A Learning System for Automatic Berg Balance Scale Score Estimation

Davide Bacciu[a], Stefano Chessa[a], Claudio Gallicchio[a], Alessio Micheli[a], Luca Pedrelli[a], Erina Ferro[b], Luigi Fortunati[b], Davide La Rosa[b], Filippo Palumbo[b], Federico Vozzi[c], Oberdan Parodi[c]

[a]*Department of Computer Science, University of Pisa, Largo B. Pontecorvo 3, Pisa, Italy.*
[b]*Institute of Information Science and Technologies "A. Faedo", National Research Council, ISTI-CNR Pisa, Via Moruzzi 1, Pisa, Italy.*
[c]*Institute of Clinical Physiology - National Research Council, IFC-CNR Pisa, Via Moruzzi 1, Pisa, Italy.*

Abstract

The objective of this work is the development of a learning system for the automatic assessment of balance abilities in elderly people. The system is based on estimating the Berg Balance Scale (BBS) score from the stream of sensor data gathered by a Wii Balance Board. The scientific challenge tackled by our investigation is to assess the feasibility of exploiting the richness of the temporal signals gathered by the balance board for inferring the complete BBS score based on data from a single BBS exercise.

The relation between the data collected by the balance board and the BBS score is inferred by neural networks for temporal data, modeled in particular as Echo State Networks within the Reservoir Computing (RC) paradigm, as a result of a comprehensive comparison among different learning models. The proposed system results to be able to estimate the complete BBS score directly from temporal data on exercise #10 of the BBS test, with ≈ 10 seconds of duration. Experimental results on real-world data show an absolute error below 4 BBS score points (i.e. below the 7 % of the whole BBS range), resulting in a favorable trade-off between predictive performance and user's required time with respect to previous works in literature. Results achieved by RC models compare well also with respect to different related learning models.

Overall, the proposed system puts forward as an effective tool for an accurate automated assessment of balance abilities in the elderly and it is characterized by being unobtrusive, easy to use and suitable for autonomous usage.

Keywords: Balance Assessment, Reservoir Computing, Echo State Network, Learning with Temporal Data, Berg Balance Scale.

1. Introduction

All European countries are experiencing aging of their populations, with a decrease in the number of people of working age per retiree. By 2050, an estimated 35% of the European population will be over the age of 60, compared to 20% in 2005. Health trends among older people are mixed: severe disability is declining in some countries but increasing in others, while mild disabilities and chronic diseases are generally increasing. The aging process is characterized by a constant decline of body functions and is frequently associated to a series of impairments involving reduction in mobility and cognitive decline [1]: these aspects work synergistically increasing the risk of falls. These, and related injuries, represent one of the major causes of morbidity/mortality [2] and decline of elderlies' quality of life [3, 4]. Commonly, falls are considered the result of multifactorial elements working synergistically together, such as vision problems, lower limb weakness, altered mobility and somatosensory function alterations [5, 6, 7]. Prevention of falls should be one of the first defense lines to support an active aging. Accordingly, the balance assessment of elderly is assuming great relevance in clinical practice, with the development of several screening tools and tests that are used to assess stability or its deterioration: these include both simple clinical measures and also sophisticated technologies [8]. One of the common and easiest functional tests frequently used in medical practice is the Berg Balance Scale (BBS) test. Initially, this was proposed for balance assessment in elderly population but it has been frequently used in subjects with stroke [9], Parkinson's disease [10], brain injury [11], and multiple sclerosis [12]. The test is composed by 14 items, in the following also referred to as *exercises*, with a score ranging from 0 to 4 points. The maximum BBS score is 56 and the test duration time is $\approx 15\text{-}20$ minutes. A score of 45 is indicated as a threshold for subjects at high risk of fall [13]; each reduction of 1 point in BBS score is correlated to an increased risk of 6-8% to fall [14].

Recently, within the aims of the DOREMI European project[1], a technological platform to support and motivate elderly people to perform physical activity has been developed, targeted at a reduction in sedentariness, cognitive decline and malnutrition, at the same time promoting an improvement in the quality of life and social inclusion. The general architecture of DOREMI is described in [15], whereas an analysis of the platform reliability under the influence of human factors in the real setting is provided in [16]. The DOREMI project is focused on the development of a systemic solution for healthy aging able to prolong the functional capacities of the elderly. Trough an integrated control of psychologically related socio-physical disabilities, vital signs combined with nutritional behavior, physical activity and social interaction, it should be possible to counter-

---

act cognitive and physical decline. This hardware and software-based platform is able to monitor users physical activity levels, integrating this information with nutritional parameters. One of the main innovative objectives of the DOREMI platform consists in the development of an automated system for balance assessment. In the proposed approach, the balance assessment system is an easy-to-use, cost-effective and unobtrusive solution for early pre-frail risk detection and frailty prevention. This system, being part of elderly daily activities, could support the remote identification of potential negative evolution of the physical health status, thus favoring the development of user-specific treatments. This innovative DOREMI solution leverages the Wii Balance Board, a low-cost, portable and widely available force platform, able to evaluate the user weight distribution at the four corners of its surface, developed by Nintendo for the Wii gaming console. This device has been compared to laboratory-grade force platforms [17][18] and its accuracy proved to be acceptable for the employment in numerous scientific studies involving balance assessment [17][19] and gait or balance rehabilitation [20][21][22]. In [22], the Wii Balance Board device accuracy was tested against a laboratory grade device on the measurement of COP (Center of Pressure) path length. The authors have shown that the Wii Balance Board exhibits excellent test-retest reliability for COP path length, also proving concurrent validity with laboratory-grade force platforms. In [18], the authors compared the Wii Balance Board against a gold-standard force plate produced by AMTI on the simultaneous measurement of COP displacement, also implementing an improved device calibration algorithm. The relation between COP related features (e.g. mean location, root-mean-square displacement and mean frequency) and the evaluation of the balance stability in elderly patients has been investigated in [23], which presents a comparative study on stability-related measures using regression methods on data collected from a force platform with the aim of estimating the probability of patients' fall. However, it is worth to note that the whole temporal signal generated by a force platform, such as the Wii Balance Board, potentially contains richer information than the above mentioned static parameters, thereby allowing to envisage approaches that try to directly and automatically exploit such richness of signal dynamics.

In this paper we propose a novel system for automatic assessment of balance abilities in elderly, able to estimate the overall BBS score of a user based on the stream of input signals gathered from the Wii Balance Board during the execution of only 1 BBS exercise out of 14. The major scientific goal of this work is to assess the feasibility of accurately estimating the overall BBS score by exploiting the temporal series from pressure sensors gathered during a single exercise execution by the user, and to provide an experimental validation of the proposed system on real-world data.

Such a scientific challenge requires to address the fundamental questions of whether such temporal series contain enough information to be correlated with the full BBS score and whether a machine learning model can efficiently exploit such an information to automatically estimate the score. Keeping in mind these objectives, we resort to the class of Recurrent

Neural Networks (RNNs) [24], which is widely recognized as particularly appropriate for processing and extracting relevant dynamic knowledge from noisy temporal data. In particular, as a result of an extensive and comprehensive comparison among different learning models, we take into consideration the Reservoir Computing (RC) paradigm [25, 26] and specifically the Echo State Network (ESN) model [27, 28], which represents a state-of-the-art approach in the context of efficient learning in temporal domains [29]. Moreover, in our analysis, we investigate specific approaches aiming at tailoring the proposed system to the learning task at hand and to its challenges. On the side of the learning model design and settings, this includes taking into consideration the possible left-right symmetry during the execution of the balance exercise, as well as the integration of the temporal input pressure signals with static clinical data of the subjects. Besides, another relevant aspect of analysis concerns the careful selection of the physical exercise to be performed, in order to account for both safety of exercise execution, under a clinical perspective, and richness of the gathered signals, on the information processing side. The proposed approach is experimentally validated on real-world data collected through a measurement campaign on 21 volunteer users, also through a performance comparison with alternative learning models in the field of machine learning and neuro-computing.

Overall, the system described in this paper represents an automatic solution for the assessment of balance abilities in elderly people, which is unobtrusive, safe, easy-to-use (even without the supervision by clinicians), and requires the execution of only 1 of the 14 BBS exercises, thereby allowing to practically save time for monitoring the balance stability. To the best of our knowledge, the approach proposed in this paper represents the first attempt to estimate the BBS score of a subject from temporal data collected by using a non-intrusive external (unworn) device during the execution of a single balance exercise of very short duration (≈ 10 seconds). Alternative recent literature approaches are based on more intrusive solutions in the sense that they require the use of wearable devices and the execution by the user of a wider set of exercises for the BBS score estimation. Moreover, they did not explore machine learning models tailored to the direct treatment of the temporal dimension of the gathered signals, such as RNN. In particular, among the contributes for the automatic (unobtrusive and objective) BBS score estimation, the pioneering work described in [30] proposes a k nearest neighbor (k-NN) approach to estimate the BBS score of a subject by using data collected from a tri-axial accelerometer placed on the lower back, during the consecutive execution of various items of the BBS test. In [31] it is proposed an automatic system in which the BBS score of a subject is estimated by summing up the output of 14 Multi Layer Perceptrons, each of which is individually and specifically trained on data pertaining to one of the different BBS exercises. The system described in [31] uses feature-based data from 5 inertial body-fixed sensors (3-axial accelerometers and angular velocity) and requires the subject to execute the complete BBS test (i.e. all the 14 balance exercises), thereby resulting in a comprehensive but more intrusive and time-consuming approach.

This paper consistently extends the preliminary work presented in [32], i) by introducing a complete description of the *system* including the hardware, the modules for data collection and the pre-processing; ii) by considering a larger set of BBS exercises individually used for system assessment, iii) by making publicly available a dataset for BBS estimation from temporal data collected by the pressure sensors on a balance board[2]; iv) by discussing the relevant clinical aspects of interest for the application; v) by providing a complete and extensive experimental analysis of the proposed approach for the problem of BBS score estimation, taking into account also specifically devised methodologies based on the characteristics of the problem[3]; vi) by extending the experimental investigation to alternative models in the area of neural networks for temporal data processing. All the above mentioned aspects were lacking in [32] and represent a scientific support to the provided results.

The rest of this paper is organized as follows. Section 2 describes the Wii Balance Board and the software infrastructure used for data gathering, whereas Section 3 illustrates the pre-processing techniques employed to prepare the real-world dataset used in this work. Section 4 presents a complete overview of the proposed system for automatic assessment of balance abilities and the description of the learning models considered. The bulk of this paper's contribution is presented in Section 5, which describes the results of the experimental evaluation of the proposed approach for BBS score estimation. A discussion on the achieved results is presented in Section 6. Finally, Section 7 draws the conclusions.

2. The Balance Board

The Wii Balance Board is a gaming device developed by Nintendo for the Wii console. It was first released in Japan in 2007 along with the Wii Fit, an exercise video game consisting in activities to be performed with the peripheral. The device is a force platform with a width of 32,5cm, lenght of 52 cm and height of 7,5cm, weighing $\approx$ 4 Kg, similar to a household body weight scale. The device, shown in Figure 1, is able to evaluate vertical forces at the four corner ends of its surface, i.e front left (FL), front right (FR), back left (BL), back right (BR), by using four strain-gauge pressure sensors placed inside the 4 feet bumpers. It is able to communicate the weight measured at the four corners via a Bluetooth 2.0 interface to a nearby Bluetooth receiver. Optimal data trasmission rates can be reached by keeping the balance board and the receiver at a distance of 4 meters maximum. Under these conditions, the device provides a variable sample rate of $\approx$ 100 samples/second. Data is sampled by the device at a rate of 100 samples/second and stored

in an internal memory buffer. A Bluetooth client, paired with the balance board, is able to read the latest sample stored in the internal buffer over a low level interface with polled read operations. Each data sample, which consists in the weight measured at the four corners and the binary status of the frontal button, is usually employed to compute the CoP (Center of Pressure), i.e. the vertical projection of the center of mass on the balance board plane.
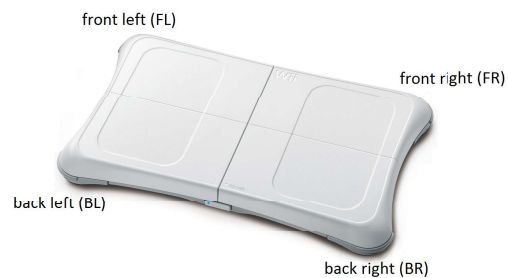

Figure 1: The Wii Balance Board.

