Automatic Fall Risk Estimation using the Nintendo Wii Balance Board

Gert Mertes$^{1,2,3}$, Greet Baldewijns$^{1,2,3}$, Pieter-Jan Dingene$^{1}$, Tom Croonenborghs$^{1,4}$ and Bart Vanrumste$^{1,2,3}$

$^1$KU Leuven, Campus Geel, AdvISe, Geel, Belgium
$^2$KU Leuven, ESAT-STADIUS, Leuven, Belgium
$^3$iMinds Medical Information Technology Department, Ghent, Belgium
$^4$KU Leuven, Department of Computer Science, DTAI, Heverlee, Belgium

Keywords: Wii Balance Board, Center of Pressure, Fall Risk Classification, Machine Learning, Support Vector Machine, K-Nearest Neighbours.

Abstract: In this paper, a tool to assess a person’s fall risk with the Nintendo Wii Balance Board based on Center of Pressure (CoP) recordings is presented. Support Vector Machine and K-Nearest Neighbours classifiers are used to distinguish between people who experienced a fall in the past twelve months and those who have not. The classifiers are trained using data recorded from 39 people containing a mix of students and elderly. Validation is done using 10-fold cross-validation and the classifiers are also validated against additional data recorded from 12 elderly. A cross-validated average accuracy of 96.49% ± 4.02 is achieved with the SVM classifier with radial basis function kernel and 95.72% ± 1.48 is achieved with the KNN classifier with $k = 4$. Validation against the additional dataset of 12 elderly results in a maximum accuracy of 76.6% with the linear SVM.

1 INTRODUCTION

A third of all older persons aged 65 or older fall at least once a year (Milisen et al., 2004), (Robertson and Gillespie, 2013). Approximately 10% of these fall incidents result in serious injuries. Moreover, 7% of emergency room visits are due to fall incidents (Tinetti, 2003).

Fall incidents, however, not only result in physiological injuries but also have an impact on the psychological health of the person that fell. After all, fall incidents can lead to fear of future falls which in turn can cause the elderly to move less and therefore spend more time indoors (Tinetti and Williams, 1998), (Milisen et al., 2004), (Noury et al., 2008). The resulting isolation and lack of exercise can in turn reduce the muscle strength of the older person which causes an increase of the fall risk. Preventing fall incidents would therefore not only contribute to reducing fall related health care costs, but would also greatly contribute to the quality of life of older persons.

When an elevated fall risk can be detected at an early stage, preventive measures can be taken to reduce this risk and hence reduce the number of fall incidents. These possible measures include an adaptation of the medication regime or the implementation of an exercise programme. The home environment of the older person could furthermore be adapted to further reduce the risk (e.g.: removing loose carpets, installing new light fixtures, etc.).

To date several methods to assess the fall risk of a person such as the Timed-get-Up-and-Go (TUG) test (Martin, 2011), (Podsiadlo and Richardson, 1991) and the Tinetti Mobility test (Tinetti et al., 1986) already exist.

During the TUG test, the elderly is asked to rise from a chair, walk three meters, turn around, return to the chair and sit down. The fall risk is subsequently quantified using the manually recorded time to complete the test combined with the observations of the health care professional (e.g.: visible shuffling, posture during the walk, etc.). Typically, the elderly is considered at risk when the time needed to complete the test exceeds 14 seconds or if the person has a very unstable gait pattern (Large et al., 2006).

The Tinetti test in turn evaluates a person’s balance and gait using a series of exercises. Balance is evaluated using nine exercises including, among others, standing up from a chair and the ability to withstand a nudge to the chest. Gait is evaluated by ob-
serving the older person’s step length, step symmetry, walking stance, etc. Each exercise is observed and subsequently graded on a scale of 0 to 2 after which the grades are summed into a final score. A score below a certain threshold (19 out of 28 points) indicates a balance or gait problem (Tinetti et al., 1986).

