Sensor-based Fall Risk Assessment – an Expert ‘to go’

M. Marschollek¹; A. Rehwald²; K. H. Wolf³; M. Gietzelt³; G. Nemitz³; H. Meyer zu Schwabedissen³; R. Haux²

¹Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig – Institute of Technology and Hanover Medical School, Hanover, Germany; ²Institute of Technology and Hanover Medical School, Braunschweig, Germany; ³Braunschweig Medical Center, Department for Geriatric Medicine, Braunschweig, Germany

Keywords
Accidental falls, motor activity, assisted living facilities, sensors

Summary
Background: Falls are a predominant problem in our aging society, often leading to severe somatic and psychological consequences, and having an incidence of about 30% in the group of persons aged 65 years or above. In order to identify persons at risk, many assessment tools and tests have been developed, but most of these have to be conducted in a supervised setting and are dependent on an expert rater.

Objectives: The overall aim of our research work is to develop an objective and unobtrusive method to determine individual fall risk based on the use of motion sensor data. The aims of our work for this paper are to derive a fall risk model based on sensor data that may potentially be measured during typical activities of daily life (aim #1), and to evaluate the resulting model with data from a one-year follow-up study (aim #2).

Methods: A sample of n = 119 geriatric inpatients wore an accelerometer on the waist during a Timed ‘Up & Go’ test and a 20 m walk. Fifty patients were included in a one-year follow-up study, assessing fall events and scoring average physical activity at home in telephone interviews. The sensor data were processed to extract gait and dynamic balance parameters, from which four fall risk models – two classification trees and two logistic regression models – were computed: models CT#1 and SL#1 using accelerometer data only, models CT#2 and SL#2 including the physical activity score. The risk models were evaluated in a ten-times tenfold cross-validation procedure, calculating sensitivity (SENS), specificity (SPEC), positive and negative predictive values (PPV, NPV), classification accuracy, area under the curve (AUC) and the Brier score.

Results: Both classification trees show a fair to good performance (models CT#1/CT#2): SENS 74%/58%, SPEC 96%/82%, PPV 92%/74%, NPV 77%/82%, accuracy 80%/78%, AUC 0.83/0.87 and Brier scores 0.14/0.14. The logistic regression models (SL#1/SL#2) perform worse: SENS 42%/58%, SPEC 82%/78%, PPV 62%/65%, NPV 67%/72%, accuracy 65%/70%, AUC 0.65/0.72 and Brier scores 0.23/0.21.

Conclusions: Our results suggest that accelerometer data may be used to predict falls in an unsupervised setting. Furthermore, the parameters used for prediction are measurable with an unobtrusive sensor device during normal activities of daily living. These promising results have to be validated in a larger, long-term prospective trial.

1. Introduction

Falls are a predominant problem in our aging population. The incidence of falls within 12 months among persons aged 65 years or more has been reported to be 25.1% for men and 37.0% for women [1]. Although many falls do not have severe consequences, about 3–5% of them result in fractures [1, 2], and about 20% of fallers need medical intervention [2]. Apart from the physical consequences, which often lead to a lasting reduction or loss of mobility and independence, the so-called ‘post-fall syndrome’ as reported by Tinetti et al. is dreaded [3]: The fear of falling induced by a fall event leads to a vicious circle with less activity, leading to an increased fall risk and further falls, social isolation or other harmful consequences with regard to the person’s health status, e.g. a higher risk to suffer from a cardiovascular event. The annual costs of falls and their consequences in the U.S. have been estimated at more than $19 billion [4].

Driven by the enormous impact of falls among the elderly, many tools to identify persons with a high fall risk have been developed and evaluated extensively. Gates et al. review 29 different assessment tools [5], among them e.g. the prominent Performance-Oriented Mobility Assessment (POMA) by Tinetti [6], coming to the conclusion that no explicit recommendation for anyone test may be given. Among the assessment tools for inpatients, the St. Thomas Risk Assessment Tool in Falling Elderly Inpatients (STRATIFY) [7] and the Timed ‘Up & Go’ test (TUG) are frequently used [8]. In a systematic analysis of prospective studies with several tools, Oliver et al. conclude that even the best tools currently cannot identify a high percentage of fallers correctly [9]. Similar results have
been reported e.g. by Kim et al., who – using the Hendrich II Fall Risk Model [10] – reach a maximum specificity of 61.5% in a sample of non-geriatric patients [11].

