task | 11.94 | < 0.01 | 0.253 | 0.261 | Medium
Sensor placement | 14.68 | 0.01 | 0.278 | 0.290 | Medium
Feature category | 15.82 | < 0.01 | 0.288 | 0.301 | Medium

$\chi^2$: Pearson’s chi-squared statistic for the association test in which the null hypothesis is ‘no association’

C: Cramer’s coefficient

V: Cramer’s coefficient

a A C (V) of 0.100 (0.1), 0.287 (0.3) and 0.447 (0.5) are considered as evidence of small, medium and large association, respectively

Table III summarizes the frequency of features per feature category, task and sensor placement for the complete listing of features (column A) and for the subset of features showing significant trends (column B).

C. Statistical Analysis of Inertial Sensor-Based Features

The results from the Pearson’s chi-squared tests and the measures of association revealed statistically significant associations between feature significance and feature category, sensor placement and task (Table IV).

Furthermore, the computed Pearson residuals for the three-way table containing feature category, task and sensor placement as covariates revealed strong to very strong associations for 9 triads. Table V summarizes these results. As an example, the double arrow, ‘↑↑’, for the triad ‘angular velocity-walking-shins’ means that significant features are much more likely to arise from this combination. Conversely, the single arrow, ‘↑’, for the triad ‘angular velocity-walking-lower back’ means that significant features are less likely to arise from this combination. The ‘-’ symbol indicated that the significance of a feature is not particularly affected by its category, sensor placement or task.

D. Meta-Analysis of Inertial Sensor-Based Features

Based on the selection criteria for the meta-analysis, 20 features were pooled using the methods described above. Table VI shows the trends and values for those features, as well as the number of subjects in each group. It also shows the task and the sensor placement as covariates revealed strong to very strong associations for 9 triads.
TABLE V
ASSOCIATION TREND AND STRENGTH FOR ALL POSSIBLE TRIADS OF FEATURE CATEGORY, TASK, AND SENSOR PLACEMENT

<table>
<thead>
<tr>
<th>Feature category</th>
<th>Sensor placement</th>
<th>Task</th>
<th>Association strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear acceleration</td>
<td>Upper back</td>
<td>SS TUG Walking</td>
<td>- - - ↓ ↑ ↑</td>
</tr>
<tr>
<td>Spatial</td>
<td>Lower back</td>
<td>SS TUG Walking</td>
<td>- - - ↓ ↑</td>
</tr>
<tr>
<td>Temporal</td>
<td>Shins</td>
<td>SS TUG Walking</td>
<td>- - - ↓</td>
</tr>
</tbody>
</table>

SS: Sit-to-Stand / Stand-to-Sit; TUG: Timed Up and Go test
↓ ↑ ↑: substantially stronger negative (positive) association for a specific triad of feature category, task and sensor placement
↓ ↑: strong negative (positive) association for a specific triad of feature category, task and sensor placement
- - -: either negative or positive non-significant association for a specific triad of feature category, task and sensor placement

sensor placement for each feature.

Linear acceleration features included in the meta-analysis were: Root Mean Square (RMS) value (expressed in g-force units) of acceleration signal in the mediolateral (ML) direction assessed at the lower back during quiet standing with both eyes open and eyes closed (ML RMS of acceleration). This feature is related to postural stability during standing.

Spatial features included in the meta-analysis were: number of steps during the Timed Up and Go (TUG) test, and step length as estimated from inertial signals measured during the walking stage of the TUG test or other walking task.

Temporal features included in the meta-analysis were: cadence (i.e. steps per minute); gait speed; step time; stance time; swing time; stride time; total time to complete the TUG test; single and double support time, i.e. the time during which only one foot and both feet are in contact with the walking surface, respectively, expressed as a percentage of a gait cycle; and the Coefficient of Variation (CV) for step, stance, swing, stride, single and double support times. The CV is the ratio of the standard deviation and mean for a given feature, expressed as a percentage; hence, it is a standardized measure of dispersion of the distribution of feature values.

All the spatial and temporal features included in the meta-analysis are widely used in clinical gait analysis [39].

