On The Feasibility of Multimodal Dialog Based Remote Balance Assessment

Nikhil Sukhdev*, Oliver Roesler*, Michael Neumann, Meredith Bartlett, Doug Habberstad, and Vikram Ramanarayanan

Abstract A variety of tests exist to assess balance, which are usually administered by a clinician in clinic. This limits how often they can be administered and requires patients to travel to the clinic, which can often be inconvenient or difficult, especially in cases involving movement disorders. To address this issue, we evaluate the feasibility of using a multimodal dialog-based platform for remote balance assessment. The platform uses the participants' webcams to track their body poses while a virtual agent guides them through a series of tasks to assess balance inspired by the Berg Balance Scale (BBS). We recorded 62 assessment sessions from 58 healthy participants to evaluate whether participants were able to perform the tasks as instructed and to evaluate whether the obtained data can be used to automatically extract analytically valid metrics to assess balance. The results show that the assessment is feasible but participant compliance with instructions is crucial to ensure good overall data quality. Furthermore, we show that the time taken to stand from a seated position (TSS) can be accurately calculated if participants comply with instructions to perform the task correctly.

N. Sukhdev
Modality AI, Inc., San Francisco, USA, e-mail: nikhil.sukhdev@modality.ai

O. Roesler
Modality AI, Inc., San Francisco, USA, e-mail: oliver.roesler@modality.ai

M. Neumann
Modality AI, Inc., San Francisco, USA, e-mail: michael.neumann@modality.ai

M. Bartlett
Modality AI, Inc., San Francisco, USA, e-mail: meredith.bartlett@modality.ai

D. Habberstad
Modality AI, Inc., San Francisco, USA, e-mail: doug.habberstad@modality.ai

V. Ramanarayanan
Modality AI, Inc., San Francisco, USA, e-mail: vikram.ramanarayanan@modality.ai

* These authors contributed equally.

Key words: multimodal dialog system, remote patient monitoring, balance assessment

1 Introduction

Balance impairment reduces mobility, independence, and overall the quality of life of affected people and places a higher burden on caregivers and the health system (Atteya et al, 2019). Furthermore, postural instability is a major cause of falls for a variety of neurological conditions, like Parkinson's Disease (PD) (Wood et al, 2002), Huntington's Disease (Busse et al, 2009), or Multiple Sclerosis (Coote et al, 2020), leading to twice as many falls in people with neurological conditions than age-matched healthy controls (Stolze et al, 2004). A variety of interventions exist that can be applied to reduce the risk of falling, e.g. through exercise or medication (Allen et al, 2022; Gillespie et al, 2012; Stevens and Lee, 2018), however, this requires people to be aware that they are having balance problems.

There are several tests that can be used to assess balance, like the BESTest (Horak et al, 2009), the Fullerton Advanced Balance Scale (Rose et al, 2006), or the Berg Balance Scale (BBS) (Berg, 1992), which are disease-agnostic, while there are also a variety of disease specific tests, which include balance related items, like the MDS-UPDRS (Goetz et al, 2008) for PD. All tests have in common that they are usually administered by a clinician limiting the frequency in which they can be administered and causing inconvenience to the patients because they have to travel to a clinic, which might also require the help of a caregiver.

To address these problems, several studies have investigated the possibility to automatically conduct balance assessments directly at the patients' homes and without the supervision of a clinician. For example, Wei and Dey (2019) employed a force plate to assess balance through a calculation using the center of mass and pressure. Other studies, e.g. Abujrida et al (2017); Arora et al (2015); Bot et al (2016); de Lima et al (2016); Lipsmeier et al (2018), avoided the need for special hardware, such as force plates, through the use of smartphone apps, which utilized the sensors embedded in smartphones, to monitor progression in PD through a variety of tests that assessed among other things also gait and balance. While some of the studies have shown that the collected data can be used to distinguish patients from healthy controls, the tests are not as comprehensive as the more holistic assessments used by clinicians, like the BESTest or the BBS, which consist of many everyday tasks and also evaluate whether someone needs external support, e.g. to stand up or to stand on one leg.

Recently, Morinan et al (2022) used computer vision to assess balance using the arising from a chair task of the MDS-UPDRS, which is also part of other assessments like the BBS. More specifically, a smartphone app was used to record the participant standing up, afterwards several body landmarks were obtained and used to extract balance related metrics. Although the task can potentially be done without the presence of a clinician, the described study was done in clinic and designed in a way that a clinician guides and conducts the assessment.