Even if the Wii Balance Board is considered a fairly cheap force platform, the device has been employed in numerous scientific studies involving balance assessment [17][19] and rehabilitation addressing several disabilities [20][21][22], being a portable and widely available device with acceptable accuracy.

We developed a Java-based solution for collecting Wii board data using a minimal set of devices with a scarce set of computing resources. The set of components is depicted in Figure 2. The software modules are based on the Open Service Gateway initiative (OSGi) platform. OSGi represents a modular system and a service platform for the Java programming language [34], allowing an easy integration with existing state-of-the-art Ambient Assisted Living (AAL) platforms [35, 36]. In our scenario, a Linux-based system hosts an OSGi container running a set of developed Java components. The WiiBoard Abstraction layer bundle performs a Bluetooth device search when started, looking for devices in the Bluetooth transmission range (10m). This component leverages a WiiBoard Java Driver library sourced from the Wiibrew project [4] and customized in order to improve communication channel reliability and performance. The driver library implements the device-specific communication protocol. It connects to the Wii Balance Board via a Bluetooth Logical Link Control and Adaptation (L2CAP) connection, also performing the initial device pairing and calibration. Bluetooth communication capabilities are provided by the BlueZ official Linux Bluetooth protocol stack BlueZ [5] and a Bluetooth Class 2 interface dongle. When a Wii Balance Board is found by the Wiiboard Abstraction Layer, the component generates a proxy software instance (WiiBoard OSGi service), i.e. a software component providing an Application Programming Interface (API) that allows clients to listen for data being generated by the physical device. The component is

---

[2]Data and related description is available at the website address indicated in [33].

[3]Note that with respect to the preliminary work in [32], the extended experimental assessment in this paper provides novel results under a broader set of experimental conditions, both from the point of view of the learning model architecture and of the design of the resulting learning task. Such results ultimately lead to a final setup that involves a different learning model than in [32].

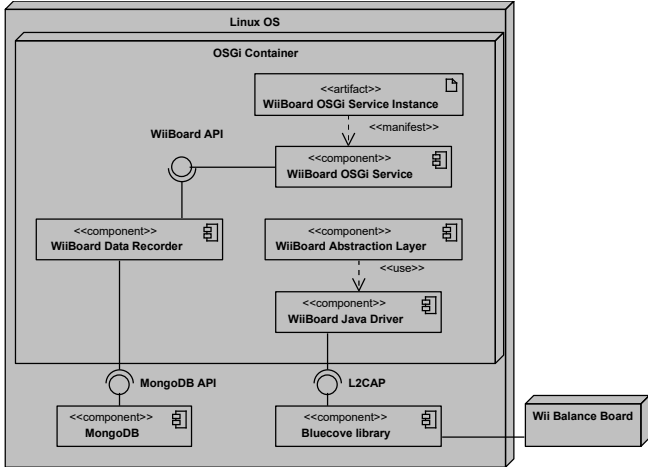[4]http://wiibrew.org/
[5]http://www.bluez.org/

Figure 2: WiiBoard Data collection system deployment diagram.

registered as an OSGi service in the container, making it available to multiple software modules running in the OSGi environment. Data produced by the Wii Balance board is recorded on a MongoDB [6] non relational database for later analysis. The WiiBoard Data Recorder component starts recording data as soon as a WiiBoard OSGi service is available. The system is able to run on low-end computing devices, such as credit card-sized computing devices (e.g., the Raspberry Pi [7]). In our laboratory setup, we used a laptop with the aforementioned software stack. The source code of the components is released with open source license and available from public repositories [8].

3. Dataset and Pre-processing

In the proposed system, the user BBS score is estimated from data generated by the execution on the Wii Balance Board of a single BBS exercise. For the design and validation of our system we restricted our attention on 3 out of the 14 exercises of the complete BBS test, namely exercises #6, #7, and #10. Such exercises have been chosen because they showed high inter-group variability in scores, greater difficulty to maintain stability during clinical examination, absence of technical issues for data acquisition with the balance board system (e.g. #11, #12, #13, #14 cannot be performed on a balance board), and do not require additional objects to be performed by subjects (e.g. the execution of exercises #1, #3, #4, #5, #9 and #12 requires additional objects). In order to train and validate our approach, we therefore set up a measurement campaign, conducted on 21 volunteers aged between 65 and 80 years. We collected the weight signal measured at the four corners of the balance board during the execution of the selected BSS exercises, which proved to be safe in all the 21 volunteers. Moreover, to take into account possible variations in the exercise execution, for each user we gathered data pertaining to a maximum of 10 repetitions of each

exercise. More details on the recruitment process and the protocols used during the measurement campaign are reported in Appendix A.

Data gathered from the Wii balance board presents a variable frequency of $\approx 100Hz$. In order to obtain a fixed frequency to be used as input for the learning modules (see Section 4), the time series relative to each exercise are resampled with a linear interpolation, avoiding aliasing or distortion to input data [37]. We denote the obtained weight signals at time step t by $u^r_{FL}(t), u^r_{FR}(t), u^r_{BL}(t), u^r_{BR}(t) \in \mathbb{R}$, where the subscripts FL, FR, BL and BR indicate the corresponding balance board corner, i.e. front left, front right, back left and back right, respectively. For each resampled sequence (at $100Hz$) of each board corner, we computed the Central Moving Average (CMA) over $N_{CMA} = 20$ samples, as shown in Equation 1:

$$\bar{u}^r_C(t) = \frac{1}{N_{CMA}} \sum_{i=-N_{CMA}/2}^{N_{CMA}/2} u^r_C(t+i), \tag{1}$$

where for each corner $C \in \{FL, FR, BL, BR\}$, $\bar{u}_C(t)$ is the un-weighted mean of the signal taken from an equal number $N_{CMA}/2$ of data on either side of a central value $u^r_C(t)$. This ensures that variations in the mean are aligned with the input data rather than being shifted in time, as in the case of using only "past" data. Hence, a central moving average can be computed, using data equally spaced on either side of the point in the time series where the mean is calculated [38]. The main advantage of the CMA over other techniques, as Exponential Moving Average (EMA), is that it offers a smoothed resulting timeseries, less prone to whipsawing up and down in response to slight or temporary changes in weight. The biggest CMA weakness is that it is slower to respond to rapid changes in data, but, in the considered scenario, this cannot happen due to the dynamics of the selected exercise [39]. The sequence of weight values corresponding to each exercise execution is then sub-sampled at a fixed frequency of $5Hz$ in order to achieve a trade-off among transmission rate (affecting the available bandwidth in real home installations), storage space, and signal resolution. For each time step t, the obtained weight values are collected into a column vector $\mathbf{u}(t) = [u_{FL}(t) \, u_{FR}(t) \, u_{BL}(t) \, u_{BR}(t)]^T \in \mathbb{R}^4$.

During the measurement campaign, users were asked to step up onto the balance board and do the various exercises with the balance board continuously collecting data. This generates data also when not needed. For this reason, pre-processing included a segmentation phase to eliminate data corresponding to a total weight under a threshold τ_w (equal to 4 kilograms in our setting). Figure 3 shows an excerpt of the data gathered in correspondence of an execution of exercise #10 before and after pre-processing, where it can be seen that the pre-processed sequence starts when the total weight is over τ_w. Figure 3 also shows that the application of the CMA filter smooths out the irregular roughness of the raw input providing a clearer signal.

An example of the temporal signals gathered in correspondence of a complete exercise execution (including the phases of getting on and off the balance board) is reported in Figure 4, which illustrates the temporal evolution of the user weight values at the four corners of the balance board during three rep-
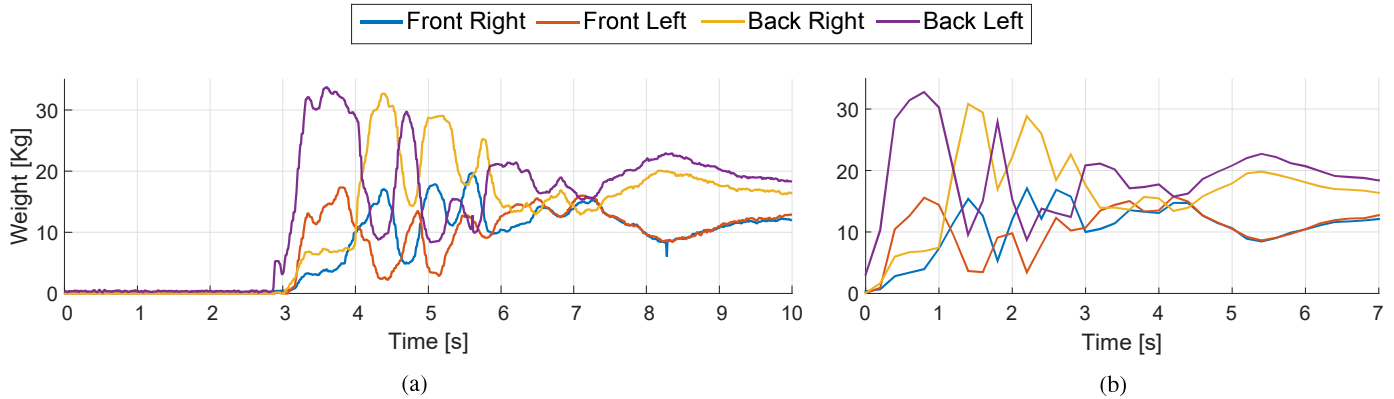
Figure 3: Plots of (a) the raw user weights from an exercise sequence and (b) the corresponding sequence after pre-processing.

etitions by the same user of exercise #10. Note that these three exercise repetitions correspond to the same ground-truth BBS score. Figure 4 shows how difficult would be to identify a pattern for the specific BSS score by human visual inspection, also due to the noisy nature of the signal. This is further complicated by the different ways the same exercise can be executed by the user, e.g. the way they step up onto the balance board (in terms of which foot is used first to step up and step down the board), the physical conditions of the users that can lead to lurching during the exercise, and the total duration of the execution.

As a result of the measurement campaign and data pre-processing we obtained Balance datasets for the definition of 3 regression tasks on sequences, one for each BBS exercise considered. The datasets contain couples of the type $(\mathbf{s}, y_{target})$, where $\mathbf{s} = [\mathbf{u}(1)\mathbf{u}(2)\ldots\mathbf{u}(L)]$ is the pre-processed 4-dimensional input sequence of length L containing the stream of weight values recorded by the balance board during the exercise execution, and $y_{target} \in [0, 56]$ is the corresponding target BBS score of the user, representing the ground-truth information evaluated by a clinician during the measurement campaign (by summing up the scores for all the 14 BBS exercises). Overall, the Balance datasets contain a total number of 470 sequences. The length of the collected sequences varies in the range of an average of 15 seconds for exercise #6, 65 seconds for exercise #7, and 10 seconds for exercise #10. The Balance datasets and their description are made publicly available at [33].

4. Learning the BBS Score using RC

The overall operation of the proposed system for automatic BBS score estimation is graphically sketched in Figure 5. While a subject executes a BBS exercise on the Wii Balance Board, the sensor stream is gathered and collected into a database as described in Section 2. Afterwards, data are pre-processed as described in Section 3 and then used as input for the neural network model that computes the overall BBS score estimate.

