

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/329388524>

On-line Human Gait Stability Prediction using LSTMs for the fusion of Deep-based Pose Estimation and LRF-based Augmented Gait State Estimation in an Intelligent Robotic Rollator

Preprint · December 2018

DOI: 10.48550/arXiv.1812.00252

CITATIONS

0

READS

324

6 authors, including:



Petros Koutras

National Technical University of Athens

47 PUBLICATIONS 768 CITATIONS

SEE PROFILE



Jack Hadfield

National Technical University of Athens

10 PUBLICATIONS 124 CITATIONS

SEE PROFILE



Xanthi S. Papageorgiou

UBITECH

59 PUBLICATIONS 656 CITATIONS

SEE PROFILE



Costas Tzafestas

National Technical University of Athens

175 PUBLICATIONS 2,092 CITATIONS

SEE PROFILE

LSTM-based Network for Human Gait Stability Prediction in an Intelligent Robotic Rollator

Georgia Chalvatzaki, Petros Koutras, Jack Hadfield,

Xanthi S. Papageorgiou, Costas S. Tzafestas and Petros Maragos

School of Electrical and Computer Engineering, National Technical University of Athens, Greece

{gchal, xpapag}@mail.ntua.gr, {pkoutras, ktzaf, maragos}@cs.ntua.gr

Abstract—In this work, we present a novel framework for on-line human gait stability prediction of the elderly users of an intelligent robotic rollator using Long Short Term Memory (LSTM) networks, fusing multimodal RGB-D and Laser Range Finder (LRF) data from non-wearable sensors. A Deep Learning (DL) based approach is used for the upper body pose estimation. The detected pose is used for estimating the body Center of Mass (CoM) using Unscented Kalman Filter (UKF). An Augmented Gait State Estimation framework exploits the LRF data to estimate the legs' positions and the respective gait phase. These estimates are the inputs of an encoder-decoder sequence to sequence model which predicts the gait stability state as Safe or Fall Risk walking. It is validated with data from real patients, by exploring different network architectures, hyperparameter settings and by comparing the proposed method with other baselines. The presented LSTM-based human gait stability predictor is shown to provide robust predictions of the human stability state, and thus has the potential to be integrated into a general user-adaptive control architecture as a fall-risk alarm.

I. INTRODUCTION

A. Motivation

The worldwide population aged over 65 rises exponentially according to recent reports of the United Nations [1]. Amongst others the mobility problems prevail in the elder society. Ageing and many pathologies invoke changes in walking speed and stability [2], while 30% of the aged population is reported to have fallen every year. Especially, changes in gait speed are related to the functional independence and mobility impairment of the elderly [3], and are closely connected to fall incidents.

In the last 15 years robotics research has focused on robotic mobility assistive devices, aiming to provide postural support and walking assistance [4]–[8], as well as sensorial and cognitive assistance to the elderly [9]. Their goal is to increase the user mobility, while avoiding the anxiety and frustration imposed by the common walking aids. In our previous work [10], [11], we have shown that for a robotic rollator, that aims to support patients of different mobility status, user-adaptation is important. Specifically, a handy system should be able to assess the mobility state of the user and adapt the control strategies accordingly,

This research has been cofinanced by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH CREATE INNOVATE (project code:T1EDK- 01248 / MIS: 5030856).

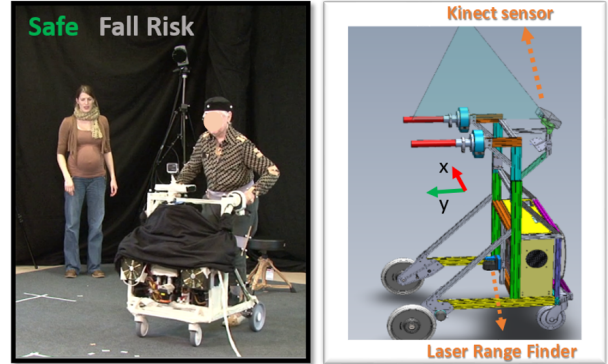


Fig. 1: **Left:** Elderly patient walking supported by a robotic rollator. The predicted states for Gait Stability may be Safe walking or Fall Risk. **Right:** a CAD drawing of the rollator with RGB-D and LRF sensors.

and also monitor the rehabilitation progress and provide fall prevention. Although computing various gait parameters and the on-line classification of the pathological gait status of the user plays a significant role in the adaptation of a context-aware controller, there are also issues of gait stability that should be addressed.

The purpose of this paper is to present, analyse and evaluate a novel and robust method for on-line gait stability analysis of elderly subjects walking with a mobility assistant platform (Fig. 1), fusing the information from an RGB-D camera that captures the upper body and a LRF monitoring the legs motion in the sagittal plane. Specifically, we use a DL approach for detecting the upper body pose and track the respective CoM through time along with an augmented gait state estimation from the LRF [12]. We propose a LSTM-based network for predicting the stability state of the patient, by classifying his gait as safe or fall-risk at each instance. We present a new method for predicting the gait stability state of robotic rollator users with variable pathological gait conditions, using only non-wearable sensors, in order to supply the robotic system with an alarm regarding fall risk episodes. The main goal is to integrate this information into a user-adaptive context-aware robot control architecture for the robotic assistant platform that would prevent possible falls.

