Domain-Adaptive Learning System for Predicting Berg Balance Scale

Hsin-Lung Wu, *Member, IEEE*, Bor-Shing Lin, *Senior Member, IEEE*, and Han-Yuan Chang

***Abstract*— Developing automatic balance evaluation systems is one of the important issues in healthcare. One of the common clinical scales for assessing balance is Berg balance scale (BBS). The state-of-the-art BBS evaluation systems are usually constructed by using deep learning techniques and inertial measurement units (IMU) worn on several body parts. Although existing BBS evaluation systems have proven to be useful in assessing balance, they still have the so-called cross-position balance assessment (CPBA) problem where only labelled IMU data on certain body part (called the source IMU domain) is available and the system needs to learn how to predict BBS scores based on unlabeled IMU training data on the other body parts (called the target IMU domain). To address the CPBA problem for evaluating BBS scores, this paper proposes two novel models called domain-adaptive regression for BBS scores based on the gait data (DARBS-G) and based on the movement data in which the participants performed the task to stand on one foot (DARBS-F) to assess balance under BBS scores. The proposed DARBS-G and DARBS-F models can transfer knowledge learned from the source IMU domain data to any target IMU domain by minimizing the domain classification loss and the BBS prediction loss at the same time. Experimental results show that the proposed DARBS framework outperforms the conventional BBS prediction model without domain adaption training. The result also demonstrates the propose DARBS approach can deal with the CPBA problem well.**

***Index Terms*— Berg balance scale, inertial measurement units (IMU), domain adaptation, deep adversarial training**

I. Introduction

WALKING imbalance problems may cause a lot of symptoms such as dizziness and, as a result, the patients may not feel well. This can greatly interfere with the patient's daily life and may lead to falls, which can cause broken bones and other injuries [1]. Nowadays, older adults may suffer from walking imbalance with high probability and hence regularly assessing the balance is important for them. There are many common balance assessment tools consisting of several well-known clinical scales such as the Berg Balance Scale (BBS) [2], timed up and go (TUG) [3], and single-leg stance (SLS) [4]. To give a careful evaluation of human body balance, several studies [5]–[7] suggested to adopt multiple scales to assess balance. However, because these scales consist of many tasks, their testing approaches are time-consuming. As an example, it requires patients to carry out 14 tasks and takes 15 to 20 min to fulfill BBS assessment. Moreover, patients often complete the evaluation by a physical therapist (PT). This greatly increases the workload of PTs according the Physical Therapy Workforce Study in the United States [8]. Therefore, many methods are proposed to construct efficient automated balance assessment systems. The basic idea consists of two phases. The first phase is to use devices such as Wii pressure plate, Microsoft Kinect, and wearable inertial measurement units (IMUs) to collect motion data. The second phase is to utilize machine learning (ML) methods to train and predict balance scale scores. Following this trend, Bacciu et al. [9] firstly proposed a framework based on Wii pressure plates to determine BBS scores and obtained a prediction system whose mean absolute error (MAE) was 3.8. In [10], Johnson et al. presented an approach based on Microsoft Kinect 2 to gather image data of six sitting tasks and predict the BBS scores. In addition, by their work, it only requires two sitting tasks to predict BBS scores and has MAE 1.16. Other than Wii pressure plate and Microsoft kinect 2, Badura and Pietka [11] used IMU to gather data by requiring participants to wear five IMUs and perform 14 BBS tasks. Then they used the collected data to train 14 models of the same structure to predict each task score. The average MAE of their method was 1.47. In [12], Shahzad et al. proposed a prediction method using only one IMU to collect the motion data when participants performed three tasks. The MAE of their method was 1.44. In [13], Lin et al. gave a thorough study in which they asked participants to wear IMUs on seven parts of the body and perform 17 test tasks and concluded that the participant is required to wear only one IMU on a specific part of the body and carry out two BBS tasks in order to predict the BBS score. The MAE of their proposed system was 1.274.