Other tests include the Four Test Balance Scale (Gardner et al., 2001), the Functional Reach test as described by (Duncan et al., 1992), the Five Times Sit to Stand Test proposed by (Guralnik et al., 1995), the Berg Balance Scale developed by (Berg et al., 1989) and the Balance Evaluation Systems Test by (Horak et al., 2009). While providing insight in a person’s balance, all of the previously mentioned tests require specialised personnel to perform the test. Furthermore, these tests are mostly administered when a fall incident has already taken place and are therefore not always useful in preventing future falls.

The aim of the presented study is therefore the development of an easy to use tool which can detect an elevated fall risk at an early stage without the intervention of a health care professional. The tool will make it possible for informal care givers (e.g.: children or neighbours of the older person) to have an early assessment of the fall risk of their care receiver. If the tool shows an elevated fall risk, a more in depth assessment of the causes of this detected risk should be initiated by a health care professional which can be followed by the installation of preventive measures.

For this purpose, a system using a force plate is presented. In recent years, force plates have gained popularity as a way to measure a person’s balance. Force plates offer the advantage of being able to quickly measure a person’s balance without requiring specialised personnel. (Melzer et al., 2004) found that it is possible to discriminate between elderly who have recently experienced falls and non-falling elderly persons using parameters of mediolateral (ML) sway extracted from the Center of Pressure (CoP) trajectory. (Piirtola and Era, 2006) reviewed nine prospective studies where force platform measurements have been used as predictors of falls among elderly people. They found that in five studies, fall-related outcomes were associated with features measured with the force platform. These systems, however, use expensive force plates which are not accessible for regular care givers.

The system proposed in this paper uses a cheap alternative for these force plates namely the Nintendo Wii Balance Board (Wii BB). The Wii BB can be used by care givers to quickly and automatically estimate if a person is at risk of falling and is, due to the popularity of the Wii console, currently already present in a lot of homes. Compared to current classification methods, e.g. the TUG, and previous studies, our system does not require trained personnel, no manual interpretation of the data is required and no expensive equipment needs to be present.

Fifty-one participants were asked to stand on the Wii BB for 40 seconds while the pressure on the sensors was sampled. The pressure measured on the force plate is used to calculate the CoP trajectory. Features are subsequently extracted from the CoP trajectories and used to automatically classify if a person is at risk for falling.

In the remainder of the paper the data acquisition, preprocessing, feature extraction and classification methods using the data of the participants is presented in section 2. This is followed by a results section in which the classification results of a Support Vector Machine (SVM) and k-Nearest Neighbours (KNN) classifier are presented. These results are subsequently discussed in depth in section 4. Finally we conclude that a good classification can be reached using both the SVM as the KNN classifier.

2 METHODS

2.1 Dataset

2.1.1 Wii Balance Board

The Wii Balance Board is a game controller developed by Nintendo and introduced in 2007 for the Wii video game console. It ships with the game Wii Fit in which users are required to do exercises with or around the Wii Balance Board, such as Yoga or Aerobics. The game provides visual feedback as well as track the user’s performance over time.

The board is shaped like a regular household body scale. It contains four pressure sensors, one in each corner as illustrated by figure 1, to measure the forces exited by the user on the board. With these 4 forces, the CoP can be calculated using the momentum balance equations of the board. Communication is han-
died by a wireless Bluetooth link. In order to capture the raw sensor data with MATLAB, we used the open-source WiiLab library developed by (Brindza et al., 2009).

2.1.2 Participants

The dataset used for training purposes consists of recordings from 39 people. Participants were sorted in three groups: students, elderly without fall history and elderly with fall history. Participant characteristics are shown in table 1.

Sixteen students aged 18 to 22 were recruited on campus to act as control group. They declared not to have any postural balance issues nor experienced a fall due to balance issues.

The group elderly without fall risk consists of 15 living at home elderly aged 59 to 79 who declared not to have fallen in the past 12 months. They were recruited by the researchers among family and friends.

Data from eight elderly people was recorded at a nursing home. These people declared to have fallen at least once in the past 12 months; 6 of which fell at least twice during this time period.