B. Related Work

Fall detection and prevention is a hot topic in the field of assistive robotics [13]. Most of the proposed control

strategies for robotic assistive platforms in literature do not deal with the problem of fall prevention and research works focus on navigation and obstacle avoidance [14]–[16]. However, there exist some targeted research focusing on incorporating strategies for preventing or detecting fall incidents and facilitating user’s mobility. In [17], [18] the authors developed an admittance controller for a passive walker with a fall-prevention function considering the position and velocity of the user, utilizing data from two LRFs. They model the user as a solid body link, in order to compute the position of the center of gravity [19], based on which they applied a braking force on the rollator to prevent falls. A fall detection for a cane robot was presented in [20], [21], that computes the zero moment point stability of the elderly, using on-shoe sensors that provide ground force reactions.

Regarding the extraction of gait motions, different types of sensors have been used [22], [23]. Gait analysis can be achieved by using Hidden Markov Models for modelling normal [24] and pathological human gait [25], and extracting gait parameters [26]. Recently, we have developed a new method for online augmented human state estimation, that uses Interacting Multiple Model Particle Filters with Probabilistic Data Association [12], which tracks the users’ legs using data from a LRF, while it provides real-time gait phase estimation. We have also presented a new human-robot formation controller that utilizes the gait status characterization for user adaptation towards a fall preventing system [11].

Gait stability is mostly analysed by using wearable sensors [27], like motion markers placed on the human body to calculate the body’s CoM and the foot placements [28], and force sensors to estimate the center of pressure of the feet [29]. Gait stability analysis for walking aid users can be found in [30]. Regarding stability classification, an early approach can be found in [31], where the authors use the body skeleton provided by the RGB-D Kinect sensor as input and perform action classification to detect four classes of falling scenarios. However, the system was tested only with a physical therapist performing different walking problems.

Human pose estimation is a challenging topic due to the variable formations of the human body, the parts occlusions, etc. The rise of powerful DL frameworks along with the use of large annotated datasets opened a new era of research for optimal human pose estimation [32]. Most approaches provide solutions regarding the detection of the 2D pose from color images by detecting keypoints or parts on the human body [33], [34] achieving high accuracy. The problem of 3D pose estimation is more challenging [35], as the detected poses are scaled and normalized. Recent approaches aim to solve the ambiguity of 2D-to-3D correspondences by learning 3D poses from single color images [36], [37]. Another relevant research topic concerns the tracking of human poses [38], but while they achieve improved levels of accuracy compared to previous methods, due to the contribution of DL, the high estimation error makes it prohibitive to integrate it into a robotic application that requires high accuracy and robustness. A recent application of pose estimation for a robotic application can be found in [39].

Our contributions presented in this work is the design of a novel deep-based framework for on-line human gait stability state prediction using the detection of the upper body 3D pose and the respective CoM estimation from an RGB-D camera and the human gait states estimated from LRF data. Differing from the common gait stability analysis methods in literature, we propose an LSTM-based network for fusing the multi-modal information, in order to decode the hidden interaction of the body’s CoM with the legs motion and the gait phases, in order to predict the gait stability of the elderly considering two possible classes: safe and fall-risk walking. The proposed on-line LSTM-based human gait stability predictor is evaluated using multi-modal data from real patients. To justify the model selection, we present an exploratory study regarding the network architecture, the selected hyperparameters and compare the performance of our framework with baseline methods. The results demonstrate the great efficiency of the LSTM-based approach to provide robust predictions of the human stability state, and show its potential to be integrated into a general user-adaptive control architecture as a fall-risk alarm.

II. HUMAN GAIT STABILITY

During walking the body is in a continuous state of instability [40]. Gait stability is described by the interaction of the position and velocity of the CoM w.r.t. the Base of Support (BoS) in the horizontal plane. The BoS is the imaginary parallelogram formed by the contact of at least one foot with the ground. During double support phases the BoS covers its largest area, while in single leg support phases (one foot in stance phase while the other swings through) the BoS covers smaller areas. Biomechanics considers an inverted pendulum model to describe the CoM-BoS interaction [41].

When the projection of the CoM lies inside the BoS the body is stable. However, during gait the CoM is outside the BoS for the single leg support phases, for the most time in a gait cycle. Each foot contact, initiating a new gait cycle, prevents a potential fall [28]. One indicator of human stability is the distance of the CoM to the boundaries of the BoS, which is also the stability measure used in this study.

When the CoM is inside the BoS their respective distance is called stability margin, and when the CoM lies outside the BoS the distance is called CoM separation. The measures of these distances are indicative of the stability of a person while walking. Although a human supported by a walking aid has an enlarged BoS, the reported high fall incidents of walking aid users [42], along with the fact that users often disengage their hands from the aid to perform several actions, led us to consider the general notion of the human-centric BoS in this particular paper.

III. METHOD

In Fig. 2 an overview of the proposed LSTM-based human gait stability prediction framework is presented. The proposed method uses multimodal RGB-D and LRF data. The RGB-D data are employed for a deep-based pose detection and the estimation of the CoM position. The LRF data are used in an augmented gait state estimation framework