The drawback of the mentioned assessment scales and tests however is that their use demands a certain time and effort, ranging from the simple TUG, which may take only one minute, to the complex POMA demanding up to ten minutes or more of a physiotherapist’s time. These tests have to be conducted in a supervised environment and may therefore suffer from influences such as the Hawthorne effect. The overall aim of our research work therefore is to develop an objective and unobtrusive method to determine individual fall risk based on the use of accelerometer sensor data. The aims of our work for this paper are:

● to derive fall risk models based on sensor data that may potentially be measured during typical activities of daily life (aim #1), and
● to evaluate the resulting models with data from a one-year follow-up study (aim #2).

2.1 Related Work

Only few studies using accelerometer sensor data for fall prediction have been published so far. Giansanti makes use of a wearable device incorporating three uniaxial accelerometers and three gyroscopes [12]. This device is worn on the lower back above the lumbar vertebra L5, and the study participants (n = 390) follow a fixed test protocol: standing with eyes open on level ground, standing on a foam cushion with eyes open and closed. After a classifier training phase with three groups, each with 30 participants, altogether 300 participants were classified into the groups ‘age < 65 years, no mobility impairment’, ‘age ≥ 65 years, no mobility impairment’, and ‘age ≥ 65 years, mobility impairment’ (Tinetti level 3). The author reports a sensitivity of ≥ 93% and a specificity of ≥ 93.9%, but does not report on positive and negative predictive values [12]. Laessøe et al. investigate fall risk in a one-year cohort study with n = 94 participants aged between 70 and 80 years, measuring e.g. gait cadence and gait variability with a triaxial accelerometer [13]. The authors conclude that it is not possible to distinguish between fallers and non-fallers. It has to be noted that persons with musculoskeletal diseases, pain during the performance of activities of daily living, a fall event within the last month, vestibular problems, cognitive impairment (Mini Mental State Examination score lower than 23 points), and those in need of care have been excluded from this study [13].

2. Methods

2.1 General Approach

With regard to our aim to assess relevant data unobtrusively during activities of daily life, we chose to measure motion while standing up and walking, thus obtaining information about a person’s gait and her or his dynamic balance. These data are the basis for the induction of classification models, using supervised machine-learning algorithms. As we aimed to evaluate our models in a one-year follow-up study and technically were – at that point in time – not able to assess activities of daily life with a sensor system over such a long period of time, we chose to use a validated activity questionnaire in addition to our sensor data.

2.2 Study Population and Test Procedure

In our study we included a population that is known to have the highest risk for falling: geriatric inpatients [14]. They were recruited at the Department for Geriatric Medicine of the Braunschweig Medical Center based on the following inclusion criteria:

● admission as an inpatient between April 24 and October 18, 2007
● ability to stand up and walk, and thus to take part in the sensor measurement
● written consent to take part in the follow-up interviews

Altogether n = 119 patients took part in the first part of the study, i.e. the measurement of standing up from a chair and walking for 20 m on an even floor in the physiotherapy department of the clinic [15]. N = 50 patients (37 women and 13 men) with a mean age of 81.3 years were included in the second part, the follow-up study. The reasons for dropout were: death (n = 17), untraceable (n = 13), no answer to the letter asking for written consent (n = 4), and several other problems such as deafness or cognitive impairment (n = 8). Four of the 50 motion data sets were unusable due to technical failures in the equipment (sensor malfunction, transmission interruption), leaving n = 46 sets for analysis. The study protocol has been approved by the Hanover Medical School ethics committee.

2.3 Sensor Equipment

The motion measurement was performed with a triaxial accelerometer sensor board (Freescale ZStar RD3152MM7260Q, 30 Hz) which was boxed and attached to the patient’s waist with a custom belt. This position was chosen on the one hand to make the device as unobtrusive as possible, possibly to be integrated into a belt buckle, and on the other hand to be close to the body’s center of gravity, thus allowing for reasonable measurement of gait parameters and balance characteristics. The data were transmitted to a laptop PC during the measurement and recorded for later processing.

2.4 Fall and Activity Assessment

In order to assess both fall events and physical activity of the enrollees during the one-year follow-up, deemed an important predictive factor for falls [16], we chose to conduct telephone interviews. The first part of the interview assessed the number of fall events and their consequences as proposed by the Prevention of Falls Network Europe (ProFaNE) consensus [17]. At first, the definition of a fall is read out to the patient, and then she or he is asked if such an event has occurred, how often, at what time, and if the fall was injurious [17]. The second
part of the interview consisted of a translated version of the modified Baecke questionnaire for older adults, modified by Voorips et al. [18]. It comprises ten questions and assesses common activities of daily life such as light or heavy housework, meal preparation and outdoor activities. It has been validated in several studies [19–21] and, amongst other things, gives a sum score of physical activity which was used in our study.