One frequency feature was included in the meta-analysis: the Harmonic Ratio (HR) of trunk acceleration in the vertical (VT) direction. The HR has been defined as the ratio of even to odd signal harmonics extracted from the spectrum of the acceleration signal and has been suggested as a measure of the stability and smoothness of trunk movement during gait [31].

Neither angular velocity nor non-linear features were included in the meta-analysis, as none of them met the criteria to be pooled; i.e. either they were reported only in one study or they were measured during different tasks or at different sensor body placements.

The relative pooling weight of each study is reported in Table VI. The results of the pooling are reported in Table VII, where also the trend of the pooled features is shown.

Four out of twenty pooled features showed a statistically significant trend associated to fallers. A significantly higher RMS value for the ML acceleration signal (MD: 0.01 g; CI95%: 0.006 to 0.014; p<0.01) during quiet standing with eyes closed. Additionally, a significantly higher number of steps (MD: 1.638 steps; CI95%: 0.384 to 2.892; p<0.01) and a significantly higher total time to complete the TUG test (MD: 2.274 seconds; CI95%: 0.531 to 4.017; p<0.01). Finally, a significantly higher step time (MD: 0.053; CI95%; 0.012 to 0.095; p<0.05).

E. Quality Appraisal of Shortlisted Studies

All the studies reported aim of the study; experimental protocol (i.e. task, sensor quantity and placement); technical specifications of the sensor; methods for signal processing, feature extraction and statistical analysis; and features’ summary statistics per group (non-fallers and fallers). However, only 7 studies reported actual p-values (e.g. 0.035 rather than <0.05) for the feature values’ differences between groups [22], [25], [29]–[33].

Moreover, only 7 studies reported inclusion/exclusion criteria of participants and distribution of potential confounders per group (e.g. age and comorbidities) [24], [25], [27], [29], [31]–[33]. Therefore, the internal validity of 6 studies remains unclear, as unreported (or unobserved) variables could explain feature differences between fallers and non-fallers.

Finally, external validity was found for all shortlisted studies, as their samples were representative of the population under investigation and the task was representative of clinical fall-risk assessment protocols or daily-life activities.

IV. DISCUSSION

This systematic review analyzed the scientific literature focusing on the use of wearable inertial sensors for risk of fall assessment and prediction, exploring the sensitivity sensor-based features to sensor placement, task and feature category.

The statistical analysis of features reported in the 13 shortlisted studies revealed significant, very strong, positive associations in 3 different triads of feature category, task, and sensor placement:

- Angular velocity – Walking – Shins
- Linear acceleration – Quiet standing – Lower back
- Linear acceleration – Stand to sit/Sit to stand – Lower back

These results suggested that these are optimal combinations when using inertial sensors to discriminate between fallers and
non-fallers. Other potentially good combinations, given their strong, positive associations are:
- Frequency – Walking - Lower back
- Frequency – Walking - Upper back
- Temporal - TUG - Shins

Conversely, our findings suggested that the use of following combinations should be avoided as they are less discriminative of fall status:
- Angular velocity – Walking - Lower back
- Frequency – Walking - Shins
- Linear acceleration – Walking - Shins

As for the meta-analysis, the results demonstrated that 4 features significantly increased (p<0.05) among fallers: the RMS acceleration in the mediolateral direction during quiet standing with eyes closed (MD: 0.01 g; CI95%: 0.006 to 0.014); the number of steps (MD: 1.638 steps; CI95%: 0.384 to 2.892) and total time (MD: 2.274 seconds; CI95%: 0.531 to 4.017) to complete the Timed Up and Go test; and the step time (MD: 0.053; CI95%: 0.012 to 0.095; p=0.01) during walking. These results suggest that these combinations of task and features may