To the best of our knowledge, there have been no studies that have investigated conducting standard clinical assessments of balance remotely using a virtual agent to guide the interaction. In this paper, we evaluate the feasibility of using a multimodal dialog platform, which can be easily accessed through an internet browser without requiring the installation of any software, to conduct remote balance assessments. More specifically, we extended the Modality Platform (Ramanarayanan et al, 2023) with a balance assessment consisting of a number of tasks inspired by the BBS to answer the following research questions:

1. Is it feasible to administer standard balance assessments used by clinicians, like the BBS, remotely through a multimodal dialog system?
2. Can the recorded data be used to extract balance related metrics that are analytically valid?

To this end, 62 assessment sessions were collected from 58 healthy participants. 49 of the participants were recruited via Prolific[1], while 9 were internal testers. During the assessment, participants were guided by a virtual dialog agent, Tina, to perform a set of tasks inspired by the BBS, although they have been slightly modified for the purpose of the remote assessment.

The remainder of the paper is structured as follows: Sections 2 and 3 describe the employed system and integrated balance assessment. The collected data and performed analyses are presented and discussed in Sections 4 and 5. Finally, Section 6 concludes this paper.

2 System

The virtual agent, Tina, guiding participants through the interactive assessment is powered by a cloud-based multimodal dialog system (Ramanarayanan et al, 2023) designed to administer automated screening interviews to detect and monitor neurological and mental health conditions. During the interactive interview Tina guides participants through a series of structured conversational exercises designed to elicit speech, facial, and limb motor behaviors while analytics modules automatically extract a variety of audio (e.g. speaking rate), facial (e.g. range and speed of movement of the lips), and finger tapping metrics in near real-time from the recorded audio and video, and store them together with information about the interaction, like interview duration or completion status, in a database. After the interactive part of the assessment, participants are asked to complete one or more surveys about the interaction or a specific condition. All obtained information can be accessed by clinicians or researchers during and after the interaction through an easy-to-use web-based dashboard for further review and analysis.

[1] https://www.prolific.co/

3 Balance Assessment

The multimodal dialog system described in the previous section has been extended to be able to guide participants through a set of exercises inspired by the BBS. The BBS was chosen because, in contrast to assessments like the MDS-UPDRS[2] for Parkinson's Disease, it is disease agnostic. Further, it is less time consuming than other disease-agnostic tests like the BESTest or the Fullerton Advanced Balance Scale (Krzysztoń et al, 2018). Overall, six of the items in the BBS have been implemented with appropriate modifications to adapt them for a remote assessment. The items were selected based on their interpretability and ease of performance in a remote non-clinical setting. All items, except the standing on one leg task, are included twice to allow recording from both a frontal and side view, while the standing on one leg item was split into two separate tasks to ensure that participants are recorded standing both on their dominant and non-dominant legs. The following section provides an overview of the structure of the balance assessment and a description of the employed tasks.

1. **Part 1: Demographics Survey**
   Before the interactive assessment, participants are asked to complete a demographics survey and indicate whether they are able to stand up on their own and have a solid chair without wheels and armrests.[3] If they indicate that they cannot fulfil one of the requirements, they will not continue with the interactive balance assessment.
2. **Part 2: Interactive Assessment - Frontal View**
   Participants are shown a demonstration of the required seating setup, which ensures that their whole body (from head to toe) is visible both when sitting and standing[4], and a demonstration and explanation of the required tasks. Afterwards, they can either re-watch the demonstrations or change their seating setup as instructed. Finally, once they are ready, Tina will guide them through the following seven tasks.

   a. **Sit-to-Stand**: Participants are told to sit upright with their arms crossed across their shoulders and then asked to stand up.
   b. **Standing Unsupported**: Participants are told that they can put their hands down and then asked to stand upright for 10 seconds[5]

[2] https://www.movementdisorders.org/MDS/MDS-Rating-Scales/MDS-Unified-Parkinsons-Disease-Rating-Scale-MDS-UPDRS.htm

[3] Participants with known balance problems or neurological conditions will also be asked to confirm that a caregiver is present to assist them with the seating setup and for safety reasons, in case a participant loses their balance.