2.5 Accelerometer Data Processing and Parameter Extraction

Following a procedure of automatic self-calibration as described in [22] and an alignment of the sensor coordinate system in case of a tilted position on the wearer’s body, several processing steps were conducted in order to calculate the following parameters:
- kinetic energy
- pelvic sway along the transversal axis
- standard deviation of gait periodicity
- mean step duration
- step length
- number of steps during Timed “Up & Go” test

The algorithms used to extract the above-mentioned parameters have been described extensively in [15, 23]. In addition, we calculated spectral density distribution parameters of the accelerometric signal (0.25–4.0 Hz band pass filter). The method has been developed by the first author, is described in [24] and yields the following parameters:
- number of peaks
- frequency, width and relative prominence of the first peak
- frequency, width and relative prominence of the dominant peak
- number of peaks above a relative threshold

2.6 Classification Model Induction and Evaluation

To induce our classification models, we used two supervised machine-learning algorithms: the classification and regression tree (CART) algorithm introduced by Breiman et al. [25] and a logistic regression algorithm. The classification tree (CT) combines the advantages of good performance with intuitive understanding by the domain expert. Furthermore, classification rules, which are well-known among clinicians, may be extracted from the model and high-risk subgroups of a population may be identified [26]. As classification trees tend to be instable when small data sets are used for model induction [27], we also induced a logistic regression model (SL) from our data, which is less subject to fluctuations. As input for our model induction algorithms we used a binary classification attribute: fall events within the last year at home (yes/no) along with all of the above-mentioned parameters. Because we assessed long-term physical activity by means of a questionnaire, and not by long-term sensor monitoring, we induced two classification models with each induction algorithm, one with the activity parameter (models CT/SL#2) and one without it for comparison (models CT/SL#1). The classification tree models were induced using the software package SPSS 16.0 with a commercial CART implementation (parameters: minimum parent size = 5, minimum child size = 3, pruned), and the logistic regression models were induced using the open source toolkit WEKA (Waikato Environment for Knowledge Analysis, version 3.7.1, simple logistic algorithm; parameters: error on probabilities = true, heuristic stop = 50, maximum number of iterations for LogitBoost = 5000, use cross valida-
tion = true, no weight trimming). Prior to the induction of the logistic regression model, in order to exclude parameters with low information, we reduced their number by using a feature selection algorithm (wrapper subset evaluator, employing the simple logistic regression algorithm as described above).

The classification models were evaluated by means of a ten-times tenfold cross validation, a well-accepted procedure for small data sets [27]. For each model, sensitivity (SENS), specificity (SPEC), positive and negative predictive values (PPV/NPV) and classification accuracy were calculated and rounded to full numbers. Apart from these, we computed the area under the curve (AUC) – using WEKA for the logistic regression models and SPSS for the classification trees – as well as the Brier score, which is the mean squared difference between the outcome and the predicted probabilities for each instance [28]. In addition, receiver operating curves (ROC) were plotted using SPSS for models CT#2 and SL#2 (both including long-term physical activity).

3. Results

Figure 1 shows the decision tree induced with the sensor-based parameters (model CT#1), and Figure 2 shows the tree including the long-term activity parameter (model CT#2) assessed by the modified Baecke questionnaire. In model #1 (Fig. 1), the first split is performed on the width of the dominant peak, followed by the number of peaks and finally the frequency of the dominant peak. If long-term activity is added (model CT#2, Fig. 2), the first split is done on this attribute, followed by further splits on number of peaks and width of the dominant peak on the same level. The latter correspond with model CT#1, albeit the thresholds are slightly different (≤7 resp. ≤6 for number of peaks, and 1.04 vs. 1.17 for width of the dominant peak).

The classification performance of both decision trees is shown in Tables 1 (model CT#1) and 2 (model CT#2). The overall classification accuracy is almost equal in both models, as are the AUC (0.83 resp. 0.87) and the negative predictive values (77% vs. 82%). The sensitivity, however, is higher (74% vs. 58%) if long-term activity is incorporated. The specificity, in turn, falls from 96% to 82%, as does the positive predictive value (92% to 74%). The Brier score is equal for CT#1 and CT#2 (0.14).

\[
-0.41 \times (\text{number of peaks}) - 2.56 \times (\text{width of dominant peak}) + 5.33 \quad (1)
\]

\[
-0.52 \times (\text{number of peaks}) - 2.39 \times (\text{width of dominant peak}) - 0.06 \times (\text{physical activity score}) + 6.64 \quad (2)
\]

The logistic regression models SL#1 (Table 3, Eq. 1) and SL#2 (Table 4, Eq. 2) show overall worse classification performances in comparison with the classification trees, with Brier scores of 0.23 resp. 0.21 and an AUC of 0.65 resp. 0.72. Model SL#2 including the physical activity parameter, distinguishes between fallers and non-fallers slightly better than SL#1 (accuracy 70%/65%, SENS 58%/42%, SPEC 78%/82%, PPV 65%/62%, NPV 72%/67%). The parameters used match those of the classification tree models (number of peaks, width of dominant peak, physical activity score).