<table>
<thead>
<tr>
<th>Feature (units)</th>
<th>Author, Year</th>
<th>Task</th>
<th>Sensor location</th>
<th>Trend</th>
<th>Weight (%)</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Non-Fallers</th>
<th>Fallers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear acceleration features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML RMS acceleration (g)</td>
<td>Doheny, 2012</td>
<td>Quiet standing (EO)</td>
<td>Lower back</td>
<td>-</td>
<td>50.7</td>
<td>21</td>
<td>0.03</td>
<td>0.01</td>
<td>19</td>
<td>0.03</td>
</tr>
<tr>
<td>ML RMS acceleration (g)</td>
<td>Greene, 2012</td>
<td>Quiet standing (EO)</td>
<td>Lower back</td>
<td>↑↑</td>
<td>49.3</td>
<td>55</td>
<td>0.04</td>
<td>0.01</td>
<td>65</td>
<td>0.06</td>
</tr>
<tr>
<td>Spatial features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of steps (steps)</td>
<td>Weiss, 2011</td>
<td>TUG (Walking)</td>
<td>Lower back</td>
<td>↑</td>
<td>43.6</td>
<td>18</td>
<td>10.61</td>
<td>1.80</td>
<td>23</td>
<td>11.52</td>
</tr>
<tr>
<td>Step length (m)</td>
<td>Weiss, 2011</td>
<td>TUG (Walking)</td>
<td>Lower back</td>
<td>↓</td>
<td>49.3</td>
<td>18</td>
<td>0.56</td>
<td>0.08</td>
<td>23</td>
<td>0.53</td>
</tr>
<tr>
<td>Cadence (steps/min)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↓↓</td>
<td>50.2</td>
<td>142</td>
<td>108.00</td>
<td>19.30</td>
<td>207</td>
<td>99.20</td>
</tr>
<tr>
<td>Gait speed (m/s)</td>
<td>Doi, 2013</td>
<td>10MWT</td>
<td>Lower back</td>
<td>↓↓</td>
<td>32.3</td>
<td>57</td>
<td>0.98</td>
<td>0.34</td>
<td>16</td>
<td>0.63</td>
</tr>
<tr>
<td>Step time (s)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↑↑</td>
<td>33.6</td>
<td>50</td>
<td>0.86</td>
<td>0.26</td>
<td>50</td>
<td>1.23</td>
</tr>
<tr>
<td>Swing time (s)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↑↑</td>
<td>29.8</td>
<td>21</td>
<td>0.68</td>
<td>0.10</td>
<td>19</td>
<td>0.70</td>
</tr>
<tr>
<td>Stride time (s)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↓</td>
<td>37.1</td>
<td>142</td>
<td>0.80</td>
<td>0.20</td>
<td>207</td>
<td>0.80</td>
</tr>
<tr>
<td>Total time (s)</td>
<td>Greene, 2010</td>
<td>TUG</td>
<td>Shins</td>
<td>↓↓</td>
<td>52.0</td>
<td>18</td>
<td>1.11</td>
<td>0.11</td>
<td>19</td>
<td>0.70</td>
</tr>
<tr>
<td>Single support time (%)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↓↓</td>
<td>20.2</td>
<td>21</td>
<td>1.11</td>
<td>0.11</td>
<td>19</td>
<td>0.70</td>
</tr>
<tr>
<td>Double support time (%)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↓↓</td>
<td>55.4</td>
<td>142</td>
<td>50.00</td>
<td>20.00</td>
<td>207</td>
<td>40.00</td>
</tr>
<tr>
<td>CV of step time (%)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↑↑</td>
<td>54.4</td>
<td>21</td>
<td>24.67</td>
<td>17.08</td>
<td>19</td>
<td>24.47</td>
</tr>
<tr>
<td>CV of stance time (%)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↑↑</td>
<td>55.4</td>
<td>142</td>
<td>50.00</td>
<td>20.00</td>
<td>207</td>
<td>40.00</td>
</tr>
<tr>
<td>CV of swing time (%)</td>
<td>Greene, 2010</td>
<td>TUG (Walking)</td>
<td>Shins</td>
<td>↑↑</td>
<td>55.4</td>
<td>142</td>
<td>50.00</td>
<td>20.00</td>
<td>207</td>
<td>40.00</td>
</tr>
<tr>
<td>Frequency features</td>
<td>Doi, 2013</td>
<td>10MWT</td>
<td>Lower back</td>
<td>↓↓</td>
<td>50.3</td>
<td>57</td>
<td>2.69</td>
<td>0.93</td>
<td>16</td>
<td>2.07</td>
</tr>
</tbody>
</table>