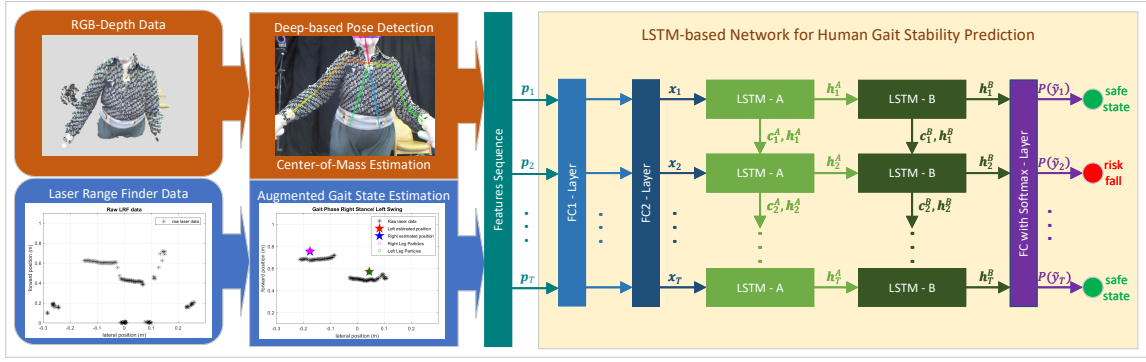


Fig. 2: Overview of the proposed LSTM-based On-line Human Gait Stability Prediction framework.

for estimating the legs' position and the gait phase. These estimated human-motion related parameters constitute the features of the LSTM-based Network for predicting the gait stability state at each time instance, using a binary description of the user's state as "Safe" or "Fall Risk" state. The components of the whole framework are described below.

A. Augmented Gait State Estimation

The augmented human gait state estimation was proposed in [12]. It is a novel framework for efficient and robust leg tracking from LRF data along with human gait phases predictions (Fig. 2). Our approach uses two Particle Filters (PFs) and Probabilistic Data Association (PDA) with an Interacting Multiple Model (IMM) scheme for a real-time selection of the appropriate motion model of the PFs. The IMM is a first-order Markov model with the following gait states: a) Left Double Support (DS), b) Left Stance/ Right Swing, c) Right DS, d) Right Stance/ Left Swing. Each state refers to both legs and imposes a different motion model for the PFs. The legs are tracked, and a maximum likelihood estimation predicts the current gait phase. A thorough analysis of the methodology is provided in [12]. From this method the augmented gait state $\mathbf{s}_t^{\text{gait}} = [x^l \ y^l \ x^r \ y^r \ \text{phase}]^T$, consisting of the estimated legs' positions along the axes (l, r refer to left and right leg respectively) and the respective gait phase, is used in the LSTM-based network (Fig. 2).

B. Deep-based Pose Detection and CoM Estimation

For the upper body pose estimation we employ the RGB images and the respective depth maps provided by the Kinect sensor, which is mounted on the rollator (Fig. 1). The 2D positions of the keypoints are detected using the Open Pose (OP) Library [34] with fixed weights. OP uses a bottom-up representation of associations of the locations and orientations of the limbs through the images and a two branch multi-stage convolutional neural network to predict the keypoints' 2D positions. The third dimension of the keypoints is obtained by depth maps. The depth maps need to be transformed to the image plane using the calibration matrix of the camera. For this purpose we have applied the method of [39].

Despite the high performance of the OP framework, the close proximity of the human body to the Kinect sensor, the occlusions of body parts, such as the head or large part of the

arms, from the camera's field of view (Fig. 2), or even the high reflectivity due to ambient light, lead to many detection losses and misdetections. Thus, tracking is required. When the pose is detected, we use the torso keypoints to compute the 3D position of the torso center, as the median of the keypoints. From the experimental analysis of the motion markers, which are used as Ground Truth (GT), we have modelled the upper body as ellipsoid with its center being the CoM. We use statistics about the ellipsoid fitting to translate the torso center in a position representing the CoM. Finally, the detected CoM is transformed from the camera coordinate frame to the robot frame.

The detected CoM positions are the observations of an UKF that tracks and estimates the CoM state through time. The UKF serves many purposes; it is used to model and predict the nonlinear CoM motion [43], [44], it filters the noisy observations, it compensates the different frame rates between the Kinect and the LRF sensor by giving predictions of the CoM states during the periods when the sensor does not transmit measurements, and also provides estimates when we do not get/accept the corrupted keypoints detections.

The tracking framework employs the well-known prediction and update equations of the UKF, described in [45]. The CoM motion model includes the following kinematic equations:

$$\begin{aligned} q_t^x &= q_{t-1}^x + v_{t-1} \cdot \cos(\omega_{t-1}^z \cdot \Delta t) \cdot \Delta t \\ q_t^y &= q_{t-1}^y + v_{t-1} \cdot \sin(\omega_{t-1}^z \cdot \Delta t) \cdot \Delta t \\ v_t &= v_{t-1} + \eta_v \\ \omega_t^z &= \omega_{t-1}^z + \eta_{\omega^z} \end{aligned} \quad (1)$$

where t is the discrete time, q_t^x, q_t^y the position along the axes, v_t, ω_t^z the linear and angular velocities, Δt is the time interval in which we make predictions and η_v, η_{ω^z} are the linear and angular velocity noises modelled as zero-mean white Gaussian with standard deviations $\sigma_v = 0.98$ m/sec and $\sigma_{\omega^z} = 1.88$ rad/sec respectively, computed from the motion markers data.

The linear observation model in UKF considers only the detected positions q_t^x, q_t^y of the CoM. In the observation model the noise vectors are modelled as white Gaussians with standard deviations 0.15 m for the q^x variable and 0.2 m for the q^y variable, learned from experiments (higher variability in measuring the q^y from the camera's depth map). Only the estimated CoM position from the UKF is fed to the LSTM-based network (Fig. 2); let us denote it as $\mathbf{s}_t^{\text{CoM}} = [q^x \ q^y]^T$.