After pre-processing, the balance signals are processed by RNNs [24], and in this context we take as reference modeling approach the randomized neural networks [29] RC paradigm
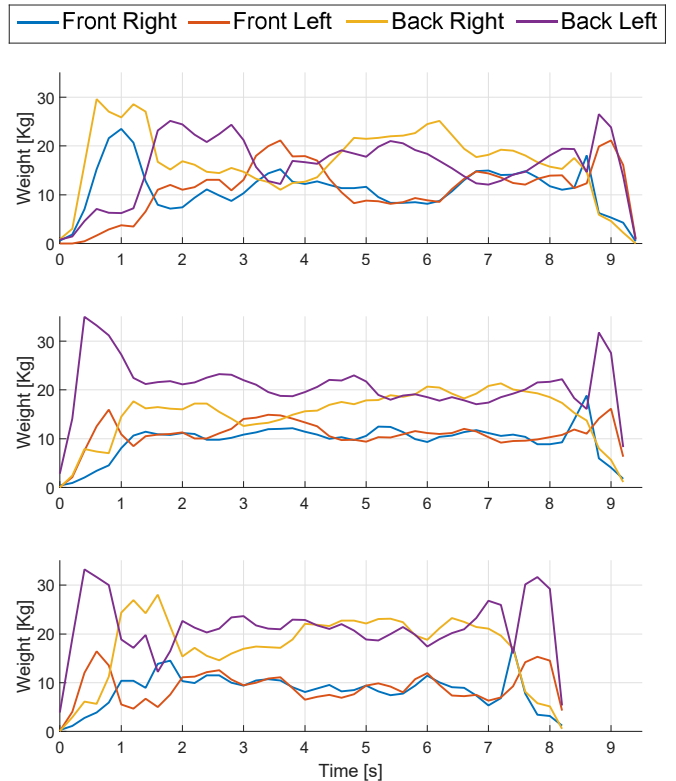

Figure 4: Plot of the pre-processed data from three repetitions of BBS exercise #10 by the same user with different ways of execution. Plot (a): exercise #10 performed placing the right foot first while stepping up onto the balance board and the right foot first while stepping down. Plot (b): exercise #10 performed placing the left foot first while stepping up onto the balance board and the right foot first while stepping down. Plot (c): Exercise #10 performed placing the left foot first while stepping up onto the balance board and the right foot first while stepping down. The exercise lasted less than usual.
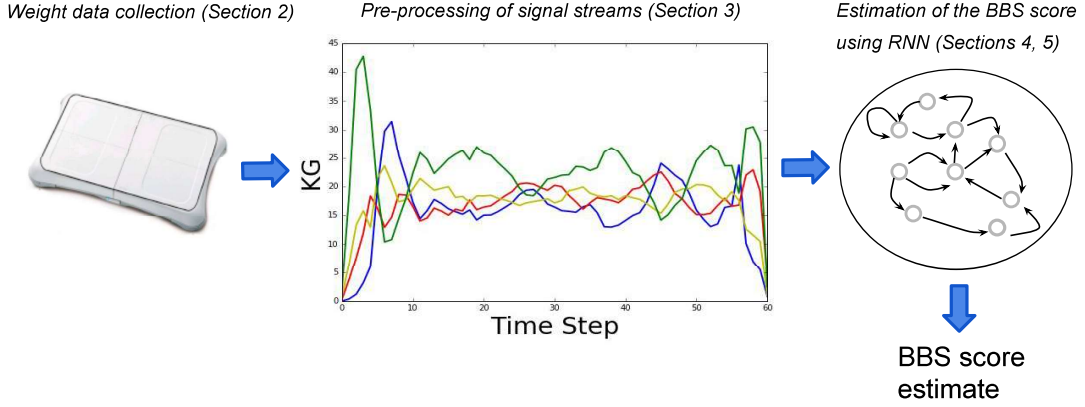
Figure 5: Graphical sketch of the overall operation of the proposed system for automatic BBS score estimation.

[25, 26]. Within the RC paradigm, we take into consideration the ESN [27, 28], which is an effective yet extremely efficient neural network model for learning in temporal domains, with a solid theoretical support (see e.g. [40, 41]) and great success in practical applications, with hundreds of successful experimental studies reported in literature [42, 26]. In particular, in this paper we use the Leaky Integrator ESN (LI-ESN) [43], a variant of the standard ESN characterized by an experimentally proved ability to effectively process temporal noisy information generated by heterogeneous sensor sources, as comparatively investigated with respect to the standard ESN model in [44]. Recent works have indeed reported successful applications of LI-ESNs in the areas of ambient assisted living, including learning tasks related to human activity recognition [45, 46, 47], prediction of the user's environmental context [48, 49], and indoor robot localization [50, 51], as also experimentally shown by the results of the RUBICON European project[9] [52, 53].

LI-ESNs implement discrete-time dynamical systems, exploiting the memory over past inputs provided by a reservoir of non-linear neurons encoding signal dynamics, whose activations are then linearly combined by an output function. The general architecture of a LI-ESN, shown in Figure 6, consists of an input layer with N_U units, a non-linear recurrent reservoir layer with N_R sparsely connected units and a linear feed-forward output layer with N_Y units. In our application, the input is represented by the 4-dimensional stream of values gathered by the balance board, whereas the output consists in the estimated BBS overall score, thereby we use $N_U = 4$ and $N_Y = 1$. Given an input sequence of length L, i.e. $\mathbf{s} = [\mathbf{u}(1)\mathbf{u}(2)\dots\mathbf{u}(L)]$, at each time step t, the reservoir computes the state $\mathbf{x}(t) \in \mathbb{R}^{N_R}$ according to the following state transition function:

$$\mathbf{x}(t) = (1 - a)\mathbf{x}(t - 1) + a\,\tanh(\mathbf{W}_{in}\mathbf{u}(t) + \hat{\mathbf{W}}\mathbf{x}(t - 1)) \quad (2)$$

where $\mathbf{u}(t) \in \mathbb{R}^{N_U}$ is the input at time t, $\mathbf{W}_{in} \in \mathbb{R}^{N_R \times N_U}$ is the input weight matrix (possibly including a bias term), $\hat{\mathbf{W}} \in \mathbb{R}^{N_R \times N_R}$ is the recurrent reservoir weight matrix, and tanh denotes the
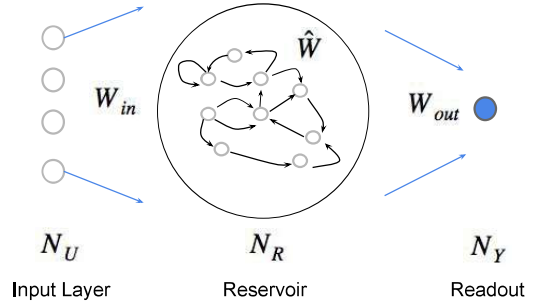

Figure 6: Architecture of a LI-ESN.

element-wise applied hyperbolic tangent non-linearity. The parameter a in equation 2 is the leaking rate parameter, taking values in the range $[0, 1]$, and controlling the speed of reservoir dynamics in reaction to the input [43, 25]. Note that if $a = 1$, equation 2 reduces to the case of standard ESN, while in the case of $a < 1$ an exponential moving average is applied to the reservoir states over time, acting as a low-pass filter on the network dynamics. This filtering effect is modulated by the value of the leaking rate parameter a, such that smaller values of a correspond to reservoir dynamics that respond more slowly to the input signal. As such, LI-ESNs provide a suitable mean of dealing with temporal data gathered from sensor devices, covering a whole spectrum of different rates at which the input signals can change with respect to the sampling frequency.

For regression tasks in which only one output value is required in correspondence of an entire input sequence, such is the case of our application to the BBS score estimation, the output corresponding to $\mathbf{s}$ is computed from a state $\chi(\mathbf{s}) \in \mathbb{R}^{N_R}$ that encodes the entire input sequence as processed by the reservoir. We consider two approaches to compute $\chi(\mathbf{s})$. In the first one, referred to as *root state mapping*, the last reservoir state computed for $\mathbf{s}$ is considered as representative of the whole encoding process, i.e.

$$\chi(\mathbf{s}) = \mathbf{x}(L). \quad (3)$$

A second approach consists in averaging the states computed for every time step of the input sequence, referred to as *mean state mapping* and representing a technique that has proved to be effective in several application contexts [54, 55]. The mean state mapping is computed according to:

$$\chi(\mathbf{s}) = \frac{1}{L} \sum_{t=1}^{L} \mathbf{x}(t). \tag{4}$$

In the following, when we want to indicate the use of a specific state mapping we use the notations LI-ESN-R and LI-ESN-M, respectively, for root state mapping and mean state mapping.

The output of the LI-ESN is computed by the readout component as a linear combination of the elements of $\chi(\mathbf{s})$, according to the following equation:

$$y(\mathbf{s}) = \mathbf{W}_{out}\chi(\mathbf{s}) \tag{5}$$

where $\mathbf{W}_{out} \in \mathbb{R}^{N_Y \times N_R}$ is the readout-to-reservoir weight matrix (possibly including a bias term). The readout is the only trained component of the LI-ESN architecture, typically by means of efficient linear methods such as ridge regression [25]. The reservoir is initialized under the constraints of the echo state property (ESP) [27, 28, 40] and then it is left untrained. The ESP is related to contractivity conditions of the state transition function in equation 2, which bounds the network state dynamics within a region of the state space characterized by Markovian properties [40, 56]. A reservoir initialization condition related to the spectral radius of matrix $\tilde{\mathbf{W}} = (1-a)\mathbf{I} + a\hat{\mathbf{W}}$ is often used in literature [25] for practical applications:

$$\rho(\tilde{\mathbf{W}}) < 1 \tag{6}$$

where, in the following, we use the symbol ρ to refer to the spectral radius of matrix $\tilde{\mathbf{W}}$. Thereby, a widely adopted process to initialize the reservoir of a LI-ESN consists in randomly choosing the values in $\hat{\mathbf{W}}$ from a uniform distribution, and then re-scaling them such that equation 6 is satisfied. Moreover, the weight values in $\mathbf{W}_{in}$ are typically randomly chosen from a uniform distribution over $[-scale_{in}, scale_{in}]$, where $scale_{in} \in \mathbb{R}$ is an input scaling parameter. Further details concerning initialization, characterization of state dynamics and readout training of RC networks can be found in [25, 40].

In the following Section 5, the effectiveness of the proposed approach is assessed also in comparison to other neural network models, taking into consideration a complete pool of approaches characterized by a progressively increasing ability to represent the temporal information involved in the task. On the general side, approaches based on neural networks for processing temporal information proved successful in a broad range of real-world application domains (see e.g. [24, 57, 58, 59, 60, 61, 62]). In particular, in this work we considered Multi Layer Perceptrons (MLPs) [57], Time Delay Neural Networks (TDNNs) [63] and Simple Recurrent Networks (SRNs), also known as Elman networks [64, 65, 66]. MLPs are suitable for treating static information only, i.e. vectors of fixed-length, thus in order to deal with sequential data a preliminary process of data concatenation with ad-hoc padding to a maximum possible length is required. TDNNs are networks based on the same MLP architecture used in conjunction with a buffering approach for treating temporal data. Accordingly, TDNNs implement time delay lines with a window of finite dimension that is directly related to the extent of the model's finite memory (a larger window leads to a longer memory) and to the number of model's free (trainable) parameters (a larger window leads to an increase in the number of free parameters). Instead of resorting to a fixed-size a-priori decided window, SRN networks implement discrete time dynamical systems by representing the temporal context through state variables formed in a recurrent architecture with feedback connections, allowing the system to potentially maintain input history information for arbitrary periods of time. Differently from the RC approach, MLPs, TDNNs and SRNs require a training process that involves all the weights in the network's architecture. This process is typically implemented by means of gradient descent algorithms, whose specialization for recurrent models is know as back-propagation through time (BPTT) [67]. As a result, these models are characterized by a higher computational cost during training than RC networks. Moreover, due to the potentially higher number of free parameters (with respect to RC), such architectures can be more prone to training data overfitting, leading to an increased risk of loosing generalization in the predictive performance.

For a further comparison with literature results, in Section 5 we also consider a k nearest neighbor (k-NN) algorithm similar to the method proposed in [30], using a preliminary data processing approach in line with the one in [30] in order to treat sequential input data through a fixed-size vectorial representation.