EO: Eyes Open; EC: Eyes Closed; TUG: Timed Up and Go test; 10MWT: 10-Meters Walking Test
↓↓↑↑: significantly lower (higher) for subjects in the fallers (high-risk) subgroup
↑↑↑↑: lower (higher) for subjects in the fallers (high-risk) subgroup
<: No difference between subgroups
be useful more effective for fall risk assessment.

Additionally, 5 features exhibited a consistent trend across the selected studies. These features were: step time, CV for step time, CV for stride time and CV for single support time, which showed a higher value for fallers when compared to non-fallers; and double support time, which showed a lower value for the same group. However, these trends were not found statistically significant when pooled in the meta-analysis. It may be explained by the high values of standard deviation reported by Green et al [23], which was included in the pooling for these features. No clear explanation for such variability within that study can be inferred from the paper.

In contrast, 7 features showed an opposite trend across the selected studies. These features were: step length, cadence, gait speed, harmonic ratio in the vertical direction, CV for stance time, CV for swing time, and CV double support time. Importantly, for 4 of these features the methods used to classify subjects as (non-)fallers were also inconsistent between studies: step length and cadence were pooled from [25] and [29], in which the classification methods were retrospective fall history and fall risk assessment tool (Timetti scale), respectively. Gait speed was pooled from [25], [28] and [31], the latter adding prospective fall occurrence to the diversity of classification methods. Finally, harmonic ratio on the vertical direction was pooled from [29] and [31], combining subjects classified as fallers via two different methods as well. This fact may represent an important source of between-study heterogeneity, as reflected by the high values of I^2 (>95%) and low significance levels (p < 0.01) obtained in the heterogeneity test for these features. Unfortunately, the low number of studies reporting on the same feature made unfeasible to explore possible sources of heterogeneity using quantitative approaches (e.g. via subgroup analysis stratified by study and/or patient characteristics).

Moreover, 5 features showed an ambiguous trend across the selected studies, as they were reported with no mean difference between non-fallers and fallers in one study, while exhibiting a trend (significant or not) in another study. These features were: the RMS value of acceleration in the mediolateral direction with eyes open, and stance time, swing time, stride time, and single support time during walking.

All in all, the evidence gathered in this review suggests that assessing the Timed Up and Go test using wearable sensors located on the shins through angular velocity, temporal (e.g. total time and step time) and spatial (e.g. number of steps) features may represent an optimal combination to discriminate fallers from non-fallers. Additionally, the triad “linear acceleration-quiet standing-lower back” seems to be a sensible choice as well.

Nevertheless, it should be stressed that these results are limited, as they are based only on features reported in the 13 papers included in the review. Hence, they are unable to provide a representative inference of all features used and all studies published, but not included in the review. It means that there might be some other sensor-based features that are discriminant