Furthermore, data from an additional 12 elderly, 11 females and 1 male, was recorded at a second nursing home. Three persons, with an average age of 84.89 ± 6.88, declared to have fallen at least once in the past 12 months. The remaining eight elderly, with an average age of 83.67 ± 4.93, had no history of falls. This additional dataset is used to validate the trained classifiers.

The medical ethical commission of the KU Leuven university hospital approved this study and all participants gave their written informed consent.

2.1.3 Procedure

Participants were asked to stand on the balance board and perform standardised exercises. The exercises required the participant to stay as rigid as possible with their arms positioned next to their body and looking straight ahead.

In the first exercise the participant places the feet on predefined outlines located on the balance board, as shown in figure 1. This stance is hereby referred to as the **rigid stance**. In the second exercise the participant was asked to stand in the center of the balance board with feet and knees pressed together, referred to as the **narrow stance**.

According to (Melzer et al., 2010), visual feedback influences postural sway. Participants were therefore asked to perform the exercises with both eyes open and eyes closed. The narrow stance exercise with closed eyes proved to be impossible to main-

tain for the duration of the measurement for all 8 of the elderly people who had previously fallen and was therefore dropped from the study.

Each participant thus performed three different exercises where each exercise was performed three times, resulting in nine measurements per participant. Each measurement lasted 40 seconds, during which the data from the 4 pressure sensors was sampled at a frequency of 64 Hz.

In the remainder of this paper, each exercise is addressed using the name of the stance and whether the eyes were open or closed, i.e.: **Rigid Open** (RO), **Rigid Closed** (RC) and **Narrow Open** (NO).

2.2 Pre-processing

Before extracting the CoP trajectory, the raw data is pre-processed. The first and last 5 seconds of each measurement are trimmed in order to discard any transient effects, resulting in measurements of 30 seconds. The trimmed data is then filtered using an 8th order Butterworth filter with a cut-off frequency of 10 Hz to remove any high frequency noise, which (Salavati et al., 2009) identified as the optimal cut-off frequency for CoP measurements.

The CoP trajectory in the ML and anterior-posterior (AP) directions is then calculated using formulas 1 and 2, derived from the momentum balance equations of the board. These formulas contain the forces in each corner of the balance board and its width and length (see figure 1).

\[
COP_{ML} = \frac{F_{TR} + F_{BR} - F_{TL} - F_{BL} L}{F_{TR} + F_{BR} + F_{TL} + F_{BL} 2} \quad (1)
\]

\[
COP_{AP} = \frac{F_{TL} + F_{TR} - F_{BL} - F_{BR} W}{F_{TR} + F_{BR} + F_{TL} + F_{BL} 2} \quad (2)
\]

2.3 Feature Extraction

From the CoP trajectories, we extract the mean velocity, standard deviation of the velocity and the standard deviation of the amplitude in both the ML and AP directions. Our choice of features is based on the findings of (Piirtola and Era, 2006) and (Melzer et al., 2004) which show that both velocity and amplitude can be used as an indicator of falls. While other features such as the 95% confidence ellipse area have also proven successful, we hypothesised, based on the results of the before mentioned studies, that two-dimensional features of velocity and amplitude would allow for a classifier with a sufficiently high accuracy.
Table 1: Characteristics of the participants included in the training data. [mean ± std.dev.]

<table>
<thead>
<tr>
<th></th>
<th>Students (N = 16)</th>
<th>Elderly non-fallers (N = 15)</th>
<th>Elderly fallers (N = 8)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>19.53 ± 1.77</td>
<td>71.87 ± 5.80</td>
<td>85.40 ± 5.83</td>
</tr>
<tr>
<td><strong>Gender [f/m]</strong></td>
<td>4/11</td>
<td>12/3</td>
<td>7/1</td>
</tr>
<tr>
<td><strong>Height [cm]</strong></td>
<td>175 ± 7</td>
<td>160 ± 8</td>
<td>152 ± 11</td>
</tr>
<tr>
<td><strong>Weight [kg]</strong></td>
<td>86.85 ± 12.55</td>
<td>71.87 ± 5.8</td>
<td>64.40 ± 7.98</td>
</tr>
</tbody>
</table>

2.4 Classification

For each measurement per exercise and type of feature, a SVM classifier is trained. Two different types of SVM classifiers are evaluated, one with a linear and one with a Gaussian radial basis function (rbf) kernel. KNN classifiers, with $k = 3$ and $k = 4$, are also trained and validated.