5. Experimental Results

The predictive performance of the learning models taken into account is evaluated through a 7-fold cross validation process, splitting the available data according to a 3 persons-out approach, i.e. such that each fold contains all the data pertaining to 3 out of the 21 volunteer users. Note that the persons-out approach adopted in the cross-validation scheme is of particular relevance for the purposes of this type of real-world applications. Such approach has indeed the advantage of enabling the estimation of the performance of future assessments on new subjects during the real operational use of the proposed system (and thus completely unseen in the training phase).

In our computational experiments with LI-ESNs, we used reservoirs with 10% of connectivity and spectral radius $\rho = 0.99$. The values of the other relevant LI-ESN hyper-parameters, including the reservoir dimension N_R, the input scaling $scale_{in}$, the leaking rate a, and the readout regularization for ridge regression training λ_r, were chosen from the ranges reported in Table 1 through a model selection process on a validation set, by means of an extra level of 6-fold cross validation on the training set of each external fold. For each reservoir hyper-parametrization, we independently generated 5 reservoir guesses, and the predictive performance in the different cases has been averaged over such guesses.

Hyper-parameter	Values considered for model selection
reservoir dimension N_R	10, 20, 50, 100, 200
input scaling $scale_{in}$	0.1, 0.5, 1
leaking rate a	0.1, 0.3, 0.5, 0.7, 1.0
readout regularization λ_r	0.001, 0.01, 0.1, 1, 10

Table 1: Range of LI-ESN hyper-parameters values considered for model selection.

For what concerns the values of the hyper-parameters used for MLP, TDNN, SRN and k-NN were selected from the ranges reported in Table 2, according to the same double cross-fold validation scheme used for the LI-ESN models, and generally resulting in a number of networks free parameters in a range that is compatible with the case of LI-ESN. In particular, for MLP we considered an input size of maximum dimension (i.e. 96), whereas for TDNN we considered a variable length of the input delay order (i.e. of the input window) and a maximum possible size of the hidden delay order (i.e. of the window for hidden units). As a general experimental setup for MLP, TDNN and SRN, we considered different sizes of the hidden layers, and trained the networks using gradient descent algorithm with momentum, varying the weight-decay regularization coefficient as specified in Table 2, and using learning rate and momentum coefficient respectively equal to 0.001 and 0.5. As in the experimental setting for LI-ESNs, also for MLPs, TDNNs and SRNs for each hyper-parametrization we generated 5 network guesses (from random weights initialization) and averaged the results over them. For what concerns the k-NN, as specified in Table 2, we considered values of k in the range 3-10 and we implemented the distance measures considering the Euclidean distance (ED), the Tanimoto coefficient (TC) and the correlation coefficient (CC), in order to allow a direct and complete comparison with the results of the approach reported in [30].

Hyper-parameter	MLP	TDNN	SRN	k-NN
Input delay order	96	5, 25, 50	-	-
Hidden delay order	-	96	-	-
Output delay order	-	-	-	-
Hidden units	10, 20, 30			-
Regularization	0.1, 0.5, 1			-
k		-		3, 5, 7, 10
Distance function		-		ED, TC, CC

Table 2: Range of hyper-parameters values considered for MLP, IDNN, TDNN, SRN and k-NN for model selection with double cross-fold validation. The distance functions considered for the k-NN model were the Euclidean distance (ED), the Tanimoto coefficient (TC) and the correlation coefficient (CC).

Regarding the experimental setup adopted for static models able to treat input information only in form of vectorial data of fixed size, we adopted the following pre-processing strategies.

For MLPs, the time steps elements of each input sequence were concatenated into a vector of fixed size equal to the maximum possible sequence length in the dataset, i.e. 96, where shorter sequences were completed by left-padding with zero values. For k-NN, in order to be able to compare the sequences using a distance measure, an approach similar to the one adopted in [30] was used, i.e. all the sequences were normalized to the average length in the dataset (i.e. 52) by means of either interpolation or decimation, subsequently concatenating all the elements to form a (1-dimensional) vectorial representation.

We considered 3 regression tasks (see Section 3) on the data collected during the execution of exercises #6, #7 and #10 of the BBS test, in the following referred to as *BBS-6*, *BBS-7* and *BBS-10*, respectively. The performance achieved by the learning models on the considered regression tasks has been computed by means of the Mean Absolute Error (MAE) and of the correlation coefficient R, respectively providing a measure of the absolute deviation and of the strength of linear relationship between the estimated and the ground-truth BBS score. Moreover, such a choice on the one hand also allows a direct comparison with the literature results in [30, 31], and on the other hand it enables the results evaluation from a clinical viewpoint. Assuming the dataset under consideration contains N sequences, MAE and R are defined by the following equations:

$$MAE = \frac{1}{N} \sum_{\mathbf{s}} |y_{target}(\mathbf{s}) - y(\mathbf{s})| \qquad (7)$$

$$R = \sqrt{1 - \frac{\sum_{\mathbf{s}}(y_{target}(\mathbf{s}) - y(\mathbf{s}))^2}{\sum_{\mathbf{s}}(y_{target}(\mathbf{s}) - \frac{1}{N}\sum_s y(\mathbf{s}))^2}}, \qquad (8)$$

where in correspondence of each input sequence $\mathbf{s}$, $y(\mathbf{s})$ and $y_{target}(\mathbf{s})$ denote the output of the learning model and the (ground-truth) target, respectively.

A preliminary experimental analysis of the results achieved on BBS-6, BBS-7 and BBS-10 allowed us to choose some common aspects of the experimental setting concerning the input sampling frequency and normalization. As regards the sampling frequency of the input signals gathered from the Wii Balance Board, as described in Section 3, the pre-processing phase has made available data at both the frequencies of $100Hz$ and $5Hz$. Between these two possibilities we selected the $5Hz$ frequency as it allowed us to achieve a considerable reduction in terms of the required training time, without a loss of fitting compared to the $100Hz$ case, at the same time relieving the overall system load in terms of data storage and transmission (see Section 3). As regards the normalization of the input data, we found out that the absolute weight of the subjects at the 4 corners of the balance board can greatly affect the amplitude of the oscillations of the balance signals. Therefore, each signal dimension in each input sequence $\mathbf{s}$ has been individually normalized to zero mean and unitary standard deviation before being used as input for the learning model. This normalization process allowed us to deal with time-series data in which signals amplitudes are in the same order of magnitude, so that the postural sways and the shape of pattern of the input sequences are no longer dependent on the absolute user weight.

The following sub-sections describe the experimental results achieved by the considered learning models. First, we show the progressive performance improvement brought about by the considered features of the experimental settings. In particular, in order to avoid cluttering of the presented results, we focus on showing the impact on the performance achieved by the best (i.e. the selected) model[10]. Then, we provide a complete comparison among all the learning models with respect to the use of the experimental features.

Specifically, the proposed performance assessment is targeted at the following aspects: the selection of the BBS exercise to be performed among the 3 considered (Section 5.1), the experimental assessment of the use of a weight sharing architectural variant (Section 5.2), the impact on the predictive performance of augmenting the input with users' clinical data (Section 5.3) and the joint use of weight sharing and clinical information (Section 5.4). A performance comparison among the considered learning models under the general final experimental settings is performed in Section 5.5. Finally, a discussion on the selected model and on the analysis of the BBS score estimation on subjects is provided in Section 5.6.

5.1. Selection of the BBS Exercise

As described in Section 3, exercises #6, #7 and #10 of the BBS test were chosen by clinical staff for reasons of safety and feasibility of execution on the Wii Balance Board, while in the considered models we require the user to perform only one BBS exercise.

Being the most important aspect in this regard, we first observe that from a clinical point of view, exercise #10 is particularly relevant since it integrates voluntary motor ability of the trunk and head, testing amplitude, gross quality, rapidity and independence in mobility, together with ability to maintain position or complete movement tasks. Moreover, as demonstrated by Kornetti et al. in [68], #10 is one of the BBS exercises, together with exercise #13 "tandem stance", exercise #12 "alternating foot" and exercise #14 "standing on one leg", that identify people who are able to reach cut-off value of 45 and successfully complete the test. Given such clinical considerations, we selected exercise #10 for our proposed system implementation.

In addition to this, it is also interesting to point out that, for what regards the generated data streams, exercise #10 is characterized by a richer temporal signal than #6 and #7. Indeed, performing of exercise #10 involves the execution of multiple motion patterns, including both static phases of standing (as in the cases of exercises #6 and #7) and phases of dynamism (which are missing in exercises #6 and #7). This aspect is further underpinned by the inherent discriminating ability that is allowed among different users in the three cases, quantitatively assessed, for instance, through the computation of the values of Pearson coefficients. Specifically, we computed inter-subject Pearson

correlation between signals gathered during exercise execution by the different subjects (averaged over the exercise repetitions in the datasets). The mean correlation values achieved for exercises #6, #7 and #10 are respectively 0.42, 0.44 and 0.40, clearly showing that exercise #10 is featured by a lower inter-subject correlation.

As a further support to the clinical considerations, we compared the predictive performance achieved by the learning models on the 3 tasks BBS-6, BBS-7 and BBS-10. Results generally show that a better validation performance is achieved on the BBS-10 task, providing experimental evidences that BBS exercise #10 enables a more accurate estimation of the total BBS score. While a complete view of the results can be found in the supplementary material [33], for the sake of conciseness here we show the impact of the different BBS exercises on the selected model, i.e. LI-ESN. We could also note that LI-ESN-R led to better results than LI-ESN-M, allowing us to focus the report of the experimental analysis in the rest of this Section to the case of root state mapping. Table 3 reports the values of MAE and R obtained by LI-ESN-R on the BBS-6, BBS-7 and BBS-10 tasks. From such results we can see that the best performance on the validation set was achieved on the BBS-10 task, which also corresponded to the smallest generalization MAE and the larger R value.

Task	TR MAE	VL MAE	TS MAE	TS R
BBS-10	3.56 ± 0.12	4.21 ± 0.14	4.80 ± 0.40	0.68
BBS-7	3.74 ± 0.11	4.74 ± 0.17	5.05 ± 0.32	0.51
BBS-6	4.43 ± 0.13	5.04 ± 0.19	5.53 ± 0.43	0.53

Table 3: Training (TR), validation (VL) and test (TS) MAE obtained by LI-ESN-R on the BBS-6, BBS-7 and BBS-10 tasks. R values on the test set are reported as well.

Overall, the clinical and experimental considerations described in this sub-section testify the goodness of the choice of the BBS exercise #10 as balance exercise to be performed by the user, thereby allowing us to restrict the focus of the following experimental assessment on the task BBS-10 only.