---

**TABLE VII**

<table>
<thead>
<tr>
<th>Feature (units)</th>
<th>Heterogeneity</th>
<th>Weighted Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F (%)</td>
<td>Q</td>
</tr>
<tr>
<td><strong>Linear acceleration features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML RMS of acceleration, EO (g)</td>
<td>93.6</td>
<td>15.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of steps (steps)</td>
<td>73.9</td>
<td>3.83</td>
</tr>
<tr>
<td>Step length (m)</td>
<td>96.5</td>
<td>28.58</td>
</tr>
<tr>
<td><strong>Temporal features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cadence (steps/min)</td>
<td>97.1</td>
<td>35.01</td>
</tr>
<tr>
<td>Gait speed (m/s)</td>
<td>97.6</td>
<td>84.47</td>
</tr>
<tr>
<td>Step time (s)</td>
<td>88.1</td>
<td>25.18</td>
</tr>
<tr>
<td>Stance time (s)</td>
<td>0</td>
<td>0.35</td>
</tr>
<tr>
<td>Swing time (s)</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Stride time (s)</td>
<td>57.3</td>
<td>4.68</td>
</tr>
<tr>
<td>Total time (s)</td>
<td>79.6</td>
<td>4.91</td>
</tr>
<tr>
<td>Single support time (%)</td>
<td>53.3</td>
<td>2.14</td>
</tr>
<tr>
<td>Double support time (%)</td>
<td>79.4</td>
<td>4.85</td>
</tr>
<tr>
<td>CV of step time (%)</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>CV of stance time (%)</td>
<td>0</td>
<td>0.69</td>
</tr>
<tr>
<td>CV of swing time (%)</td>
<td>73</td>
<td>3.7</td>
</tr>
<tr>
<td>CV of stride time (%)</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>CV of single support time</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>CV of double support time</td>
<td>69.9</td>
<td>3.32</td>
</tr>
<tr>
<td><strong>Frequency features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VT Harmonic ratio (n.u.)</td>
<td>95.9</td>
<td>24.44</td>
</tr>
</tbody>
</table>

EO: Eyes Open; EC: Eyes Closed
↓↓ (↑↑): significantly lower (higher) for subjects in the fallers (high-risk) subgroup
↓ (↑): lower (higher) for subjects in the fallers (high-risk) subgroup
MD: Mean difference; CI95%: Confidence Interval at 95%; n.u.: dimensionless
Bold values indicate statistically significant trends (p < 0.05)
between non-fallers and fallers but were not included in this systematic review as they were not reported as required by the inclusion criteria. This may be the case of some features reported in [34]–[38].

Finally, a comment regarding heterogeneity in “hit rate” (i.e. the ratio of all features to significant features expressed as a percentage) reported in the shortlisted studies is deemed relevant to this review. In some studies reporting a relatively high number of features (i.e. 28 or more) a hit rate ranging from 25 to 66% was achieved [23], [26], [27]. In contrast, some studies reporting a low number of features (i.e. 7 or less) achieved hit rates above 85%, with two studies reporting a surprising 100% [29], [31], [33]. From these studies, it was not clear if the authors investigated a low number of features or if they investigated a large number of features but only reported the most significant ones. Even if reporting bias (a.k.a. selective reporting) should not be concluded from this finding, it should at least make us aware of the potential presence of this practice in our field. This practice could undermine the findings of future studies, making more difficult to converge to meaningful conclusions.

V. CONCLUSIONS

In conclusion, this paper demonstrated that there are high and significant interactions among sensor placement, task and feature category to assess the risk of falling. This systematic review provided a framework for future study design, highlighting dependences among those factors. In addition, the review generated a comprehensive inventory of the features reported so far from inertial sensors for fall risk assessment in older adults, summarizing their trends and whether these were found statistically significant or not in each study. The statistical analysis of those features demonstrated that the triad ‘angular velocity-walking-shins’ has shown more discriminative power between non-fallers and fallers than others. Finally, the meta-analysis demonstrated that 4 features resulted significantly different between non-fallers and fallers. However, most features were not included in the meta-analysis because they were not reported with sufficient homogeneity in at least 2 studies, suggesting that future studies are required to produce more evidence that allows to conduct a more comprehensive meta-analysis. Future studies should consider the evidence resulting from our review, in particular for: 1) the selection of the features to be further explored; 2) the sensor placement; and 3) the task used to assess the risk of falling. Those studies could also benefit from the adoption of some practices more common in clinical research, such as the definition of participant inclusion/exclusion criteria, inclusion of potential confounders in the analysis and ultimately the extensive publication of the full study protocol prior to the study conduction. These practices aim to reduce the risk of bias and confounding, thus giving more validity to the study.

ACKNOWLEDGMENTS

This study brought together existing data extracted from 13 original papers included in the review and cited in the reference section. These data are provided in full as supplementary material accompanying this paper.

The authors would like to thank Miss Pasqua Meccariello for her valuable contributions to the initial steps of this work, carried out during her time as visiting researcher at the School of Engineering, University of Warwick.

REFERENCES