Each exercise was performed three times, resulting in three sets of classifiers per exercise per feature (a set being 2 SVM and 2 KNN classifiers).

The data recorded from 39 people containing 16 students, 15 elderly without and 8 elderly with fall history, as described in section 2.1.2, is used to train the classifiers. All classifiers are of the binary variant (i.e.: they can only distinguish between two classes). Data from students and elderly without fall history is grouped together and labelled as the "non-fallers" class, while the elderly with fall history are labelled as the "fallers" class.

2.5 Validation

Initial validation of the classifiers is done using 10-fold cross validation, resulting in accuracy scores for each classifier. Classifier and kernel parameters are then tuned to achieve the highest possible accuracy. It is worth noticing that the classifiers are trained with one measurement per exercise per participant. This prevents multiple measurements of the same person to be in different folds which would bias the results.

The validation dataset described in 2.1.2 of 12 additional elderly persons is used to validate the trained SVM classifiers. Data from this dataset is automatically labelled by the SVM classifier and confusion matrices are created from the results. This is done for each exercise separately. The choice to only validate the SVM classifier is based on the achieved results in combination with the advantages of the SVM as compared to the KNN classifier, which are outlined in section 4.

The validation dataset was recorded in a different nursing home with no connection between the two groups of participants.

3 RESULTS

Table 2 shows the average cross-validated accuracy of the classifiers per exercise together with the standard deviation. The three measurements per exercise are averaged to produce this accuracy. The highest accuracy of 96.49% ± 4.02 is achieved with the rbf SVM using the NO exercise and standard deviation of the velocity as feature. The difference between features, exercises and classifiers seems to be statistically irrelevant, but we can see that the standard deviation of the amplitude performs the worst of the three features.

Figure 2 illustrates the separability of the training data. The training data of the first linear SVM RO classifier (RO1) is shown. Each point represents a participant’s velocity’s standard deviation for the first RO exercise. The x and y axis represent the velocity in the ML and AP directions respectively. The support vectors are circled and the linear hyperplane is visualised.

Figure 3 shows the effect of the length of the measurement on the accuracy of the SVM classifier. Illustrated here is the average accuracy of the SVM classifier. As mentioned in section 2.5, only the linear SVM is used.

Table 4 shows the confusion matrices of the validation experiment with the validation dataset. As mentioned in section 2.5, only the linear SVM is
Table 2: Average cross-validated accuracy [%] of the classifiers for each exercise and feature. [mean ± std.dev.]