Besides using IMUs to gather motion data of BBS tasks, several works used IMUs to collect the participants' gait data to predict their BBS score. In [14], in order to predict the BBS score with ML methods, Similä et al. gathered the participants' gait data with an IMU when they walked at least 10 m. Nevertheless, the MAE of their method was 3.53. In [15], Lin et al. used 7 IMUs to collect gait data of the participants by asking them to walk forward in a straight line for 15 m. Then, authors proposed a deep learning (DL) prediction model combining a convolutional neural network (CNN) and a long

Manuscript received ???. This work was supported in part by the National Science and Technology Council, Taiwan, under Grant NSTC-113-2221-E-305-014; and in part by National Taipei University under Grant 2023-NTPU-IJRP-No.0002.

Hsin-Lung Wu, Bor-Shing Lin, and Han-Yuan Chang are with Department of Computer Science and Information Engineering, National Taipei University, New Taipei City 237303, Taiwan (e-mail: hsinlung@mail.ntpu.edu.tw; bslin@mail.ntpu.edu.tw;hanyuan@gmail.com).

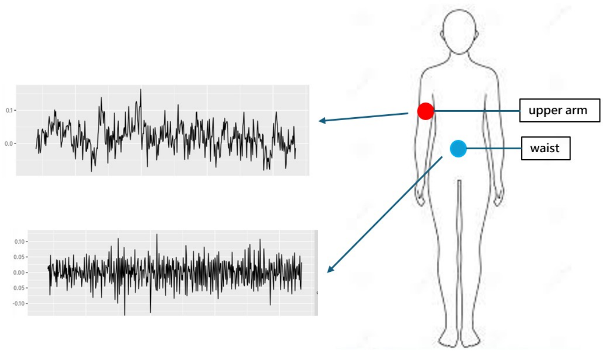

Fig. 1: The cross-position balance assessment problem (CPBA problem): We do not have labeled IMU motion data on the waist (the blue circle where we call it the target domain) while we only have labeled IMU data on the arm (the red circle where we call it the source domain). Our goal is to obtain the BBS prediction on the target domain based on the labeled motion data from the source domain.

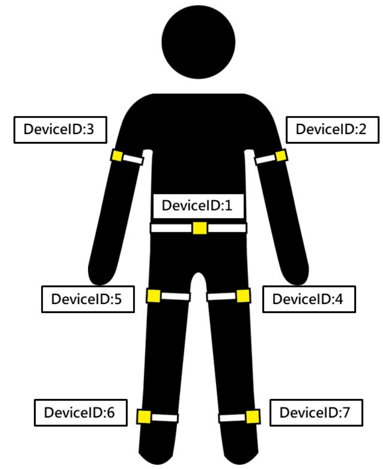

Fig. 2: ID numbers of the wearable IMUs located at different body parts.

short-term memory (LSTM) network model to predict BBS score by using the collected gait data from the IMU in the left thigh of the participant. The MAE of their CNN-LSTM model was 1.2562.

Although existing automatic balance evaluation systems have proven to be useful in assessing balance, they still have several major problems. One of key issues is the cross-position problem which is also an important issue of human activity recognition [18]. Figure 1 illustrates a situation of cross-position problem. Imagine that a patient is suffering from a severe brain disease which is highly related to activities. It is hard for us to equip his all body with IMUs to acquire the BBS scale scores since this will result in his unnatural activities. In fact, we can only label the activities on certain body parts of this patient. Suppose that a PT wants to check the activity information of the patient on the waist (the blue circle where we call it the target domain) which only contains IMU reading data without labels. How to use the IMU information with labeled BBS scale scores on other body part such as arm (the red circle where we call it the source domain) to help obtain the BBS prediction on the target domain? This problem is referred to as the cross-position balance assessment (CPBA).