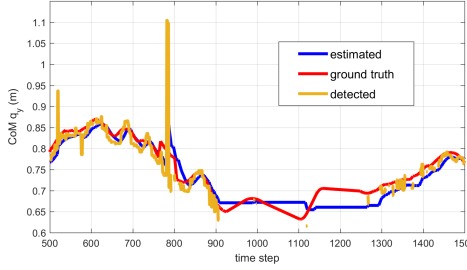


Fig. 3: CoM forward displacement.

C. LSTM-based Network for Gait Stability Prediction

In our learning based method for gait stability prediction we employ a Neural Network (NN) architecture based on LSTM units [46]. LSTM constitutes a recurrent NN that can effectively learn long-term dependencies by incorporating memory cells \mathbf{c}_t that allow the network to learn when to forget previous hidden states and when to update hidden states given new information. The overall architecture is an encoder-decoder sequence-to-sequence model, considering only past input vectors to make predictions. It consists of two Fully Connected (FC) layers, two LSTM layers and a last FC layer followed by Softmax (Fig. 2).

Input representation: Let $\mathbf{p}_t = \begin{bmatrix} \mathbf{s}_t^{CoM} & \mathbf{s}_t^{gait} \end{bmatrix}$, be the standardised observations (with zero mean and unit variance) at each time instant t and $\{\mathbf{p}_t\}_1^T$ the sequence of our observations in a temporal window of length T . These observations are transformed to the LSTM inputs $\{\mathbf{x}_t\}_1^T$ by feeding them to a network with two FC layers:

$$\mathbf{x}_t = \rho(\mathbf{W}_{FC2} \cdot \rho(\mathbf{W}_{FC1} \cdot \mathbf{p}_t + \mathbf{b}_{FC1}) + \mathbf{b}_{FC2}), \quad (2)$$

where \mathbf{W}_{FC1} , \mathbf{W}_{FC2} and \mathbf{b}_{FC1} , \mathbf{b}_{FC2} are the weight and the biases of the two linear layers and $\rho(\cdot)$ is a Rectified Linear Unit (ReLU) nonlinearity, defined as $\rho(\mu) = \max(0, \mu)$. These two FC layers have the role of an encoder of the features, helping to encode the nonlinear function of the CoM with the legs for detecting instability. In this way, we learn a static transformation and find a better representation for the observations before feeding them to the LSTM unit that models time dependencies.

LSTM Unit: LSTM is composed of an input gate \mathbf{i} , an input modulation gate \mathbf{g} , a memory cell \mathbf{c} , a forget gate \mathbf{f} and an output gate \mathbf{o} . LSTM takes the computed inputs \mathbf{x}_t at each time step t , the previous estimations for the hidden state \mathbf{h}_{t-1} and the memory cell state \mathbf{c}_{t-1} , in order to update their states using the equations:

$$\begin{aligned} \mathbf{i}_t &= \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i) \\ \mathbf{f}_t &= \sigma(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f) \\ \mathbf{o}_t &= \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o) \\ \mathbf{g}_t &= \phi(\mathbf{W}_{xg}\mathbf{x}_t + \mathbf{W}_{hg}\mathbf{h}_{t-1} + \mathbf{b}_g) \\ \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \\ \mathbf{h}_t &= \mathbf{o}_t \cdot \phi(\mathbf{c}_t) \end{aligned} \quad (3)$$

where we have the equations of the four gates: input gate $\mathbf{i}_t \in \mathbb{R}^N$, forget gate $\mathbf{f}_t \in \mathbb{R}^N$, output gate $\mathbf{o}_t \in \mathbb{R}^N$, input modulation gate $\mathbf{g}_t \in \mathbb{R}^N$ that modulates the memory

TABLE I: Demographics

subject	Sex	Age	Height (cm)	Weight (kg)
1	F	80	153,5	64,1
2	F	77	164	89,5
3	F	80	140,5	73,1
4	M	85	170	75
5	M	81	178	61,4

TABLE II: Average RMSE for the CoM estimation

subject	1	2	3	4	5	mean
average RMSE (cm)	2,65	6,13	1,2	6,09	2,5	3,71

cell $\mathbf{c}_t \in \mathbb{R}^N$ and the hidden state $\mathbf{h}_t \in \mathbb{R}^N$ with N hidden units. Symbol \odot represents element-wise multiplication, the function $\sigma: \mathbb{R} \rightarrow [0, 1]$, $\sigma(\mu) = \frac{1}{1+e^{-\mu}}$ is the sigmoid non-linearity and $\phi: \mathbb{R} \rightarrow [-1, 1]$, $\phi(\mu) = \frac{e^{\mu} - e^{-\mu}}{e^{\mu} + e^{-\mu}}$ is the hyperbolic tangent non-linearity. \mathbf{W}_{xn} , \mathbf{W}_{hn} , with $n = \{i, f, o, g\}$ are the weight matrices of the input and recurrent connection of each gate and \mathbf{b}_n denotes the bias vectors for each gate. The parameters of the \mathbf{W}_{xn} , \mathbf{W}_{hn} and \mathbf{b}_n are learned during the training of the NN.