[4] To achieve this, participants need to place their recording device at a distance of approximately two meters from their seat.

[5] The duration of 10 seconds was chosen to ensure that the complete assessment does not take longer than 10 minutes to allow its integration into existing assessments of speech, facial, and limb motor behaviors. Whether 10 seconds are sufficient to extract clinically meaningful metrics will be investigated in future work.

c. **Standing Unsupported With Eyes Closed**: Participants are asked to close their eyes and keep standing until they hear further instructions. After 10 seconds they are told that they can open their eyes.
d. **Standing On One Leg Right**: Participants are asked to lift up their left leg (for 10 seconds) until they hear further instructions.
e. **Standing On One Leg Left**: Participants are asked to lift up their right leg until they hear further instructions. After 10 seconds they are told that they can put their foot down.
f. **Stand-to-Sit**: Participants are asked to cross their arms and sit down.
g. **Sitting Unsupported**: Participants are asked to sit upright, with their arms crossed, and without taking support of the backrest for 10 seconds.

3. **Part 3: Interactive Assessment - Side View**
Similar to the frontal view, participants are first shown demonstrations of the required seating setup and tasks. Afterwards, they can either re-watch the demonstrations or change their seating setup as instructed. Finally, once they are ready, Tina will guide them through the following five tasks, which use the same instructions as for the frontal view.

a. **Sit-to-Stand**
b. **Standing Unsupported**
c. **Standing Unsupported With Eyes Closed**
d. **Stand-to-Sit**
e. **Sitting Unsupported**

4. **Part 4: UX Survey**
After the interactive assessment, participants are asked to complete a user experience (UX) survey to obtain information about the usability of the system, e.g. the perceived performance of the system or the clarity of instructions. For participants with neurological conditions a disease-specific questionnaire will usually be shown, e.g. the MDS-UPDRS for PD patients.

The whole balance assessment takes about 10 minutes and can therefore be combined with other tasks designed to assess participants' speech or facial movements to provide a holistic assessment. The video captured during the assessment – recorded with a resolution of 320x240 pixels and 15fps to ensure that the system also works seamlessly for people with bandwidth-limited internet connections – is used to compute a variety of balance related metrics in two steps. First, MediaPipe Pose, which is based on BlazePose (Bazarevsky et al, 2020), is used to extract 33 3D body landmarks, and (2) a variety of metrics, like the time it takes participants to stand up or how long they were able to stand on one leg, are computed. Figure 1 illustrates the setup for the frontal tasks as well as the extracted body landmarks.

Fig. 1 Participant performing the Sit-to-Stand task. The right panel illustrates the extracted body landmarks.

4 Feasibility Analysis

To evaluate the feasibility of conducting a remote balance assessment via a multi-modal dialog system, the 51 assessment sessions from 49 healthy participants (24 females, 25 males; mean age 41 years, range 18 - 68 years) recruited via Prolific[6] were used, while the 11 assessments belonging to internal testers were discarded because they had a much higher compliance rate than the crowd-sourced participants. 7 assessments ended after the initial survey because the participants indicated that they did not have an appropriate chair and 3 assessments were terminated by the participants during the setup phase for the second part of the assessment because they did not have enough space. 2 of the remaining 41 sessions that started the interactive part of the assessment were terminated early. One of them was terminated during the third task without any obvious reason, while the other was terminated after the second part was completed and the participant was not able to adjust their seat for the third part, i.e. the side view. However, the latter participant did another successful assessment a few minutes later after changing their setup.

The overall task compliance, i.e. whether participants did the individual balance tasks as instructed, was very good at 86.8%. The main non-compliant behavior involved movements away from the camera after standing up or movements towards the recording device to adjust position so that the whole body is visible. Figure 2 shows that most non-compliant behavior occurred during the Sit-to-Stand and Stand-to-Sit tasks, while complete compliance was observed for the Standing Unsupported task. Furthermore, 77.8% of all non-compliant behavior was observed during the frontal view tasks as compared to the side view tasks. The reason might be that participants