<table>
<thead>
<tr>
<th>Feature</th>
<th>Exercise</th>
<th>SVM (lin)</th>
<th>SVM (rbf)</th>
<th>KNN (3)</th>
<th>KNN (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>std. dev. velocity</td>
<td>RO</td>
<td>94.87 ± 2.57</td>
<td>92.31 ± 2.57</td>
<td>94.01 ± 3.92</td>
<td>95.72 ± 1.48</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>94.87 ± 2.57</td>
<td>91.45 ± 3.92</td>
<td>91.40 ± 2.88</td>
<td>93.09 ± 1.54</td>
</tr>
<tr>
<td></td>
<td>NO</td>
<td>92.10 ± 2.63</td>
<td>96.49 ± 4.02</td>
<td>92.98 ± 1.52</td>
<td>93.85 ± 3.04</td>
</tr>
<tr>
<td>mean velocity</td>
<td>RO</td>
<td>95.72 ± 3.92</td>
<td>94.87 ± 2.57</td>
<td>94.01 ± 2.96</td>
<td>94.87 ± 4.45</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>93.11 ± 1.40</td>
<td>93.97 ± 1.45</td>
<td>91.38 ± 1.42</td>
<td>93.97 ± 1.45</td>
</tr>
<tr>
<td></td>
<td>NO</td>
<td>92.98 ± 1.52</td>
<td>94.73 ± 2.63</td>
<td>92.10 ± 0.00</td>
<td>93.85 ± 3.04</td>
</tr>
<tr>
<td>std. dev. amplitude</td>
<td>RO</td>
<td>86.32 ± 2.96</td>
<td>83.75 ± 3.92</td>
<td>82.05 ± 5.13</td>
<td>84.61 ± 6.78</td>
</tr>
<tr>
<td></td>
<td>RC</td>
<td>87.04 ± 2.77</td>
<td>86.19 ± 3.08</td>
<td>88.75 ± 4.13</td>
<td>89.60 ± 4.67</td>
</tr>
<tr>
<td></td>
<td>NO</td>
<td>84.30 ± 0.15</td>
<td>86.84 ± 2.62</td>
<td>83.42 ± 6.17</td>
<td>83.54 ± 4.11</td>
</tr>
</tbody>
</table>

Table 3: Average cross-validated accuracy [%] of the classifiers per feature, averaged across the exercises. [mean ± std.dev.]

<table>
<thead>
<tr>
<th>Feature</th>
<th>SVM (lin)</th>
<th>SVM (rbf)</th>
<th>KNN (3)</th>
<th>KNN (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>std. dev. velocity</td>
<td>93.95 ± 1.60</td>
<td>93.41 ± 2.70</td>
<td>92.80 ± 1.31</td>
<td>94.22 ± 1.35</td>
</tr>
<tr>
<td>mean velocity</td>
<td>93.94 ± 1.55</td>
<td>94.52 ± 0.49</td>
<td>92.50 ± 1.36</td>
<td>94.23 ± 0.56</td>
</tr>
<tr>
<td>std. dev. amplitude</td>
<td>85.89 ± 1.42</td>
<td>85.59 ± 1.63</td>
<td>84.74 ± 3.54</td>
<td>85.92 ± 3.24</td>
</tr>
</tbody>
</table>

Figure 3: Average cross-validated accuracy of the linear SVM classifier plotted against measurement length. The dotted lines are the upper and lower limits of the standard deviation. Exercise: RO, feature: std. dev. velocity.

validated. Each matrix contains the summed results of the three measurements per exercise, i.e.: the true and false negatives and positives of the measurements were added together per exercise. To recap, each person performed three different exercises, where each exercise was executed three times. For the RO and RC exercise this results in 3 measurements x 12 persons = 36 results. The NO confusion matrix contains a total of 31 results. This is due to the fact that several participants in the validation dataset failed to execute this exercise.

4 DISCUSSION

The standard deviation of the velocity and the mean velocity perform almost equally well as features, with only a marginal difference between the two. Table 3 shows the accuracy of each classifier averaged across all exercises. The linear SVM performs best with the standard deviation of the velocity as feature, while the SVM with rbf kernel performs best using the mean velocity. The standard deviation of the amplitude scores the lowest. In section 2.3 it was hypothesised that our choice of features would result in a sufficiently high accuracy, which is confirmed by the results.

The rbf SVM offers the highest average accuracy when using the NO exercise and standard deviation of the velocity, but scores lower than the linear SVM with the other two exercises. The KNN classifier scores highest with all exercises and features when
While the KNN classifiers score almost equally high as compared to the SVM classifiers, the former is prone to be the slowest of the two and require more memory when a lot of data points are present since it is a form of lazy learning (Atkeson et al., 1997).