In the last layer (Fig. 2), we estimate the classes of gait stability y_t at each time step t , $y_t \in [0, 1]$ with “0” being the Safe class and “1” the Fall Risk class, by learning a linear transformation from the hidden states \mathbf{h}_t to the output state $\tilde{\mathbf{y}}_t$, described by: $\tilde{\mathbf{y}}_t = \mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y$, where again \mathbf{W}_{hy} is the weight matrix and \mathbf{b}_y the bias vector of the output layer. Then, the probability of having a Fall Risk gait is given by taking the softmax:

$$P(y_t = 1 | \{\mathbf{p}_\tau\}_1^t; \mathbf{W}, \mathbf{b}) = \frac{e^{\tilde{y}_{t,1}}}{e^{\tilde{y}_{t,0}} + e^{\tilde{y}_{t,1}}}, \quad (4)$$

where \mathbf{W}, \mathbf{b} denote the trainable parameters of the whole network and $\{\mathbf{p}_\tau\}_1^t$ are the observations until time t .

IV. EXPERIMENTAL ANALYSIS AND RESULTS

A. Experimental Setup

Data Collection: The data used in this work were collected in Agaplesion Bethanien Hospital - Geriatric Center with the participation of real patients. The participants presented moderate to mild mobility impairment, according to clinical evaluation. Table I presents the participants’ demographics. The subjects walked with physical support of a passive robotic rollator, used for the purpose of data collection, in a specific hospital area, while wearing a set of MoCap markers. The visual markers data are used for ground truth extraction. The data for the Gait Stability Prediction were provided by a Kinect v1 camera for capturing the upper body, and a Hokuyo rapid LRF for detecting the legs, which were mounted on the rollator (Fig. 1).

Ground truth extraction from visual markers: For extracting the GT labels of the Gait Stability Prediction framework, we followed the methodology described in [28] for computing the CoM from visual markers. In our previous work [47] we have thoroughly described the process of detecting the gait phases from the planar foot impact. As explained in Sec. II for specifying the body stability we have to analyze the CoM-BoS interaction. Therefore, we detect

TABLE III: Performance results of different architectures of LSTM-based Prediction Models

Test dataset:	AUC						FScore						Accuracy					
	1	2	3	4	5	mean	1	2	3	4	5	mean	1	2	3	4	5	mean
window size: T=100																		
LSTM, $\ell = 1^1$, N=256	71,78	65,82	96,67	48,29	53,77	66,87	56,25	65,10	80,74	52,71	63,09	63,58	60,45	55,26	83,20	43,73	52,02	58,93
FC1+FC2+LSTM, $\ell = 1$, N=256	96,33	69,31	98,88	94,63	92,97	90,61	86,75	78,17	90,42	85,45	89,64	86,13	89,48	67,59	92,77	83,64	85,53	83,60
FC1+FC2+LSTM, $\ell = 2$, N=128	96,51	66,98	99,01	94,20	93,33	90,01	87,95	80,42	89,52	86,10	89,96	86,79	90,83	69,69	91,70	83,35	86,25	84,36
FC1+FC2+LSTM, $\ell = 2$, N=256	96,36	67,50	99,08	94,69	93,79	90,28	87,71	79,05	89,76	86,67	89,21	86,48	90,51	68,24	91,92	83,90	85,44	84,00

¹ ℓ denotes the layers

TABLE IV: Exploration of the sequence window size effect

Test dataset	AUC						FScore						Accuracy					
	1	2	3	4	5	mean	1	2	3	4	5	mean	1	2	3	4	5	mean
FC1+FC2+LSTM, $\ell = 2$, N=128																		
window size: T=50	97,26	63,46	98,79	93,90	92,78	89,24	86,97	77,74	88,07	87,84	88,90	85,09	89,43	66,35	90,38	85,13	85,03	83,26
window size: T=100	96,51	66,98	99,01	94,20	93,33	90,01	87,95	80,42	89,52	86,10	89,96	86,79	90,83	69,69	91,70	83,35	86,25	84,36
window size: T=200	96,11	68,49	98,99	91,99	93,02	89,72	86,56	79,57	90,33	83,06	89,50	85,80	89,53	68,80	92,42	79,84	85,53	83,22

the BoS according to the respective gait phase and measure the distance of the planar CoM position and evaluate the respective margins of stability. Given the respective average measures of the stability margins found in literature [28], we label each walking instance as Safe or Fall Risk. Those two states constitute the ground truth labels used for training the proposed network. We have, also, conducted statistical analysis on the CoM motion for tuning the UKF used in our framework. Further, we found the statistics of the BoS size and position w.r.t. the users’ tibia (the level at which the LRF scans the user’s legs) which were used for a baseline rule-based method described below.

Dataset statistics & Data Augmentation: We use a dataset containing data from five patients, resulting in about 11000 frames corresponding to more than 300sec of walking. The dataset consists of 73% safe states, leading to a largely unbalanced dataset. To avoid overfitting, we applied data augmentation, by applying on the GT CoM data of the Safe states additive random noise exploiting the statistics extracted for the CoM position by initial experimentation, i.e. zero mean and standard deviations $\sigma_x = 0,15cm$ in the x-direction and $\sigma_y = 0,20cm$ in the y-direction, thus increasing the chances of detecting Fall Risk states. The same noise vector was also applied on the pose-based CoM. The final dataset consists of about 22000 frames, with about 46% Fall-Risk labels. We employ a leave-one-out strategy for training/testing, i.e. data from four subjects in training and one in testing iteratively for cross-validation.