[6] https://www.prolific.co/

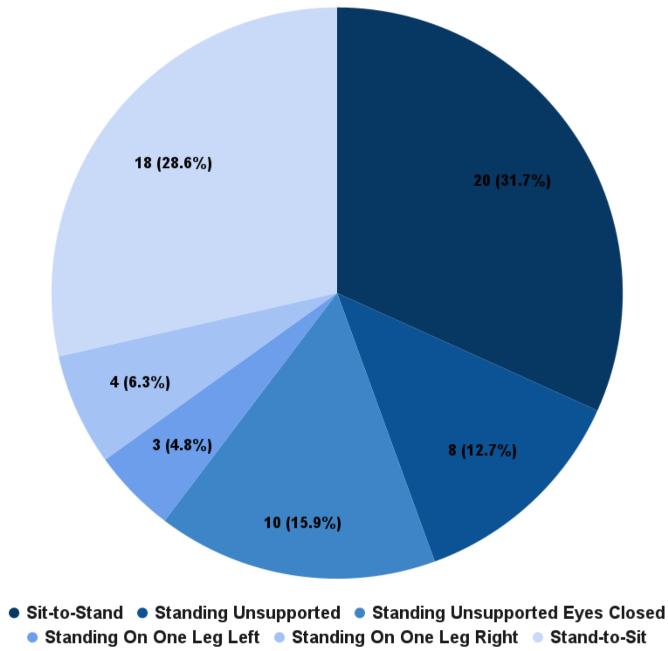

Fig. 2 Proportion of non-compliant behavior broken down by task. Overall, non-compliant behavior occurred for 63 (13.2%) of the 477 tasks recorded from all participants.

knew already what to expect when setting up their recording device and seat for the side view tasks. In contrast to the task-specific compliance, the compliance with the instructions regarding the camera setup, i.e. whether participants moved their seat so that their whole body was visible both when sitting and standing, was low. There was only one session where the participant was fully visible from head to toe for both parts of the interactive session. Interestingly, for the side view, seven participants were completely visible, which seems to confirm the learning/practice effect observed with task compliance. Furthermore, for seven assessments the participants were visible for both views from their heads to (at least) their knees, which can still allow clinicians to assess the performance. Note that while we consider data from paid crowdsourced participants here, participants in real-world deployment scenarios such as clinical trials are often more highly motivated (Sacristán et al, 2016), which would likely result in a higher compliance rate.

In general, our results show that it is feasible to conduct remote balance assessments via a multimodal dialog system because the majority of participants were able to properly perform the tasks as instructed. However, even properly performed tasks are difficult to evaluate if only a small part of the body is visible. Therefore, one must ensure adequate camera setup compliance before assessments can be conducted with patients. One option to ensure compliance might be to provide interactive feedback during the camera setup, which we will investigate in future work.

5 Analytical Validation

While the previous section has shown that utilizing a multimodal dialog system to conduct remote balance assessments is feasible, the question remains whether the obtained recordings can be used to extract meaningful automated metrics (over and above those that can be either hand annotated or assessed subjectively by a clinician). To this end, this section will analyze the analytical validity of an automatically computed time-taken-from-Sit-to-Stand (TSS) metric. The TSS metric is computed from shoulder landmarks extracted by MediaPipe Pose as follows:

1. The path of the center of the shoulders in 2D space (SC_{path}) was computed by taking the mean of the x- and y-coordinates of the left and right shoulder landmarks.
2. Min-max normalization was applied to reduce the effect of varying camera distances. Afterwards, the signal was passed through a Savitzky-Golay filter (Morgan et al, 2023) with a window length of five and a polynomial order of three to remove noise.
3. The velocity of the Sit-to-Stand movement was calculated by taking the first derivative of SC_{path}.
4. The start and end of the Sit-to-Stand activity was identified by determining the intervals during which the velocity changes for the first and last time more than a predefined threshold, respectively. More specifically, to determine the start of the Sit-to-Stand activity a sliding window with size L is used and moved to forward from the beginning of the video until the difference between the first and last frame of the sliding window is greater than threshold τ, in which case the first frame of the sliding window is considered the beginning of the Sit-to-Stand activity. To determine the end of the Sit-to-Stand activity, the window is moved backward starting from the end of the video.
5. Finally, the duration of the Sit-to-Stand activity is calculated as the number of frames in the interval determined during the previous step divided by the frame rate (in number of frames per second), i.e. $TSS = \frac{\#frames}{fps}$.