5.2. Weight Sharing Approach on Input Connections

During the exercise execution, the user can get on and off the balance board with the right or with the left foot in an arbitrary manner. The way in which the user gets on and off the board affects the shape of the input signals at the beginning and at the end of each input sequence (see e.g. Figure 4 for a graphical example). This type of information is only related to the measurement campaign and the corresponding real-world data sampling, and it is not relevant to the aim of postural balance assessment. Accordingly, we do not want the learning models to specialize on this information. A neural network architectural variant that can be adopted to reach this purpose makes use of a weight sharing (WS) approach applied to the input connections, such that the influence of the input signals coming from the left and right sides of the board is the same. Such approach enables to embed a left-right invariance in the process-

---

[10]Further information, including exhaustive details on the performance achieved by all the models in all the experimental settings are reported in the supplementary material document [33] accompanying the dataset.
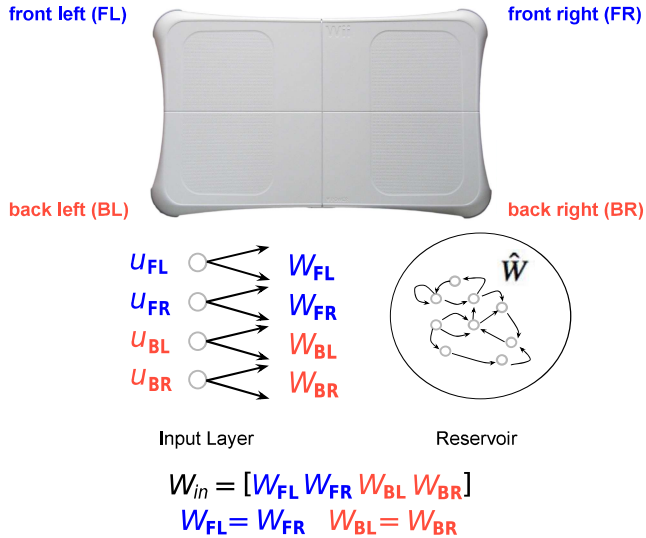
Figure 7: Graphical illustration of the adopted weight sharing approach (in the RC setting). The input weights pertaining to signals coming from the left and the right side of the balance board are shared.

ing of the input signals. Considering a row-wise aggregation of the input weights pertaining to the signals at the 4 corners of the balance board, the input weight matrix can be written as $\mathbf{W}_{in} = [\mathbf{W}_{FL}\mathbf{W}_{FR}\mathbf{W}_{BL}\mathbf{W}_{BR}]$. In this case the adopted WS approach, graphically depicted in Figure 7, consists in sharing the weights in $\mathbf{W}_{in}$ such that $\mathbf{W}_{FL} = \mathbf{W}_{FR}$ and $\mathbf{W}_{BL} = \mathbf{W}_{BR}$. The performance achieved by selected model, i.e. LI-ESN-R, on the BBS-10 task by adopting the WS technique is reported in Table 4. The positive effect of the WS approach in this application is testified by the fact that the predictive performance with respect to the case in which WS is not adopted is improved both on the validation and on the test sets. Indeed, through a comparison between Tables 4 and 3 it can be seen that the validation MAE is reduced by 0.12 BBS score points (0.21% over the whole BBS score range), corresponding to an average error reduction of 2.85% , while the test MAE is reduced by 0.77 BBS score points (i.e. 1.38% of the BBS score range), leading to an average error reduction of 16.04%.

Task	TR MAE	VL MAE	TS MAE	TS R
BBS-10 + WS	3.43 ± 0.04	4.09 ± 0.08	4.03 ± 0.13	0.71

Table 4: Training (TR), validation (VL) and test (TS) MAE obtained by LI-ESN-R on the BBS-10 task using the WS technique. The R value on the test set are reported as well.

5.3. Use of Clinical Data

A further significant experimental assessment consisted in the evaluation of the influence on the predictive performance of users' clinical data such as height, weight, age and gender. The values of such clinical data were used as input to the model by considering the height expressed in meters, the weight expressed in hundreds of Kg (i.e. dividing the weight in Kg by

100), the age expressed in hundreds of years (i.e. dividing the age in years by 100), and the gender expressed as a binary value (-1 for women and +1 for men) [11].

To avoid the introduction of artifacts related to the specific dataset on which we trained the learning models, we excluded the use of the clinical features that resulted in a high correlation with the BBS scores of the users in our sample data (without however having a known correlation in the general case or in literature). This analysis allowed us to exclude from consideration users' height and gender, restricting the focus on users' weight and age only. The augmented input has been implemented by appending at each time step of each sequence in the dataset the value of the clinical parameter of the corresponding user as a further input element. This process resulted in input sequences consisting of 5 elements per time steps, i.e. $\mathbf{u}(t)_{augmented} = [u_{FL}(t)u_{FR}(t)u_{BL}(t)u_{BR}(t)u_{clinical}]^T \in \mathbb{R}^5$, where $u_{clinical}$ is the clinical feature (weight or age). We accordingly prepared two variants of the BBS-10 task, corresponding to the cases of augmenting with users' weight or age, in the following referred to as tasks BBB-10-W and BBB-10-A, respectively. Table 5 reports the performance achieved by the selected model, i.e. LI-ESN-R, on these two tasks, showing that the best result is achieved in correspondence of the BBS-10-W task[12]. Comparing Tables 5 and 3, it is possible to observe that augmenting the input data with the user weight information ultimately leads to a reduction in the validation MAE of 0.13 BBS score points (0.23% of the BBS score range), corresponding to an average error reduction of 3.09%, whereas the improvement in the test MAE is quantifiable in 0.18 BBS score points (the 0.32% of the BBS score range), with an average error reduction of 3.75%.

Task	TR MAE	VL MAE	TS MAE	TS R
BBS-10-W	3.50 ± 0.08	4.08 ± 0.09	4.62 ± 0.30	0.69
BBS-10-A	3.67 ± 0.12	4.23 ± 0.13	4.52 ± 0.27	0.65

Table 5: Training (TR), validation (VL), test (TS) MAEs, obtained by LI-ESN-R on the BBS-10-W and BBS-10-A tasks. R values on the test set are reported as well.

5.4. Joint Use of Weight Sharing and Clinical Data

Results discussed in Sections 5.2 and 5.3 have shown the practical advantage in terms of improved performance that is singly brought about by the use of strategies related to network architectural design such as an appropriate WS technique, or when the learning model receives in input the overall user weight in addition to the time series data gathered by the balance board. In this Section, we explore the synergy of these two approaches to provide a final setup of the learning model

---

[11]These choices were made as simple scaling approaches, in order to have values approximately in the same range of the signals coming from the balance board.

[12]Detailed results on all the learning models are reported in the supplementary material [33].

for BBS score estimation. Accordingly, Table 6 shows the results obtained by the selected model, i.e. LI-ESN-R, using the WS approach on the BBS-10-W task, i.e. augmenting the input with the user weight information. Table 6 also recalls the performance achieved by LI-ESN-R without the WS approach and the augmented input information to facilitate performance comparison.

Task	TR MAE	VL MAE	TS MAE	TS R
BBS-10 (no WS)	3.56 ±0.12	4.21 ± 0.14	4.80 ± 0.40	0.68
BBS-10-W + WS	3.11 ± 0.05	3.85 ± 0.08	3.80 ± 0.17	0.76

Table 6: Training (TR), validation (VL) and test (TS) MAE obtained by LI-ESN-R on the BBS-10 task (without WS) and on the BBS-10-W task (with WS). R values on the test set are reported as well.

Results in Table 6 show that the joint effect of using WS and user weight input information is indeed superior to the single improvements due to the use of the two approaches alone. Indeed, LI-ESN-R with WS and users' weight in input achieved very close values of validation and a test MAE, respectively equal to 3.85 ± 0.08 and 3.80 ± 0.17, corresponding to an improvement of 0.37 BBS score points in the validation set (0.66% of the BBS score range), i.e. an average performance improvement of 8.79%, and of 1 point in test set (1.79% of BBS score range), with an average performance improvement of 20.83%. Moreover, the R value on the test set is overall improved of 0.08, i.e. of the 11.76%.

5.5. Comparison among the Learning Models

In this sub-section we provide a summary of the comparison among the results obtained on the BBS score estimation task, including all the learning models considered, i.e. LI-ESN, MLP, TDNN, SRN and k-NN. Specifically, whereas a detailed report on the performance achieved in all the experimental conditions is provided as part of the supplementary material [33], in light of the considerations made in previous sub-sections here we focus on the basic and on the full settings of the BBS-10 task, as described below.

A first performance comparison was conducted on the basic BBS-10 settings, i.e. without using the WS technique and the augmented input with users' overall weight. Table 7 reports the performance achieved by LI-ESN-R, MLP, TDNN, SRN and k-NN on the BBS-10 task, while Figures 8 and 9 graphically show the MAE obtained by such models on the validation and test set, respectively.

Model	TR MAE	VL MAE	TS MAE	TS R
LI-ESN-R	3.56±0.12	4.21±0.14	4.80±0.40	0.68
MLP	2.21±0.21	3.96±0.28	4.96±0.49	0.57
TDNN	2.79±0.16	3.72±0.34	4.69±0.70	0.54
SRN	3.85±0.34	4.02±0.42	4.86±0.56	0.57
k-NN	1.93±0.00	4.61±0.00	7.03±0.00	0.16

Table 7: Training (TR), validation (VL), test (TS) MAEs, achieved by LI-ESN-R, MLP, TDNN, SRN and k-NN on the BBS-10 task. R values on the test set are reported as well.
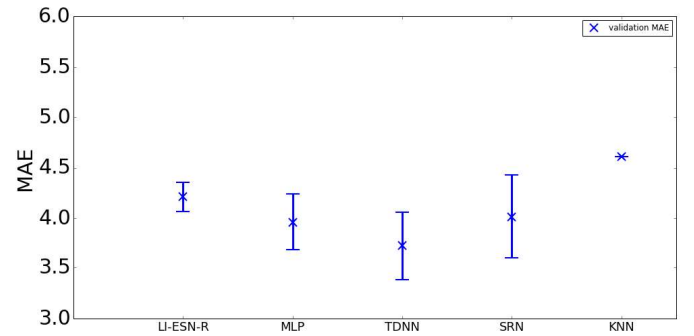

Figure 8: Validation MAE (and standard deviation represented by vertical intervals) obtained by LI-ESN-R, MLP, TDNN, SRN and k-NN on the BBS-10 task.
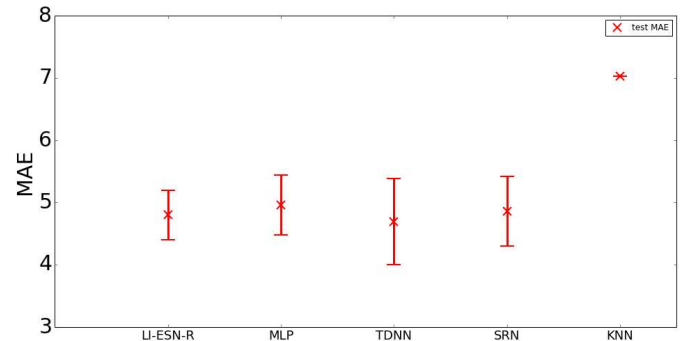

Figure 9: Test MAE (and standard deviation represented by vertical intervals) obtained by LI-ESN-R, MLP, TDNN, SRN and k-NN on the BBS-10 task.

From these results it is possible to observe that LI-ESN, MLP, TDNN and SRN show similar results on the validation set, all falling in the range of variability of each other. However, in order to select a model on the validation set (Figure 8), we note that LI-ESN shows a substantially lower performance variability (in terms of standard deviation) and a favorable ratio between the training and the validation MAE. The results are reflected on the test set performance (Figure 9), where the models show closer results on this basic setting. As can be seen from Table 7, MLP, TDNN and k-NN resulted in overfitting, whereas SRN achieved a performance close to the one of LI-ESN, but at the price of a higher standard deviation (on both validation and test sets). As a further evidence of the suitability of the LI-ESN approach for this task, it is worth noticing that

even on this basic (partial) configuration of the data representation it achieves the best R result on the test set, with a value of 0.68.

A second more significant performance comparison is reported for the complete configuration of data representation, i.e. under experimental settings analogous to those described in Section 5.4, which use WS and the input time-series augmented by the users' weight information. For what concerns the use of the WS approach on the input connections, note that this architectural variant, described in Section 5.2 in the RC setting, can be used also for MLP, TDNN and SRN, while it is not designed for the k-NN approach. The performance achieved by the considered models on the BBS-10-W task is reported in Table 8, while Figures 10 and 11 show the validation and test MAE errors, respectively. Note that for LI-ESN, MLP, TDNN and SRN, such results are referred to the use of WS.