While previous studies such as (Melzer et al., 2010) suggest that the rigid stance is not suited for measurements regarding fall related indicators, our results show otherwise. As seen in table 2, all three of the exercises result in roughly the same accuracy for each classifier and feature. While the highest average accuracy was achieved with the rbf SVM and NO exercise, this exercise was also the most difficult to carry out by the elderly participants. The RO exercise is more suited to be carried out while no specialised personnel is nearby without a significant loss in accuracy.

In figure 2 we can see that the training data is completely separable by the hyperplane. A cluster of non-fallers can easily be distinguished in the lower left corner, which is an area of low postural sway. This group contains all 16 student participants. Fallers are less grouped together and located more towards the top right corner.

As seen in figure 3, the length of the measurement has little effect on the cross-validated accuracy of the SVM classifier. It is only below 20 seconds that a decline in accuracy is observed. This loss of accuracy can be explained due to the fact that in a limited time frame the CoP trajectory might not change enough in order to extract meaningful features. Furthermore, trimming the measurement length was done by removing an equal amount of data points at the start and end of the measurement. This might explain the increase in accuracy at the 20 second mark. We found that at the start of the measurements, several participants exhibited anticipation effects such as increased stiffening of the muscles. In the same manner, participants became tired or anxious towards the end. Trimming these effects can have a positive influence on the accuracy, as long as a minimum measurement time of 20 seconds is kept. The ideal measurement time is thus located between 20 and 40 seconds.

The confusion matrices in table 4 show that the SVM model is capable of classifying all fallers with the RO exercise. While the RO and RC exercise offer the same amount of false positives and true negatives, the RC exercise has two false negatives as compared to zero for the RO exercise. A false negative in this context weighs more than the false positive. When a person without fall risk is classified incorrectly (false positive), a follow up by a doctor or care giver will give exclusion, whereas an incorrect classification for a person with actual fall risk (false negative) might give that person a false sense of security. The NO exercise performs the lesser of the three. Note that N = 31 because several participants failed to execute this exercise. This is because the narrow stance is difficult to maintain for elderly adults. The NO exercise was also recorded as the last of the three. While participants were able to take a short break between exercises, this still introduced a bias in the results.

The high amount of false positives for all three exercises is due to the fact that several non-fallers in the validation dataset had underlying conditions that influence postural balance. Two non-fallers are systematically classified as fallers and account for six of the total amount of false positives per exercise. While these people did not fall in the last year, this wrongful classification can be attributed to the fact that they had hip-prostheses. Excluding these elderly would bring the number of false positives down to seven for the RO and RC exercise. After exclusion, the RO exercise offers the highest accuracy of 76.6%. False positives may also be attributed to the fact that the average age of non-fallers in the validation dataset is higher as compared to the training dataset.

5 CONCLUSION AND FUTURE WORK

It is possible to accurately distinguish between elderly who have recently experienced falls and non-falling elderly persons using the Nintendo Wii Balance board in conjunction with machine learning algorithms. Both the SVM and KNN classifier offer good performance with our current dataset.

A few points of improvement, however, remain. Firstly, the recorded dataset is relatively small and contains data from a very specific group of people. With more data, it may be possible to further increase the accuracy of the classifiers and reduce the amount of false positives. Secondly, the algorithm may be extended to detect different pathologies such as Parkinson’s disease. This would also require more data. While a medical diagnosis would still be required, the tool could be used as a preliminary indicator. Lastly, it would be interesting to see if the algorithm can give an indication of the severity of the fall risk instead of a binary classification.

Nevertheless, our results indicate that the Nintendo Wii Balance board can be a viable and cheap alternative to pressure plates for the detection of fall risk in elderly persons.
ACKNOWLEDGEMENTS

This work is funded by the iMinds FallRisk project. The iMinds FallRisk project is co-funded by iMinds (Interdisciplinary Institute for Technology), a research institute founded by the Flemish Government. Companies and organisations involved in the project are COMmeto, Televis Healthcare, TP Vision, Verhaert and Wit-Gele Kruis Limburg, with project support of IWT.

The authors would furthermore like to thank the nursing homes, students and elderly who participated in this study.

REFERENCES