B. Evaluation Strategy

Evaluation Metrics: For the evaluation of the predicted gait stability labels, we present a thorough analysis regarding different LSTM architectures and other baseline methods by reporting the FScore, Accuracy, Precision and Recall metrics. We also evaluate the Area Under Curve (AUC), which is defined as the area under the Receiver Operating Characteristic (ROC) curve [48], [49]. The ROC curve relates true positive rates w.r.t. false positive rates at different classification thresholds, while AUC helps evaluating the classifier’s performance across all possible classification thresholds. We also demonstrate a brief validation of the CoM position estimation w.r.t. the respective extracted ground truth CoM

employing the Root Mean Square Error (RMSE).

Baseline Methods: As baseline methods we use the nonlinear Support Vector Machine (SVM) classifier with Gaussian kernel and the same inputs as the LSTM-based method, as well as a rule-based approach. The rule-based approach follows the biomechanics rules also employed for the ground truth extraction. From the augmented gait state estimation we get the leg’s positions (at knee height) and the respective gait phase. We define the BoS in the horizontal plane employing the statistics learned by the visual markers analysis. Subsequently, we apply the same thresholds for evaluating the stability margins, which were used in the ground truth labels extraction.

Implementation: We trained from scratch the proposed network with $N = 128$ hidden layers in PyTorch using an Nvidia Titan X GPU. The output units of the FC layers FC1, FC2 were $4 \cdot N$ and $2 \cdot N$ respectively. For training we employed the Adam optimizer with the Logistic Cross-Entropy Loss for binary classification:

$$\mathcal{L}(\mathbf{W}, \mathbf{b}) = - \sum_{j \in Y_+} \log P(y_t^j = 1 | \mathbf{Z}; \mathbf{W}, \mathbf{b}) - \sum_{j \in Y_-} \log P(y_t^j = 0 | \mathbf{Z}; \mathbf{W}, \mathbf{b}) \tag{5}$$

where \mathbf{Z} are the batched training samples, $y_t^j \in \{0, 1\}$ is the binary gait stability label of \mathbf{Z} , and Y_+ and Y_- are the safe and fall risk labelled sample sets respectively. $P(\cdot)$ is obtained by the softmax activation of the final layer (Eq. 4). We used mini-batches of 256 clips, with initial learning rate 0.001, momentum 0.9 and weight decay 0.01. The learning rate is divided by 10 after half of the epochs. For better regularization and faster training we also have applied dropout with probability 0.7 in the FC1 layer and batch normalization [50] after the first two FC layers and the LSTM unit. All subsystems were integrated in ROS [51].

C. Experimental Results

Validation of CoM estimation: Table II reports the average position RMSE for the CoM w.r.t. the ground truth CoM positions extracted by the visual markers. Only subjects #2 and #4 present higher errors with RMSE 6cm. Fig. 3 depicts the forward CoM displacement, as this is estimated by the UKF (blue line) given the detected CoM from the pose

TABLE V: Comparison with baseline methods

Method	Rule-based	SVM	LSTM	FC1+FC2+LSTM
Precision	83,27	80,59	66,87	88,84
Recall	77,29	88,11	64,36	85,65
FScore	80,09	83,22	63,58	86,79
Accuracy	80,81	80,62	67,27	84,36
AUC	-	86,73	58,93	90,01

(orange line) w.r.t. the extracted ground truth (red line). We aim to show the difficulty of the dataset, since the ambiguity in the proximal pose detection leads to many occlusions and misdetections which are handled well by the UKF approach. **Evaluation of the LSTM-based prediction model:** Table III explores the performance of different network architectures for fixed window size sequences of 100 instances across the whole dataset using the metrics AUC, FScore and Accuracy. We observe that the plain architecture of a simple 1-layer LSTM with N=256 hidden states cannot capture the complexity of the body motion and decode the underlying interaction of the CoM with the legs motion per gait phase. On the contrary, the other models that use the two fully connected layers FC1 and FC2 before the LSTM cells can encode the hidden parameters of the whole body motion, achieving high accuracy. More importantly all three architectures achieve high AUC over 90%, while the FScores are over 86% meaning that they can be used for providing precise predictions about the safety of the rollator user. As for per user performance, it is evident that all network architectures perform worse when predicting subject’s #2 stability state. This poor performance is highly influenced by the higher CoM estimation errors as can be seen in Table II, or even a kind of pathological gait unseen to the system, since we only had available data from five patients. However, the existence of patient #2 data in the other four training datasets did not influence the performance; indeed the deep networks could decode the different types of walking and achieve high performance in spite of the estimation noise and errors from the tracking systems.

From this exploratory study, we choose as our main network architecture the FC1+FC2+LSTMs with 2 layers of N=128 hidden variables, as it achieves the best FScore and Accuracy rate, while hitting a 90% mean AUC. In Table IV, we explore the influence of the temporal window size on the proposed model. We cross-evaluate the results for all datasets and present the AUC, FScore and Accuracy measures. We observe that the most consistent performance across all metrics is the one for input sequences of 100 instances.

The final part of the experimental evaluation presents the cross-examination of the proposed model w.r.t. baseline methods. As described previously, we compare the LSTM-based network with the rule-based method, an SVM classifier and for the sake of generality with the simple LSTM (no initial FC layers) of Table III. We evaluate the average metrics across all datasets. Inspecting the results, the rule based method achieves an accuracy of 80% and an analogous FScore. This finding is a strong indicator that our analysis about predicting the stability state by fusing the data from

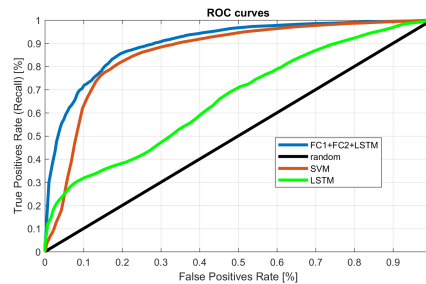


Fig. 4: ROC curve of the SVM, LSTM and fully connected LSTM along with the chance predictor.

the body pose and the gait state estimates is plausible. The rule-based method is a discrete process that does not include probabilities computation like the rest of the methods; thus the AUC could not be evaluated.