Model	TR MAE	VL MAE	TS MAE	TS R
LI-ESN-R	3.11 ± 0.05	3.85 ± 0.08	3.80 ± 0.17	0.76
MLP	3.36 ± 0.16	4.62 ± 0.48	5.76 ± 0.66	0.56
TDNN	2.99 ± 0.26	3.89 ± 0.69	5.19 ± 0.91	0.57
SRN	3.72 ± 0.38	3.69 ± 0.45	4.34 ± 0.69	0.68
k-NN	2.89 ± 0.00	4.68 ± 0.00	6.05 ± 0.00	0.23

Table 8: Training (TR), validation (VL), test (TS) MAEs, achieved by LI-ESN-R, MLP, TDNN, SRN (with WS) and k-NN on the BBS-10-W task. R values on the test set are reported as well.
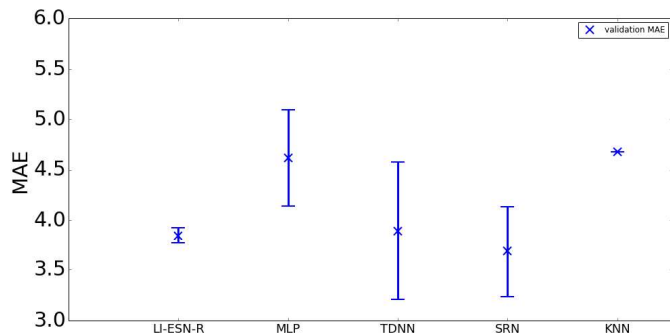

Figure 10: Validation MAE (and standard deviation represented by vertical intervals) obtained by LI-ESN-R, MLP, TDNN, SRN (using WS) and k-NN on the BBS-10-W task.
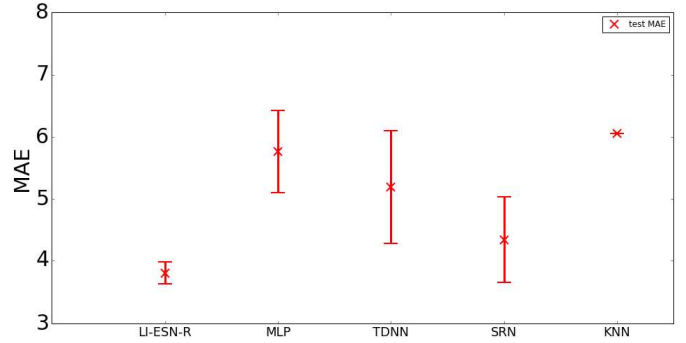

Figure 11: Test MAE (and standard deviation represented by vertical intervals) obtained by LI-ESN-R, MLP, TDNN, SRN (using WS) and k-NN on the BBS-10-W task.

The comparative performance analysis shows that the MAE achieved by LI-ESN, TDNN and SRN on the validation set are in the same range of variability (with values even closer than in the basic setting of the task), while MLP and k-NN obtained worse results. Furthermore, from Figure 10 and Table 8, we observe that in this setting LI-ESN presents the smallest standard deviation on the validation set (even smaller than in the basic setting) and a good ratio between training and validation MAE, thereby resulting in the selected model. The good performance of LI-ESN on the validation set is also reflected on the test set results (Figure 11), where in comparison to all the other learning models it clearly achieves the smallest MAE and the highest R value, confirming the goodness of the choice for the learning model in the final (complete) setting of the balance assessment task. From Table 8 it can also be observed that, as in the basic task settings, MLP, TDNN and k-NN resulted in overfitting. As a final remark concerning these results, it is worth noticing that they confirmed our choice of the selected LI-ESN learning model, also by virtue of its simplicity and efficiency, and of the complete configuration of the task, i.e. using the WS and the augmented input, fruitfully exploiting the improvements over the basic setting, as discussed in Section 5.4.

5.6. Selected Model and Analysis of BBS score Estimation on Subjects

The experimental analysis discussed in previous sub-sections highlighted that the best learning model for the BBS score estimation task is a LI-ESN-R with WS using board sensor data from exercise #10 and complemented by the subject personal weight information. The values of the LI-ESN hyper-parameters selected on the validation set for the final setup are reported in Table 9 for completeness.

Hyper-parameter	Selected value
reservoir dimension N_R	50
input scaling $scale_{in}$	0.1
leaking rate a	0.1
readout regularization λ_r	1.0

Table 9: Values of the LI-ESN hyper-parameters chosen through model selection for the final setup.

The predictive performance results achieved in the final setting are summarized in Table 10. Note that in this table, along with the MAE and R values summarizing the results in previous sub-sections, we also report the standard deviations computed on the test set under different perspectives. In particular, the uncertainty of the BBS score estimation is represented by the standard deviation on the external folds of the cross-validation, denoted by STDf in Table 10. The standard deviations with respect the reservoir guesses, the different sequences (i.e. the exercise repetitions), and the different users are respectively denoted by STDg, STDs and STDu.

VL MAE	TS MAE	STDg	STDs	STDf	STDu	TS R
3.85	3.80	0.17	2.92	1.64	2.01	0.76

Table 10: Performance results achieved in the final setting for BBS score estimation, i.e. with LI-ESN-R using WS on the BBS-10-W task. The table reports validation (VL) and test (TS) MAE, along with the standard deviation computed on the test set with respect to: the reservoir guesses (STDg), the different sequences (STDs), the external folds of the double cross-validation scheme (STDf), the different users (STDu). The R value on the test set is reported as well.

As it can be seen, the selected model achieved a validation MAE (mean of errors over the folds of the cross-validation) of 3.85 BBS score points (corresponding to the 6.88% of the total BBS score range), a test MAE of 3.80 BBS score points (corresponding to the 6.79% of total BBS score range). Such a result is indeed extremely good, considering that the generalization error is even below the threshold of 4 BBS score points, that is the score range of a single BBS exercise. Moreover, as seen in Table 10, the selected LI-ESN achieved a standard deviation of the MAE of 0.17, 2.92, 1.64 and 2.01 BBS score points with respect to the reservoir guesses, the different sequences (i.e. different exercise repetitions by the same user), the external folds in the double cross-fold validation and the different users, respectively. Note that these results are largely within the range of tolerance for clinical interpretation. A recent study [69] has, in fact, estimated that it is necessary to observe a difference of a least 8 BBS score points in order to diagnose an actual change in the postural balance ability of a subject.

The quality of the automated BSS estimate can be appreciated at a subject-level by graphically summarizing the results for all users in test. In this sense, Figure 12 shows a plot comparing the ground-truth BBS score measured by the clinicians for each subject versus the corresponding estimate provided by the LI-ESN model (computed by averaging the results obtained for each exercise repetition by the subject and for each reservoir guess). As can be seen, the points in the plot are generally distributed close to the bisector of the x-y axis, with MAE of 3.36 BBS score points, a correlation coefficient R equal to 0.8257, with p-value $p < 0.0001$. These summarized results, looking also at the distribution of errors, confirm the good quality of our RC-based system for automatic BBS estimation.
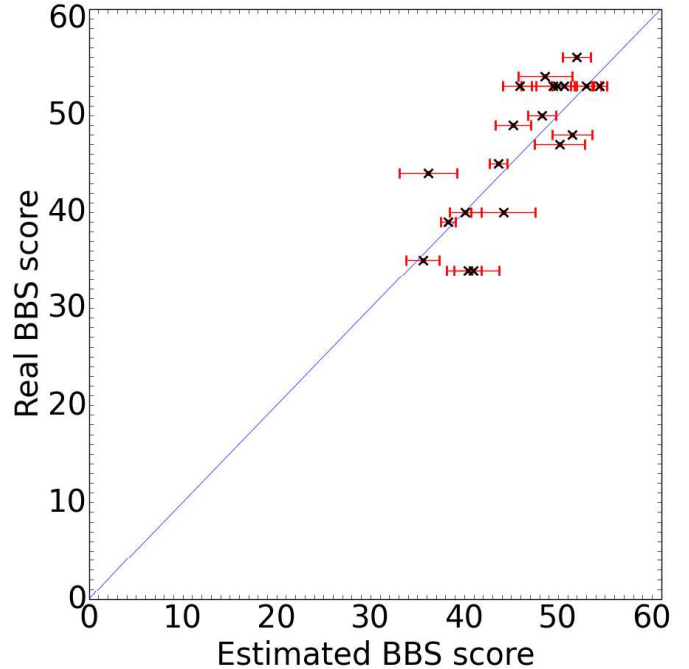

Figure 12: Real (ground-truth) versus LI-ESN estimated BBS scores plot (each point corresponds to a different user, evaluated in the test set). Standard deviations with respect to the exercise repetitions are reported in red.

Finally, it is worth mentioning that the proposed approach compares well with the ones already reported in literature for BBS score estimation tasks. In particular, the work in [30] describes a system based on a k-NN algorithm (similar to the one considered in this paper for performance comparison in Section 5.5) using data gathered during the execution of 3 BBS exercises from a 3-axial accelerometer positioned on the lower back of the user, achieving a MAE of 4.63 ± 3.89. The system described in [31] uses feature-based data gathered from 5 body-fixed sensors during the execution by the user of the entire BBS test and proposes to estimate the overall BBS score by summing up individual estimations of the 14 BBS exercises scores obtained by MLPs, obtaining a MAE of $\approx 1.5 \pm 1$. In comparison to such approaches our system outperforms the predictive performance of the work in [30], while the one in [31] results to be superior. In both cases, however, note that the literature systems present a higher intrusiveness, in the sense that they require the user to wear physical devices (1 sensor in [30], 5 sensors in [31]) and to perform a higher number of balance exercises (3 BBS exercises in [30], all the 14 BBS exercises in [31]). Thereby, the system proposed in this paper, using data collected during the execution of a single BBS exercise from an unworn device (a balance platform) is characterized by a favorable trade-off among predictive performance, required time and ease of use.

6. Discussion

We can draw two major lines of general considerations. Firstly, from the perspective of clinical trial feasibility, the proposed system is strongly time saving (≈ 10 seconds), compared

13

to complete BBS test (≈ 15-20 minutes), allowing to be replicated on a huge population during clinical analysis and totally unobtrusive also if the system is installed at user apartments. Furthermore, exercise #10 and the use of the proposed balance board proved to be safe in all the 21 volunteer users involved in our measurement campaign (see Section 3 and Appendix A), even when replicated several times. A further aspect of clinical interest is given by the results of the experimental analysis described in Section 5, showing an accurate estimation of the total BBS score when using data gathered during the execution of exercise #10. This on the one hand confirms the goodness of the choice of the BBS exercise to be performed by the users, and on the other hand it points out that the devised system actually represents an effective tool for automatic accurate assessment of balance skills of elderly people in the real-world. For example, the proposed system could be used to monitor the evolution of subjects' physical conditions or the effectiveness of specific therapies, providing useful feedback to the clinicians.

For what concerns the assessment of the predictive ability of the system, as pointed out in Section 5.6, the proposed RC-based approach allowed to estimate the overall user BBS score with an absolute deviation from the ground-truth value of 3.80 points on average on the test set, i.e. with a generalization error that is even smaller than the score range of each individual exercise of the BBS test. This result appears particularly promising and of great value also in consideration of the relevant noise in the input streams gathered from the balance board (as described in Section 3), the tolerance in the ground-truth data due to human observations and the fact that only data pertaining to a single BBS exercise is used as input for the estimation the overall BBS score. Moreover, the standard deviation achieved by our approach (see Section 5.6) is generally below 3 BBS points, a value that is largely within the range of tolerance for clinical interpretation. The RC models have experimentally proved to be effectively able to detect from the input streams pertaining to one single exercise much more information than what the clinician can express by the single exercise score, thereby making it possible to accurately learn to estimate the entire BBS score based on a single BBS exercise.

Overall, the possibility to infer the BBS scores with a good performance from the temporal data recorded in correspondence of a single BBS exercise shows the potentiality of our idea of exploiting the entire curve of the signal stream as a rich source of information for balance assessment. This confirms, in a constructive way, the positive response to the scientific challenge posed as a motivation in the Introduction. Furthermore, the experimental results illustrated in Section 5 provide a clear evidence of the effectiveness of the approaches introduced to tailor the settings of the learning methodology to the problem of balance assessment. Specifically, our results point out both the individual and the synergistic advantages brought about by the weight sharing approach on the input connections and by the integration of the input streams with clinical information.