Figures 3 and 4 show ROC curves of models CT#2 and SL#2, both including the long-term physical activity parameter.

4. Discussion

Our results, based on a one-year follow-up study, indicate that the derived fall risk models – including the long-term activity parameter – work well (aim #2), and thus that accelerometric sensor data may provide relevant information for the assessment of individual fall risk in a sample of...
Table 1  Classification results and confusion matrix for classification tree model CT#1

<table>
<thead>
<tr>
<th>model CT#1 (classification tree) – sensor-based parameters</th>
<th>confusion matrix fall within one year</th>
</tr>
</thead>
<tbody>
<tr>
<td>classification accuracy</td>
<td>80%</td>
</tr>
<tr>
<td>sensitivity</td>
<td>58%</td>
</tr>
<tr>
<td>specificity</td>
<td>96%</td>
</tr>
<tr>
<td>negative predictive value</td>
<td>77%</td>
</tr>
<tr>
<td>positive predictive value</td>
<td>92%</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.14</td>
</tr>
<tr>
<td>AUC</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Table 2  Classification results and confusion matrix for classification tree model CT#2 (including long-term physical activity)

<table>
<thead>
<tr>
<th>model CT#2 (classification tree) – sensor-based parameters + activity</th>
<th>confusion matrix fall within one year</th>
</tr>
</thead>
<tbody>
<tr>
<td>classification accuracy</td>
<td>78%</td>
</tr>
<tr>
<td>sensitivity</td>
<td>74%</td>
</tr>
<tr>
<td>specificity</td>
<td>82%</td>
</tr>
<tr>
<td>negative predictive value</td>
<td>82%</td>
</tr>
<tr>
<td>positive predictive value</td>
<td>74%</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.14</td>
</tr>
<tr>
<td>AUC</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Table 3  Classification results and confusion matrix for logistic regression model SL#1

<table>
<thead>
<tr>
<th>model SL#1 (logistic regression) – sensor-based parameters</th>
<th>confusion matrix fall within one year</th>
</tr>
</thead>
<tbody>
<tr>
<td>classification accuracy</td>
<td>65%</td>
</tr>
<tr>
<td>sensitivity</td>
<td>42%</td>
</tr>
<tr>
<td>specificity</td>
<td>82%</td>
</tr>
<tr>
<td>negative predictive value</td>
<td>67%</td>
</tr>
<tr>
<td>positive predictive value</td>
<td>62%</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.23</td>
</tr>
<tr>
<td>AUC</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Geriatric patients. In comparison with an earlier study by the authors, the aim of which was to identify patients who suffer from fall events during their inpatient stay, the results obtained here are much better [29], likely due to the longer period of observation (one year). Model CT#1 (Fig. 1) shows that only spectral density parameters are used to split the data set. A wider dominant spectral density peak (first split in Fig. 1) is associated with a lower fall risk (8% for persons with a wider peak vs. 53% for persons with a narrower peak), as was a frequency of less than 1.6 Hz of this peak (33% vs. 100%). While we may conclude that faster gait speed is associated with a lower fall risk, confirmed by numerous studies (e.g., [30–32]), we did not expect to find a higher variability in stride time to be associated with a lower fall risk, contradicting e.g. the results of Hausdorff et al. [33, 34]. This finding, however, might be attributed to the fact that several inpatients in our study sample were recovering from fractures of the lower extremities, temporarily affecting the normal gait pattern in a way that steps with the affected leg were taken more slowly and gently.

The overall activity parameter shows a clear association of a high activity level (>13 points on the modified Baecke) with a low fall risk (15% vs. 61%, Fig. 2, first split). Concerning this point, literature shows divergent results: Chan et al. report that in men aged 65 or older (prospective study, n=5995) a higher level of household activities is associated with a higher number of falls [35]; Stevens et al. state that intensive physical activity is only a risk factor for fractures induced by falls, if a person’s ability to perform activities of daily living (ADL) is reduced (odds ratio 3.2) [16]. Further research is necessary to confirm the role of a person’s physical activity level with regard to individual fall risk.