The CPBA problem is quite challenging. Firstly, it is impossible to know which target position of a patient that a PT will check. Thus, we cannot determine a source body position which is the most similar to the target position in advance. Secondly, even when we know the body parts which are similar to the target domain, it remains difficult to establish a good machine learning model using both the source and the target domains. Since IMU signals from different domains follow different distributions, there is distribution variance between them. However, traditional machine learning models are usually constructed based on the assumption that all IMU signals follow the same distribution.

To address the CPBA problem, this paper utilizes the domain adaption (DA) technique to overcome the above challenges. The main domain adaptation approach is domain adversarial training (DAT) [19] in which training data and testing data come from similar but different distributions and predictions can be made based on features that cannot be discriminated between the training (source) and test (target) domains. However, existing DAT methods mainly focus on classification tasks while the BBS score prediction is a regression task. Therefore, with the proposed domain-adaptive regression (DAR) technique, the present study develops two novel methods called domain-adaptive regression for BBS scores based on the gait data (DARBS-G) and domain-adaptive regression for BBS scores based on the movement data in which the participants performed the task to stand on one foot (right foot) (DARBS-F) to assess balance under BBS scores where, for convenience, we called the latter dataset by F-dataset. Both the gait data and the F-dataset consist of labelled source sub-dataset and unlabeled target sub-dataset. DARBS-G and DARBS-F input the gait data and the F-dataset to automatically predict the BBS scores without restricting the participants to wear IMU on a particular body parts. The scores of BBS via the gait data and the F-data can be accurately predicted, which achieve the MAE of 2.3775 s and 1.4346 s, respectively. Although the MAE of predicting BBS score based on the gait data is higher than that based on the F-dataset, its MAE is still acceptable for application since the prediction method based on the gait data is a fast and convenient way for assessing balance in a clinical setting.

II. Datasets

In order to train and validate our proposed DARBS-G and DARBS-F domain-adaptive regression models to predict BBS scores for comprehensively assessing balance, this paper utilizes the F-dataset and the gait dataset collected in [15] while the objective is to tackle the CPBA problem.

In this article, the used gait dataset and F-dataset come from the previous study of Lin et al. [15] where 136 participants were recruited from National Taipei University in New Taipei City, Taipei Medical University, and Nangang District

TABLE I: Description of BBS tasks

Task ID	Description
Task 1	Standing unsupported
Task 2	Standing with feet together
Task 3	Standing with one foot in front
Task 4	Standing with eyes closed
Task 5	Reaching forward with outstretched arm
Task 6	Retrieving object from floor
Task 7	Turning to look behind
Task 8	Turning 360 degrees
Task 9	Standing to sitting
Task 10	Sitting unsupported
Task 11	Transfers
Task 12	Sitting to standing
Task 13	Standing on one foot (right foot)
Task 14	Placing alternate foot on stool

Jiuzhuang Community Center in Taipei City, Taiwan. The experimental procedure for data collection was approved by the Ethics Committee of Taipei Medical University Hospital (Code number: N201803012). Finally, 120 participants were used in their study for research analysis. For more details about the experimental participants, please refer to the paper [15]. To complete the experimental test, all the participants wear the IMU devices in their seven body parts shown in Figure 2 where these seven IMUs were worn in the following order: lumbar spine, right upper arm, left upper arm, right thigh, left thigh, right calf, and left calf. Each IMU collects 3-axis acceleration, 3-axis angular velocity and 3-axis magnetic orientation, totaling 9-axes of data. During the experimental process, each participant followed the instructions of the PT to carry out 14 BBS tasks shown in Table I for assessing balance, and PT evaluated the BBS scores of the participants as the ground-truth values. Moreover, each participant was also required to perform the walking test by wearing the seven IMUs in the same body parts. In the beginning of the walking test, the participants stand still in the original position. Once the walking test started, the participant was asked to walk forward in a straight line for 15 m at their natural walking speed. After reaching the destination, the participant stood still and the walking test ended. In order to increase the amount of gait data, participants were asked to repeat this walking test six times with a 30 s resting period between tests.