7. Conclusions

In this paper we proposed an innovative learning system for the automatic assessment of balance abilities in the elderly. The main scientific challenge tackled by this work concerned assessing the feasibility of inferring the overall BBS score (based on the clinical evaluation on all the 14 BBS exercises) on the basis of the data streams collected by a Wii Balance Board during the execution by the subject of a single BBS exercise. The work described in this paper represents a positive answer to such a question, showing that dynamical neural networks, modeled according to the principles of the RC paradigm, allow to effectively yet efficiently exploit (in an automatic fashion) the richness of temporal dynamics contained in the data streams gathered from the balance board to extract information that is relevant for the task of overall BBS score estimation.

In order to train and experimentally validate the learning system, a measurement campaign has been carried out on 21 volunteer users, gathering data pertaining to the execution of a pool of 3 BBS exercises (i.e. #6, #7 and #10), considered as the most clinically suitable for the aims of automatic BBS score estimation from balance board data, and among these exercise #10 has been selected for the final setup of the system. The resulting Balance dataset has been made publicly available (online at [33]) and represents another relevant outcome of this work.

The experimental analysis of the proposed RC-based approach on real-world data, conducted by means of rigorous persons-out cross-fold validation, showed a very good predictive performance, allowing to accurately estimate the overall BBS score with an average generalization error of 3.80 BBS points in test. Interestingly, such a value is below the 7% of the whole BBS range (0-56) and it is even smaller than the extent of the range of assignable points for each single BBS exercise (0-4), ultimately suggesting that, with regard to the execution of one BBS exercise, the learning models have been able to extract more information than the one that can be provided by a clinician in terms of the score of a single exercise. From a clinical perspective, exercise #10 is more informative about the total BBS score and it allowed us to gather richer signal streams from the balance board. Therefore, it has been selected as balance exercise to be performed by the user in our system. Interestingly, the experimental analysis confirms the soundness of the selection of exercise #10. Under a perspective of RC network design, the results also provided experimental evidences supporting the effectiveness of introducing a weight sharing technique among the input weights for the data coming from the two sides of the balance board. Besides, from a task design point of view, augmenting the temporal input information with static user's weight data has proved useful to improve the generalization performance of the system. The relevance of the results achieved with RC networks has been further assessed through a performance comparison with other learning models, in particular in the area of neuro-computing, generally showing a better generalization performance with lower variability, and a favorable ratio among training, validation and test errors. Moreover, in comparison with literature works, our approach showed a favorable trade-off between predictive performance

and intrusiveness.

Overall, the system proposed in this work represents an automatic tool for the accurate estimation of a users' score at the BBS test (of ≈ 15-20 minutes of duration) from the execution of only one exercise of the test (of ≈ 10 seconds of duration, i.e. ≈ 1% of the duration of the whole test), thereby resulting in a tremendous time saving in the task of monitoring balance stability in elderly people. The system is characterized by limited obtrusiveness since it does not require the subject to wear any sensor. In this respect, it appears of straightforward use and particularly suitable for autonomous usage. Clinical and experimental evidence has in fact highlighted how the BBS exercise selected for our setup is of simple and safe execution and can be performed even without supervision by a clinician.

Finally, from a clinical perspective, it is worth to point out that the BBS test represents one of the first choices among tests for balance assessment, routinely used in clinical practice. Several studies have tried to improve it to better recognize people at risk of falls in terms of a) modify items rating or b) identify the most sensitive/specific among these. Our solution follows a third way: combine a specific BBS exercise with the development of a system (exploiting Wii Balance Board data and a learning system) for the automatic estimation of the overall BBS score. This approach is able to perform an objective evaluation of balance assessment, with some specific plus as time saving procedure and user safety. As a way of a possible further improvement for this work, we envisage the prospect of an even more extended data gathering campaign, e.g. by targeting also a lower range of BBS score values, enlarging the set of physical conditions and clinical parameters that are taken into account. Moreover, future developments foresee a system that will be placed in elderly people houses where it will allow daily monitoring of balance in an unobtrusive manner, alerting general practitioners or specialists if changes in the balance behavior occur.

Acknowledgment

Appendix A. Recruitment Process and Protocols for Clinical Test

Cardiologists of the Institute of Clinical Physiology IFC-CNR enrolled 21 elderly subjects. Selection was performed following the inclusion and exclusion criteria reported below:

- **Inclusion criteria**
  - Age between 65 and 80 years old
  - Balanced rate of male and female
  - No known relevant diseases
  - Written informed consent

- **Exclusion criteria**
  - Advanced cancer
  - Hearing and vision problems, which may interfere with physical activity
  - Moderate to severe aortic stenosis
  - Hypertrophic cardiomyopathy
  - NYHA III/IV heart failure
  - Chronic obstructive pulmonary disease
  - Oxygen saturation below 95%
  - Resting heart rate over 100 bpm

Cardiologists previously visited participants, in order to collect biometric data (i.e. weight, height, BMI). Table A.11 shows users' characteristics.

	Male	Female	Total
Nr. of participants	11	10	21
Age	73 ± 5	75 ± 5	74 ± 5
Weight [Kg]	85.3±15.5	74.0±13.3	79.9±15.3
Height [m]	1.76±0.06	1.59±0.04	1.68±0.10
BMI [Kg/m^2]	27.4 ± 4.1	29.5 ± 5.2	28.4 ± 4.6

Table A.11: Aggregated biometric data of study users.

After this step, each subject was evaluated through the use of Berg Balance Scale (BBS). Table A.12 shows scores' distribution for each item among the selected users.

Once defined the BBS score for each user, items #06, #07, #10 were tested with the use of the balance board.

Regarding exercise #06, approximately 6 repetitions were performed by each patient (duration of each repetition is approximately of 10 seconds; thereby the whole session has duration of approximately 60 seconds), following the protocol:

1. The operator fills a form in the Logger GUI, provided to the specialists, with the (anonymized) information regarding the patient (e.g. sex, age, weight, height, overall Berg score, and score in the specific exercise/item). This step is required only for the first repetition for each patient.
2. The patient stands in front of the balance board.

SCORE	TASKS													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
4	14	18	21	11	14	15	13	10	16	9	9	12	4	2
3	4	1	0	9	5	6	6	9	5	2	3	5	6	7
2	1	2	0	1	1	0	1	1	0	8	5	2	6	1
1	1	0	0	0	1	0	0	1	0	2	3	2	3	6
0	1	0	0	0	0	0	1	0	0	0	1	0	2	5

Table A.12: Aggregated biometric data of study users.

3. The operator clicks on the Start button of the Logger GUI, to start the data logging.

4. The patient gets on the balance board with the dominant foot.

5. The patient places both the feet on the balance board; feet should be placed as symmetrically as possible with respect to the vertical line in the middle of the board.

6. The patient closes the eyes and stands on the balance board for a maximum time of 10 seconds.

7. The patient gets off the balance board (backwards) with the dominant foot first.

8. The operator clicks on the End button of the Logger GUI to conclude the data logging.

9. The operator can add further information pertaining to the exercise repetition, filling the field notes in the Logger GUI.

Regarding exercise #07, approximately 8 repetitions were performed by each patient (duration of each repetition is approximately of 1 minute; thereby the whole session has duration of approximately 8 minutes), following the protocol:

1. The operator fills a form in the Logger GUI, provided to the specialists, with the (anonymized) information regarding the patient (e.g. sex, age, weight, height, overall Berg score, and score in the specific exercise/item). This step is required only for the first repetition for each patient.

2. The patient stands in front of the balance board.

3. The operator clicks on the Start button of the Logger GUI, to start the data logging.

4. The patient gets on the balance board with the dominant foot.

5. The patient is positioned on the balance board with the feet together; feet should be placed as close as possible to the vertical line in the middle of the board.

6. The patient remains in the same position, on the balance board with feet together, for a maximum time of 1 minute.

7. The patient gets off the balance board (backwards) with the dominant foot first.

8. The operator clicks on the End button of the Logger GUI to conclude the data logging.

9. The operator can add further information pertaining to the exercise repetition, filling the field notes in the Logger GUI.

Regarding exercise #10, approximately 8 repetitions were performed by each users (duration of each repetition is approximately of 15 seconds; thereby the whole session has duration of approximately 120 seconds) following the protocol:

1. The operator fills a form in the Logger GUI, provided to the specialists, with the (anonymized) information regarding the patient (e.g. sex, age, weight, height, overall Berg score, and score in the specific exercise/item). This step is required only for the first repetition for each patient.

2. The patient stands in front of the balance board.

3. The operator clicks on the Start button of the Logger GUI, to start the data logging.

4. The patient gets on the balance board with the dominant foot.

5. The patient places both the feet on the balance board; feet should be placed as symmetrically as possible with respect to the vertical line in the middle of the board.

6. The patient turns to look directly behind her/him over toward the right shoulder and then over the left shoulder.

7. The patient gets off the balance board (backwards) with the dominant foot first.

8. The operator clicks on the End button of the Logger GUI to conclude the data logging.

9. The operator can add further information pertaining to the exercise repetition, filling the field notes in the Logger GUI.

References

[1] M. E. Tinetti, S. K. Inouye, T. M. Gill, J. T. Doucette, Shared risk factors for falls, incontinence, and functional dependence: unifying the approach to geriatric syndromes, Jama 273 (17) (1995) 1348–1353.

[2] C. Bradley, J. E. Harrison, Hospitalisations due to falls in older people, Australia, 2003-04, Australian Institute of Health and Welfare Adelaide, 2007.

[3] R. G. Cumming, G. Salkeld, M. Thomas, G. Szonyi, Prospective study of the impact of fear of falling on activities of daily living, sf-36 scores, and nursing home admission, The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 55 (5) (2000) M299–M305.

[4] M. E. Tinetti, Preventing falls in elderly persons, New England journal of medicine 348 (1) (2003) 42–49.

[5] J. Kulmala, A. Viljanen, S. Sipilä, S. Pajala, O. Pärssinen, M. Kauppinen, M. Koskenvuo, J. Kaprio, T. Rantanen, Poor vision accompanied with other sensory impairments as a predictor of falls in older women, Age and ageing 38 (2) (2009) 162–167.

[6] R. Whipple, L. Wolfson, P. Amerman, The relationship of knee and ankle weakness to falls in nursing home residents: an isokinetic study, Journal of the American Geriatrics Society 35 (1) (1987) 13–20.

[7] M. E. Tinetti, M. Speechley, S. F. Ginter, Risk factors for falls among elderly persons living in the community, New England journal of medicine 319 (26) (1988) 1701–1707.

[8] J. Browne, N. O'Hare, Review of the different methods for assessing standing balance, Physiotherapy 87 (9) (2001) 489–495.

[9] L. Blum, N. Korner-Bitensky, Usefulness of the berg balance scale in stroke rehabilitation: a systematic review, Physical therapy 88 (5) (2008) 559–566.

[10] A. A. Qutubuddin, P. O. Pegg, D. X. Cifu, R. Brown, S. McNamee, W. Carne, Validating the berg balance scale for patients with parkinsons disease: a key to rehabilitation evaluation, Archives of physical medicine and rehabilitation 86 (4) (2005) 789–792.

[11] G. Juneja, J. J. Czyrny, R. T. Linn, Admission balance and outcomes of patients admitted for acute impatient rehabilitation, American journal of physical medicine & rehabilitation 77 (5) (1998) 388–393.

[12] D. Cattaneo, A. Regola, M. Meotti, Validity of six balance disorders scales in persons with multiple sclerosis, Disability and rehabilitation 28 (12) (2006) 789–795.

[13] D. L. Riddle, P. W. Stratford, Interpreting validity indexes for diagnostic tests: an illustration using the berg balance test, Physical therapy 79 (10) (1999) 939–948.

[14] A. Shumway-Cook, M. Baldwin, N. L. Polissar, W. Gruber, Predicting the probability for falls in community-dwelling older adults, Physical therapy 77 (8) (1997) 812–819.