The parameters that appear in the classification tree models are derived during phases of normal walking and thus are not dependent on a supervised (e.g., POMA or TUG) or self-administered test setting [36] and therefore may be obtained during common daily routine by using an unobtrusive, small wearable sensor device which can be worn over extended periods of time. Previous studies suggest that wearing small accelerometer devices is well-accepted by older persons, even if these are worn for weeks [37–39].

In comparison to the classification tree models (AUC 0.83 resp. 0.87), both logistic regression models show a lower classification performance (Tables 3 and 4), a finding that is also reflected by the ROC curves shown in Figs. 3 (model CT#2) and 4 (model SL#2). SL#2 performs slightly better than SL#1 (AUC 0.72 resp. 0.65), thus confirming the relevance of our long-term activity parameter.

In summary, our classification results show a good performance, especially if compared to a conventional fall risk assessment test that was used in parallel: If the patients’ STRATIFY score was employed to predict falls within the following year, using a cut-off value of ≥2 points.
as suggested by the developers [7], a good sensitivity of 79% would go along with poor results for specificity (26%), positive and negative predictive values (43% resp. 63%) and classification accuracy (48%). It has to be noted, however, that this score has been constructed for inpatients and not for fall prediction in a domestic environment. Nevertheless, these findings suggest that our objective ‘electronic expert rater’ yields predictive results that not only match those of conventional methods, but also sheds light on a domain that was so far out of reach for assessment.

Our classification results are slightly worse than those achieved by Giansanti [12], but were specifically aimed to identify fall risk rather than mobility impairment, and are not dependent on a fixed test protocol. Our results contradict those of Laessoe et al. [13], which may be explicable by the differing sample characteristics, with our sample of geriatric inpatients having a high fall risk and a high prevalence of comorbidities.

In conclusion, we may state that the study results on hand show a good performance in classifying fallers and non-fallers within a year of follow-up, especially when taking into account that our parameters can be measured in an unsupervised, truly pervasive [40–42], setting during normal activities of daily living, and thus – e.g. combined with a fall detection algorithm and an emergency communication system – may be regarded as part of a future smart home environment [43, 44].

### 4.1 Limitations

The sample size of our follow-up study is very small, thus limiting the generalizability of our classification results. Furthermore, our approach only considers parameters which are associated with mobility impairments, ignoring the fact that numerous other risk parameters exist which may not or merely indirectly influence mobility, e.g. impaired vision, urinary incontinence, cognitive impairment, etc. An extensive review can be found in [45]. It is likely that the predictive performance of our models would be enhanced if we incorporated these parameters into our models.

Furthermore, despite the assumption that wearing a small sensor device is less intrusive than a conventional expert fall risk assessment or the supervision of a test procedure, we cannot rule out that our approach also induces a Hawthorne effect.

From an economic perspective, it remains unclear if the prediction results are good enough to justify the implementation of costly preventive measures for the false positives. A cost-benefit analysis should be conducted, comparing direct and indirect costs of fall events with those of preventive measures. Furthermore, despite promising first results (e.g. [38]), the patients’ acceptance of our long-term sensor-based approach has to be investigated, e.g. using the Sensor Acceptance Model [46].

Finally, our models have to be tested in a large, prospective clinical trial to evaluate their true predictive potential.

---

**Table 4** Classification results and confusion matrix for logistic regression model SL#2 (including long-term physical activity)

<table>
<thead>
<tr>
<th>model SL#2 (logistic regression) – sensor-based parameters + activity</th>
<th>confusion matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>classification accuracy</td>
<td>70%</td>
</tr>
<tr>
<td>sensitivity</td>
<td>58%</td>
</tr>
<tr>
<td>specificity</td>
<td>78%</td>
</tr>
<tr>
<td>negative predictive value</td>
<td>72%</td>
</tr>
<tr>
<td>positive predictive value</td>
<td>65%</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.21</td>
</tr>
<tr>
<td>AUC</td>
<td>0.72</td>
</tr>
</tbody>
</table>

**Table 4** Classification results and confusion matrix for logistic regression model SL#2 (including long-term physical activity)

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>no</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. yes</td>
<td>11</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>pred. no</td>
<td>8</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td>sum</td>
<td>19</td>
<td>27</td>
<td>46</td>
</tr>
</tbody>
</table>

**References**

2. Kannus P, Sievanen H, Palvanen M, Jarvinen T, Parkkari J. Prevention of falls and consequent in-