To determine the optimal values for the two parameters used by the TSS algorithm, i.e. L and τ, only 22 of the 62 assessments were used because they were the only ones for which participants were both visible from their heads to (at least) their knees and compliant with the instructions of the Sit-to-Stand task. Leave-one-out cross-validation was performed, to determine the parameters leading to the lowest mean absolute error (MAE) and root mean square error (RMSE) between the TSS metric value predicted by the algorithm and the ground truth obtained through manual annotations of the start and end frames of the Sit-to-Stand activity. For L, values between 2 and 15 frames were evaluated and for τ, values between 0.02 and 0.08 with a step size of 0.001. The lowest MAE of 224ms (standard deviation: 7ms) on the training data was achieved with $L = 10$ and $\tau = 0.044$ or $\tau = 0.045$, which in turn resulted in a MAE of 224ms (standard deviation: 156ms) on the test data, while the lowest RMSE of 224ms (standard deviation: 7ms) was obtained for $L = 10$ and $\tau = 0.049$ resulting in a RMSE of 267ms (standard deviation: 145ms) on the

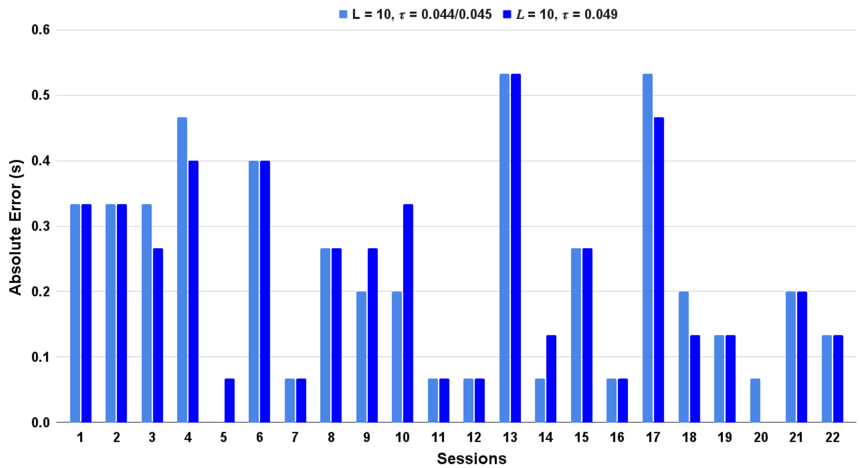

Fig. 3 Illustration of the absolute errors for all sessions and optimal parameter combinations.

test data. Figure 3 shows the absolute errors for all sessions and all three optimal parameter combinations. Since the obtained MAE and RMSE are close to the human inter-annotator MAE and RMSE (based on two annotators) of 202ms and 285ms, respectively, and less than 16% of the mean TSS of 1.73s, an argument can be made that the metric can be considered analytically valid for healthy controls. This illustrates that the recorded data can be used to extract analytically valid balance metrics if both the camera setup and task are performed correctly.

6 Conclusion

We have shown that it is in general feasible to conduct remote assessments of balance through a multimodal dialog system and that the collected data can be used to extract analytically valid balance metrics. However, the obtained results have also shown that the utility of the collected data depends strongly on participants' compliance to the instructions for the camera setup and tasks, and that especially the former is currently low. In future work, we will investigate whether interactive feedback to guide participants during the camera setup can increase compliance. Additionally, we will analytically validate further balance related metrics. Finally, we are planning to repeat both the feasibility assessment and analytical validation along with clinical validation for people with neurological conditions.

References

Abujrida H, Agu E, Pahlavan K (2017) Smartphone-based gait assessment to infer parkinson's disease severity using crowdsourced data. In: 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT), IEEE, pp 208–211

Allen NE, Canning CG, Almeida LRS, Bloem BR, Keus SH, Löfgren N, Nieuwboer A, Verheyden GS, Yamato TP, Sherrington C (2022) Interventions for preventing falls in parkinson's disease. Cochrane database of systematic reviews (6)

Arora S, Venkataraman V, Zhan A, Donohue S, Biglan KM, Dorsey ER, Little MA (2015) Detecting and monitoring the symptoms of parkinson's disease using smartphones: A pilot study. Parkinsonism & related disorders 21(6):650–653

Atteya A, Elwishy A, Kishk N, Ismail RS, Badawy R (2019) Assessment of postural balance in multiple sclerosis patients. The Egyptian Journal of Neurology, Psychiatry and Neurosurgery 55:1–5