On the other hand, we notice that the SVM classifier performs very well. This was well expected, since the SVM classifier with nonlinear kernel is known to work well in binary classification problems with relatively small datasets. However, the proposed network improves the SVM scores as it achieves about 4% better FScore and ameliorates the AUC score by approximately 3%. The plain LSTM model achieves again the poorest results, as in Table III, since it is evident that there is an underlying nonlinear relation between the CoM and the gait states, which is encoded by the fully connected layers. For better understanding of the results of this table, we plot the ROC curve for the SVM, LSTM, the proposed network, and the random predictor across all possible classification thresholds. Although, all methods perform better than random predictor, the proposed fully connected LSTM-based network outperform all the other methods at all cases.

V. CONCLUSIONS & FUTURE WORK

In this work, we have presented and experimentally evaluated a novel and robust method for on-line gait stability analysis of elderly subjects walking with a mobility assistant platform. The proposed method is fusing the information from an RGB-D camera which captures the upper body and a LRF which monitors the legs motion. We use the OP framework for detecting the upper body pose, we track the respective CoM, while we exploit the LRF data in an augmented gait state estimation framework for extracting the legs’ position and the respective gait phase. Our main contribution is the proposal of a novel LSTM-based network for predicting the stability state of the elderly, by classifying walking as safe or fall-risk at each instant.

In the future, we plan to increase our datasets for training and evaluation of the proposed model and integrate it into a user-adaptive context-aware robot control architecture with a fall-prevention functionality for an intelligent robotic assistant platform.