Each participant followed the instructions of the PT to carry out the tasks during the experimental process. 16 participants were excluded because they could not finish the task, or some of their data was missing. Finally, the collected data from 120 participants were included in the gait dataset.

As demonstrated in [13], Task 13 in Table I is one of important BBS tasks for predicting BBS scores with an MAE of 1.469 was obtained. Based on their research result, we collect the IMU movement data of BBS Task 13 of these participants as the F-dataset used in this article.

III. METHODS

A. System Overview

In this section, we propose two deep learning methods called DARBS-G and DARBS-F with the same architecture to predict BBS scores based on the gait IMU data and the movement IMU data in which the participants performed the task to stand on one foot (right foot), respectively. For DARBS-G, the participants need to walk naturally for 15 m to obtain the BBS scores while they are required to stand on one foot(right foot) for DARBS-F. The overall system architecture of DARBS-F and DARBS-G is shown in Fig. 3. The kinematic data of the participants is obtained when they walk in a straight line for 15 m or stand on one foot (right foot). One of the seven IMUs is selected as the source IMU while the rest are called target IMUs. Next, the two collected data are segmented into second-long segments to generate the gait dataset and F-dataset, respectively. Moreover, A PT provides the BBS score for each participant as the labelled value of the movement data gathered from the source IMU of the participant. Now, the dataset is divided into two parts: the source dataset and the target dataset. These two sub-datasets are input into the domain-adaptive deep learning model for domain-adaptively training to obtain the prediction model which can estimate BBS score from any IMU data.

B. Data Preprocessing

As mentioned in [15], the collected IMU data have much variance due to the differences in the participants' body shape, the shaking amplitude, and so on. Thus, following the approach in [15], we utilize Z-score standardization to mitigate the influence of the body shape difference of the participants on the IMU movement data, the formula is as follows.

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

where X is the original IMU time series data on the axis, μ is the mean of the IMU data on the axis, σ is the standard deviation of the IMU data on the axis, and X' is the standardized IMU data on the axis.

The score range of the BBS scale is between 0 and 56. We simply use min-max normalization to scale it to the range of 0 to 1. The formula for min-max normalization is as follows.

$$b' = \frac{b - b_{\min}}{b_{\max} - b_{\min}} \tag{2}$$

where b denotes the original BBS score, $b_{\max}$ and $b_{\min}$ are the maximum and minimum values of the BBS scale, and b' is the result of min-max normalization.

1) Data Segmentation: For the gait dataset and the F-dataset, according to [15], the collected IMU data is suggested to be segmented with a fixed window size. Based on the experimental result in [15], the authors suggested to set the window size by 3 seconds and this study follows their suggested window size setting. Since the sampling rate of IMU devices is 50 Hz, the generated gait dataset and F-dataset have training and testing examples with the size 150×9.

2) Source dataset and target dataset: In order to carry out the domain adversarial training, the current dataset (the gait dataset or the F-dataset) is divided into two sub-datasets: the source dataset and the target dataset. The source and the target datasets consist of the movement data generated by the source IMU and the rest of IMUs, respectively. In addition, the size of the current source dataset is much smaller than the target dataset. In order to have a good domain adversarial