Bazarevsky V, Grishchenko I, Raveendran K, Zhu T, Zhang F, Grundmann M (2020) Blazepose: On-device real-time body pose tracking. arXiv preprint arXiv:200610204

Berg K (1992) Measuring balance in the elderly: Development and validation of an instrument

Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, Doerr M, Pratap A, Wilbanks J, Dorsey E, et al (2016) The mpower study, parkinson disease mobile data collected using researchkit. Scientific data 3(1):1–9

Busse ME, Wiles CM, Rosser AE (2009) Mobility and falls in people with huntington's disease. Journal of Neurology, Neurosurgery & Psychiatry 80(1):88–90

Coote S, Comber L, Quinn G, Santoyo-Medina C, Kalron A, Gunn H (2020) Falls in people with multiple sclerosis: risk identification, intervention, and future directions. International journal of MS care 22(6):247–255

Gillespie LD, Robertson MC, Gillespie WJ, Sherrington C, Gates S, Clemson L, Lamb SE (2012) Interventions for preventing falls in older people living in the community. Cochrane database of systematic reviews (9)

Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stern MB, Dodel R, et al (2008) Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. Movement disorders: official journal of the Movement Disorder Society 23(15):2129–2170

Horak FB, Wrisley DM, Frank J (2009) The balance evaluation systems test (bestest) to differentiate balance deficits. Physical therapy 89(5):484–498

Krzysztoń K, Stolarski J, Kochanowski J (2018) Evaluation of balance disorders in parkinson's disease using simple diagnostic tests—not so simple to choose. Frontiers in neurology 9:932

de Lima ALS, Hahn T, de Vries NM, Cohen E, Bataille L, Little MA, Baldus H, Bloem BR, Faber MJ (2016) Large-scale wearable sensor deployment in parkinson's patients: the parkinson@ home study protocol. JMIR research protocols 5(3):e5990

Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, Cheng WY, Fernandez-Garcia I, Siebourg-Polster J, Jin L, et al (2018) Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 parkinson's disease clinical trial. Movement Disorders 33(8):1287–1297

Morgan C, Masullo A, Mirmehdi M, Isotalus HK, Jovan F, McConville R, Tonkin EL, Whone A, Craddock I (2023) Automated real-world video analysis of sit-to-stand transitions predicts parkinson's disease severity. Digital Biomarkers 7(1):92–103

Morinan G, Peng Y, Rupprechter S, Weil RS, Leyland LA, Foltynie T, Sibley K, Baig F, Morgante F, Wilt R, et al (2022) Computer-vision based method for quantifying rising from chair in parkinson's disease patients. Intelligence-Based Medicine 6:100046

Ramanarayanan V, Pautler D, Arbatti L, Hosamath A, Neumann M, Kothare H, Roesler O, Liscombe J, Cornish A, Habberstad D, Richter V, Fox D, Suendermann-Oeft D, Shoulson I (2023) When Words Speak Just as Loudly as Actions: Virtual Agent Based Remote Health Assessment

Integrating What Patients Say with What They Do. In: Proc. INTERSPEECH 2023, pp 678–679

Rose DJ, Lucchese N, Wiersma LD (2006) Development of a multidimensional balance scale for use with functionally independent older adults. Archives of physical medicine and rehabilitation 87(11):1478–1485

Sacristán JA, Aguarón A, Avendaño-Solá C, Garrido P, Carrión J, Gutiérrez A, Kroes R, Flores A (2016) Patient involvement in clinical research: why, when, and how. Patient preference and adherence pp 631–640

Stevens JA, Lee R (2018) The potential to reduce falls and avert costs by clinically managing fall risk. American journal of preventive medicine 55(3):290–297

Stolze H, Klebe S, Zechlin C, Baecker C, Friege L, Deuschl G (2004) Falls in frequent neurological diseases: prevalence, risk factors and aetiology. Journal of neurology 251:79–84

Wei W, Dey S (2019) Center of mass estimation for balance evaluation using convolutional neural networks. In: 2019 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, pp 1–7

Wood B, Bilclough J, Bowron A, Walker R (2002) Incidence and prediction of falls in parkinson's disease: a prospective multidisciplinary study. Journal of Neurology, Neurosurgery & Psychiatry 72(6):721–725