[15] D. Bacciu, S. Chessa, C. Gallicchio, A. Micheli, E. Ferro, L. Fortunati, F. Palumbo, O. Parodi, F. Vozzi, S. Hanke, J. Kropf, K. Kreiner, Smart environments and context-awareness for lifestyle management in a healthy active ageing framework, in: Progress in Artificial Intelligence, Springer, 2015, pp. 54–66.

[16] F. Palumbo, D. La Rosa, E. Ferro, D. Bacciu, C. Gallicchio, A. Micheli, S. Chessa, F. Vozzi, O. Parodi, Reliability and human factors in ambient assisted living environments, Journal of Reliable Intelligent Environments (2017) 1–19.

[17] R. A. Clark, A. L. Bryant, Y. Pua, P. McCrory, K. Bennell, M. Hunt, Validity and reliability of the Nintendo Wii Balance Board for assessment of standing balance, Gait and Posture 31 (3) (2010) 307–310. doi:10.1016/j.gaitpost.2009.11.012.

[18] J. M. Leach, M. Mancini, R. J. Peterka, T. L. Hayes, F. B. Horak, Validating and calibrating the Nintendo Wii balance board to derive reliable center of pressure measures, Sensors (Switzerland) 14 (10) (2014) 18244–18267. doi:10.3390/s141018244.

[19] W. Young, S. Ferguson, S. Brault, C. Craig, Assessing and training standing balance in older adults: a novel approach using the nintendo wiibalance board, Gait & posture 33 (2) (2011) 303–305.

[20] J.-A. Gil-Gómez, R. Lloréns, M. Alcañiz, C. Colomer, Effectiveness of a Wii balance board-based system (eBaViR) for balance rehabilitation: a pilot randomized clinical trial in patients with acquired brain injury., Journal of neuroengineering and rehabilitation 8 (1) (2011) 30. doi:10.1186/1743-0003-8-30.
URL http://jneuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-8-30

[21] C. H. Shih, C. T. Shih, M. S. Chiang, A new standing posture detector to enable people with multiple disabilities to control environmental stimulation by changing their standing posture through a commercial Wii Balance Board, Research in Developmental Disabilities 31 (1) (2010) 281–286. doi:10.1016/j.ridd.2009.09.013.

[22] J. E. Deutsch, M. Borbely, J. Filler, K. Huhn, P. Guarrera-Bowlby, Use of a low-cost, commercially available gaming console (wii) for rehabilitation of an adolescent with cerebral palsy, Physical therapy 88 (10) (2008) 1196–1207.

[23] B. E. Maki, P. J. Holliday, A. K. Topper, A prospective study of postural balance and risk of falling in an ambulatory and independent elderly population, Journal of gerontology 49 (2) (1994) M72–M84.

[24] J. F. Kolen, S. C. Kremer, A field guide to dynamical recurrent networks, John Wiley & Sons, 2001.

[25] M. Lukoševičius, H. Jaeger, Reservoir computing approaches to recurrent neural network training, Computer Science Review 3 (3) (2009) 127–149.

[26] D. Verstraeten, B. Schrauwen, M. d'Haene, D. Stroobandt, An experimental unification of reservoir computing methods, Neural Networks 20 (3) (2007) 391–403.

[27] H. Jaeger, H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, Science 304 (5667) (2004) 78–80.

[28] H. Jaeger, The "echo state" approach to analysing and training recurrent neural networks, Tech. rep., GMD - German National Research Institute for Computer Science, Tech. Rep. (2001).

[29] C. Gallicchio, J. D. Martin-Guerrero, A. Micheli, E. Soria-Olivas, Randomized machine learning approaches: Recent developments and challenges, in: 25th European Symposium on Artificial Neural Networks (ESANN), i6doc.com, 2017, pp. 77–86.

[30] H. Simila, J. Mantyjarvi, J. Merilahti, M. Lindholm, M. Ermes, Accelerometry-based berg balance scale score estimation, Biomedical and Health Informatics, IEEE Journal of 18 (4) (2014) 1114–1121.

[31] P. Badura, E. Pietka, Automatic berg balance scale assessment system based on accelerometric signals, Biomedical Signal Processing and Control 24 (2016) 114–119.

[32] C. Gallicchio, A. Micheli, L. Pedrelli, L. Fortunati, F. Vozzi, O. Parodi, A Reservoir Computing Approach for Balance Assessment, Vol. 9785 of Lecture Notes in Computer Science, Springer International Publishing, 2016, pp. 65–77.

[33] Balance Assessment Dataset and Supplementary Material, http://www.di.unipi.it/groups/ciml/Data/balance.html, 2017.

[34] O. Alliance, Osgi service platform, release 3, IOS Press, Inc., 2003.

[35] F. Palumbo, J. Ullberg, A. Štimec, F. Furfari, L. Karlsson, S. Coradeschi, Sensor network infrastructure for a home care monitoring system, Sensors 14 (3) (2014) 3833–3860.

[36] F. Palumbo, D. La Rosa, S. Chessa, Gp-m: Mobile middleware infrastructure for ambient assisted living, in: Computers and Communication (ISCC), 2014 IEEE Symposium on, IEEE, 2014, pp. 1–6.

[37] J. A. Parker, R. V. Kenyon, D. E. Troxel, Comparison of interpolating methods for image resampling, IEEE Transactions on medical imaging 2 (1) (1983) 31–39.

[38] R. W. Schafer, What is a savitzky-golay filter?, Signal Processing Magazine, IEEE 28 (4) (2011) 111–117.

[39] F. Johnston, J. Boyland, M. Meadows, E. Shale, Some properties of a simple moving average when applied to forecasting a time series, Journal of the Operational Research Society (1999) 1267–1271.

[40] C. Gallicchio, A. Micheli, Architectural and markovian factors of echo state networks, Neural Networks 24 (5) (2011) 440–456.

[41] P. Tiňo, B. Hammer, M. Bodén, Markovian bias of neural-based architectures with feedback connections, in: Perspectives of neural-symbolic integration, Springer, 2007, pp. 95–133.

[42] M. Lukoševičius, H. Jaeger, B. Schrauwen, Reservoir computing trends, KI-Künstliche Intelligenz 26 (4) (2012) 365–371.

[43] H. Jaeger, M. Lukoševičius, D. Popovici, U. Siewert, Optimization and applications of echo state networks with leaky-integrator neurons, Neural Networks 20 (3) (2007) 335–352.

[44] C. Gallicchio, A. Micheli, P. Barsocchi, S. Chessa, User movements forecasting by reservoir computing using signal streams produced by moteclass sensors, in: Mobile Lightweight Wireless Systems, Springer, 2012, pp. 151–168.

[45] F. Palumbo, C. Gallicchio, R. Pucci, A. Micheli, Human activity recognition using multisensor data fusion based on reservoir computing, Journal of Ambient Intelligence and Smart Environments 8 (2) (2016) 87–107.

[46] G. Amato, D. Bacciu, S. Chessa, M. Dragone, C. Gallicchio, C. Gennaro, H. Lozano, A. Micheli, G. M. OHare, A. Renteria, C. Vairo, A benchmark dataset for human activity recognition and ambient assisted living, in: Ambient Intelligence-Software and Applications–7th International Symposium on Ambient Intelligence (ISAmI 2016), Springer, 2016, pp. 1–9.

[47] F. Palumbo, P. Barsocchi, C. Gallicchio, S. Chessa, A. Micheli, Multisensor data fusion for activity recognition based on reservoir computing, in: Evaluating AAL systems through competitive benchmarking, Springer, 2013, pp. 24–35.

[48] D. Bacciu, P. Barsocchi, S. Chessa, C. Gallicchio, A. Micheli, An experimental characterization of reservoir computing in ambient assisted living applications, Neural Computing and Applications 24 (6) (2014) 1451–1464.

[49] P. Barsocchi, S. Chessa, A. Micheli, C. Gallicchio, Forecast-driven enhancement of received signal strength (rss)-based localization systems, ISPRS International Journal of Geo-Information 2 (4) (2013) 978–995. doi:10.3390/ijgi2040978.

[50] M. Dragone, C. Gallicchio, R. Guzman, A. Micheli, Rss-based robot lo-

calization in critical environments using reservoir computing., in: Proceedings of the 24th European Symposium on Artificial Neural Networks (ESANN), i6doc.com, 2016, pp. 71–76.

[51] S. Chessa, C. Gallicchio, R. Guzman, A. Micheli, Robot localization by echo state networks using rss, in: Recent Advances of Neural Network Models and Applications, Springer, 2014, pp. 147–154.

[52] G. Amato, D. Bacciu, M. Broxvall, S. Chessa, S. Coleman, M. Di Rocco, M. Dragone, C. Gallicchio, C. Gennaro, H. Lozano, T. M. McGinnity, A. Micheli, A. K. Ray, A. Renteria, A. Saffiotti, D. Swords, C. Vairo, P. Vance, Robotic ubiquitous cognitive ecology for smart homes, Journal of Intelligent & Robotic Systems (2015) 1–25.

[53] M. Dragone, G. Amato, D. Bacciu, S. Chessa, S. Coleman, M. Di Rocco, C. Gallicchio, C. Gennaro, H. Lozano, L. Maguire, M. McGinnity, A. Micheli, G. M. OHare, A. Renteria, A. Saffiotti, C. Vairo, P. Vance, A cognitive robotic ecology approach to self-configuring and evolving aal systems, Engineering Applications of Artificial Intelligence 45 (2015) 269–280.

[54] C. Gallicchio, A. Micheli, Tree echo state networks, Neurocomputing 101 (2013) 319–337.

[55] C. Gallicchio, A. Micheli, A preliminary application of echo state networks to emotion recognition, in: Proceedings of EVALITA 2014, 2014, pp. 116–119.

[56] P. Tiňo, M. Čerňanský, L. Beňušková, Markovian architectural bias of recurrent neural networks, Neural Networks, IEEE Transactions on 15 (1) (2004) 6–15.

[57] S. Haykin, Neural networks and learning machines, Vol. 3, Pearson Education Upper Saddle River, 2009.

[58] A. Graves, A.-r. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2013, pp. 6645–6649.

[59] C. Wu, K.W. Chau, C. Fan, Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques, Journal of Hydrology 389 (1) (2010) 146–167.

[60] C. Paoli, C. Voyant, M. Muselli, M.-L. Nivet, Forecasting of preprocessed daily solar radiation time series using neural networks, Solar Energy 84 (12) (2010) 2146–2160.

[61] K.W. Chau, C.L. Wu, A hybrid model coupled with singular spectrum analysis for daily rainfall prediction, Journal of Hydroinformatics 12 (4) (2010) 458–473.

[62] D. V. Prokhorov, Training recurrent neurocontrollers for real-time applications, IEEE Transactions on Neural Networks 18 (4) (2007) 1003–1015.

[63] A. Waibel, Modular construction of time-delay neural networks for speech recognition, Neural computation 1 (1) (1989) 39–46.

[64] J. L. Elman, Finding structure in time, Cognitive science 14 (2) (1990) 179–211.

[65] J. L. Elman, Distributed representations, simple recurrent networks, and grammatical structure, Machine learning 7 (2-3) (1991) 195–225.

[66] A. F. Atiya, A. G. Parlos, New results on recurrent network training: unifying the algorithms and accelerating convergence, Neural Networks, IEEE Transactions on 11 (3) (2000) 697–709.

[67] K. Doya, Recurrent networks: Supervised learning, in: The handbook of brain theory and neural networks, MIT Press, 1998, pp. 796–800.

[68] D. L. Kornetti, S. L. Fritz, Y.-P. Chiu, K. E. Light, C. A. Velozo, Rating scale analysis of the berg balance scale, Archives of physical medicine and rehabilitation 85 (7) (2004) 1128–1135.

[69] M. Conradsson, L. Lundin-Olsson, N. Lindelöf, H. Littbrand, L. Malmqvist, Y. Gustafson, E. Rosendahl, Berg balance scale: intrarater test-retest reliability among older people dependent in activities of daily living and living in residential care facilities, Physical Therapy 87 (9) (2007) 1155–1163.