REFERENCES

- [1] T. W. Bank, "Population ages 65 and above," 2017.
- [2] J. M. Hausdorff, "Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking," *Human Movement Science* 2007.
- [3] F. Garcia-Pinillos et al., "Gait speed in older people: an easy test for detecting cognitive impairment, functional independence, and health state," *Psychogeriatrics*, 2016.
- [4] S. D. et.al., "Pamm - a robotic aid to the elderly for mobility assistance and monitoring: A 'helping-hand' for the elderly," in *IEEE Int'l Conf. on Robotics and Automation*, 2000, pp. 570–576.
- [5] B. Graf, M. Hans, and R. D. Schraft, "Mobile robot assistants," *IEEE Robotics Automation Magazine*, vol. 11, no. 2, pp. 67–77, June 2004.
- [6] D. Rodriguez-Losada, F. Matia, A. Jimenez, R. Galan, and G. Lacey, "Implementing map based navigation in guido, the robotic smart-walker," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, April 2005, pp. 3390–3395.
- [7] V. Kulyukin, A. Kutiyawala, E. LoPresti, J. Matthews, and R. Simpson, "iwalker: Toward a rollator-mounted wayfinding system for the elderly," in *2008 IEEE International Conference on RFID*, April 2008, pp. 303–311.
- [8] T. Ohnuma et. al., "Particle filter based lower limb prediction and motion control for jaist active robotic walker," in *2014 RO-MAN*.
- [9] S. Jenkins et. al., "Care, monitoring, and companionship: Views on care robots from older people and their carers," *International Journal of Social Robotics*, 2015.
- [10] G. Chalvatzaki, X. S. Papageorgiou, and C. S. Tzafestas, "Towards a user-adaptive context-aware robotic walker with a pathological gait assessment system: First experimental study," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2017, pp. 5037–5042.
- [11] —, "User-adaptive human-robot formation control for an intelligent robotic walker using augmented human state estimation and pathological gait characterization," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018.
- [12] G. Chalvatzaki, X. S. Papageorgiou, C. S. Tzafestas, and P. Maragos, "Augmented human state estimation using interacting multiple model particle filters with probabilistic data association," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1872–1879, July 2018.
- [13] Y. S. Delahoz and M. A. Labrador, "Survey on fall detection and fall prevention using wearable and external sensors," *Sensors*, vol. 14, no. 10, pp. 19806–19842, 2014. [Online]. Available: <http://www.mdpi.com/1424-8220/14/10/19806>
- [14] O. Chuy et. al., "Environment feedback for robotic walking support system control," in *2007 IEEE International Conference on Robotics and Automation (ICRA)*, April 2007, pp. 3633–3638.
- [15] C. A. Cifuentes and A. Frizera, *Development of a Cognitive HRI Strategy for Mobile Robot Control*. Springer International Publishing, 2016.
- [16] M. Geravand et. al., "An integrated decision making approach for adaptive shared control of mobility assistance robots," *Int'l J. of Social Robotics*, 2016.
- [17] Y. Hirata et. al., "Motion control of intelligent passive-type walker for fall-prevention function based on estimation of user state," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*.
- [18] Y. Hirata, S. Komatsuda, and K. Kosuge, "Fall prevention control of passive intelligent walker based on human model," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2008, pp. 1222–1228.
- [19] M. Takeda, Y. Hirata, K. Kosuge, T. Katayama, Y. Mizuta, and A. Koujina, "Human cog estimation for assistive robots using a small number of sensors," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 6052–6057.
- [20] P. Di, J. Huang, S. Nakagawa, K. Sekiyama, and T. Fukuda, "Fall detection and prevention in the elderly based on the zmp stability control," in *2013 IEEE Workshop on Advanced Robotics and its Social Impacts*, Nov 2013, pp. 82–87.
- [21] P. Di, Y. Hasegawa, S. Nakagawa, K. Sekiyama, T. Fukuda, J. Huang, and Q. Huang, "Fall detection and prevention control using walking-aid cane robot," *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 2, pp. 625–637, April 2016.
- [22] J. Bae and M. Tomizuka, "Gait phase analysis based on a hidden markov model," *Mechatronics*, 2011.
- [23] C. Nickel et. al., "Using hidden markov models for accelerometer-based biometric gait recognition," in *CSPA 2011*.
- [24] X. S. Papageorgiou, G. Chalvatzaki, C. S. Tzafestas, and P. Maragos, "Hidden markov modeling of human normal gait using laser range finder for a mobility assistance robot," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 482–487.
- [25] —, "Hidden markov modeling of human pathological gait using laser range finder for an assisted living intelligent robotic walker," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 6342–6347.
- [26] X. S. Papageorgiou, G. Chalvatzaki, K. N. Lianos, C. Werner, K. Hauer, C. S. Tzafestas, and P. Maragos, "Experimental validation of human pathological gait analysis for an assisted living intelligent robotic walker," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechanics (BioRob)*, June 2016, pp. 1086–1091.
- [27] S. Page, K. Mun, Z. Guo, F. A. Reyes, H. Yu, and V. Pasqui, "Unbalance detection to avoid falls with the use of a smart walker," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechanics (BioRob)*, June 2016, pp. 1000–1005.
- [28] V. Lugade, V. Lin, and L.-S. Chou, "Center of mass and base of support interaction during gait," *Gait and Posture*, vol. 33, no. 3, pp. 406 – 411, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S096663621000442X>
- [29] C.-W. Huang, P.-D. Sue, M. F. Abbod, B. C. Jiang, and J.-S. Shieh, "Measuring center of pressure signals to quantify human balance using multivariate multiscale entropy by designing a force platform," *Sensors*, vol. 13, no. 8, pp. 10 151–10 166, 2013. [Online]. Available: <http://www.mdpi.com/1424-8220/13/8/10151>
- [30] E. Costamagna, S. Thies, L. Kenney, D. Howard, A. Liu, and D. Ogden, "A generalisable methodology for stability assessment of walking aid users," *Medical Engineering and Physics*, vol. 47, pp. 167 – 175, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S135045317301510>
- [31] S. Taghvaei, Y. Hirata, and K. Kosuge, "Visual human action classification for control of a passive walker," in *2017 7th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO)*, April 2017, pp. 1–5.
- [32] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 3686–3693.
- [33] S. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 4724–4732.
- [34] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 1302–1310.
- [35] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A simple yet effective baseline for 3d human pose estimation," in *ICCV*, 2017.
- [36] G. Pavlakos, L. Zhu, X. Zhou, and K. Daniilidis, "Learning to estimate 3D human pose and shape from a single color image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [37] G. Pavlakos, X. Zhou, and K. Daniilidis, "Ordinal depth supervision for 3D human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [38] R. Girdhar, G. Gkioxari, L. Torresani, M. Paluri, and D. Tran, "Detect-and-Track: Efficient Pose Estimation in Videos," in *CVPR*, 2018.
- [39] C. Zimmermann, T. Welschhold, C. Dornhege, W. Burgard, and T. Brox, "3d human pose estimation in rgbd images for robotic task learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018. [Online]. Available: <https://lmb.informatik.uni-freiburg.de/projects/rgbd-pose3d/>
- [40] S. M. Bruijn and J. H. van Dieën, "Control of human gait stability through foot placement," *Journal of The Royal Society Interface*, vol. 15, no. 143, 2018. [Online]. Available: <http://rsif.royalsocietypublishing.org/content/15/143/20170816>
- [41] Y. chung Pai and J. Patton, "Center of mass velocity-position predictions for balance control," *Journal of biomechanics*, vol. 30 4, pp. 347–54, 1997.
- [42] T. R. de Mettelinge and D. Cambier, "Understanding the relationship between walking aids and falls in older adults: a prospective cohort study," *Journal of geriatric physical therapy*, vol. 38 3, pp. 127–32, 2015.

- [43] M. Jacquelin Perry, *Gait Analysis: Normal and Pathological Function*, 1st ed., C. Bryan Malas, MHPE, Ed. SLACK Incorporated, 1992.
- [44] N. Belloto and H. Hu, "People tracking with a mobile robot: A comparison of kalman and particle filters." IASTED 2007.
- [45] E. A. Wan et al., "The unscented kalman filter for nonlinear estimation," in *Proc. of IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.
- [46] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [47] G. Chalvatzaki, X. S. Papageorgiou, C. S. Tzafestas, and P. Maragos, "Estimating double support in pathological gaits using an hmm-based analyzer for an intelligent robotic walker," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug 2017, pp. 101–106.
- [48] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 233–240.
- [49] P. Koutras and P. Maragos, "A perceptually based spatio-temporal computational framework for visual saliency estimation," *Signal Processing: Image Communication*, vol. 38, pp. 15 – 31, 2015, recent Advances in Saliency Models, Applications and Evaluations. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0923596515001290>
- [50] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, pp. 448–456. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3045118.3045167>
- [51] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.