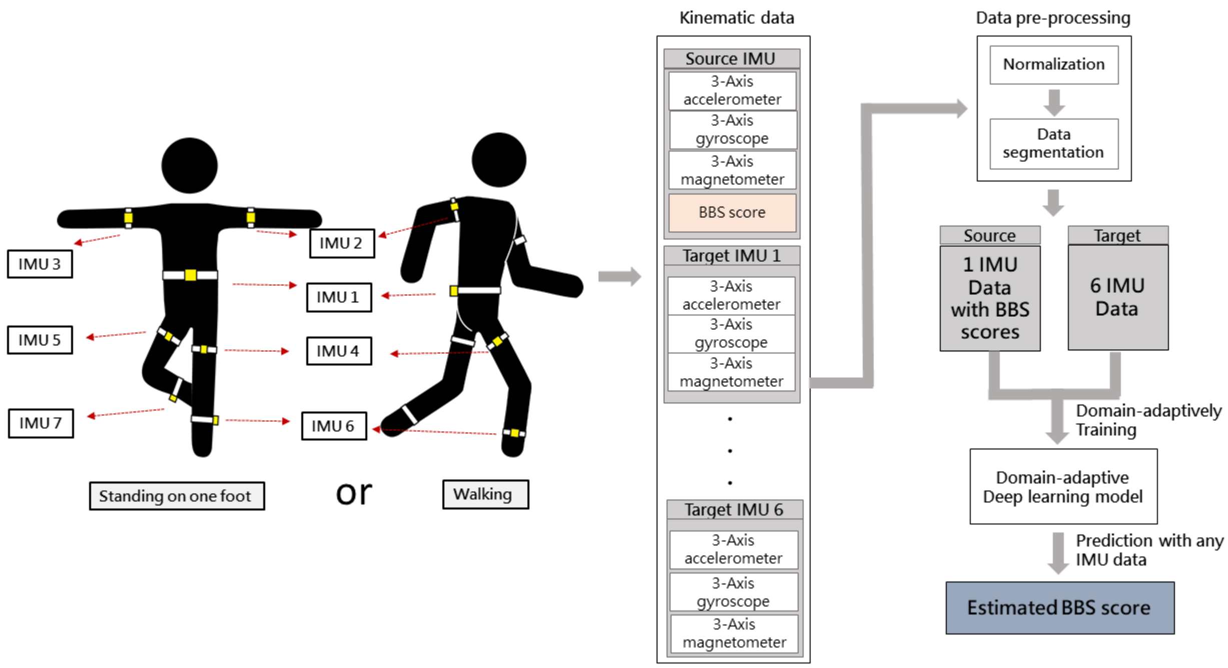

Fig. 3: The proposed overall system architecture of DARBS-F and DARBS-G.

training, we augment the original source dataset by copying all examples in it six times.

C. Model architecture and its learning procedure

This study applies the domain-adversarial neural network (DANN) framework [19] to construct the proposed system DARBS which consists of a feature extractor G_f, a balance predictor G_b, and a domain predictor G_d. The whole procedure can be divided into two stages: the training stage and the testing stage which are illustrated in Fig. 4 and Fig. 5, respectively.

In the training stage as shown in Fig. 4, on one hand, the balance predictor learns to predict BBS scores from the labelled source dataset. On the other hand, the domain predictor includes a gradient reversal layer (GRL) which connects to the feature extractor, makes the input unaltered in the forward pass, and multiplies the gradient by a certain negative constant during backpropagation. By combining these two predictors, it is possible to extract domain-invariant and balance-predictive features simultaneously.

Precisely, let $S = \{(x_i^s, b_i) : i = 1, \dots n_s\}$ denote the labelled source dataset and $T = \{(x_i^t) : i = 1, \dots n_t\}$ denote the unlabeled target dataset where $v_i \in [0,1]$ is the normalized BBS score. We consider the following objective function:

$$
\begin{aligned}
E(\theta_f, \theta_b, \theta_d) = &\sum_{\substack{i=1\\ d_i=0}}^{n_s} L_b(G_b(G_f(x_i;\theta_f);\theta_b), b_i) - \\
&\lambda \sum_{\substack{i=1\\ d_i\in\{0,1\}}}^{n_t} L_d(G_d(G_f(x_i;\theta_f);\theta_d), d_i) \\
&\triangleq \sum_{\substack{i=1\\ d_i=0}}^{n_s} L_b(\theta_f, \theta_b) - \lambda \sum_{\substack{i=1\\ d_i\in\{0,1\}}}^{n_t} L_d(\theta_f, \theta_d)
\end{aligned} \tag{3}
$$

where $L_b(\cdot,\cdot)$ is the loss for BBS score prediction and $L_d(\cdot,\cdot)$ is the loss for the domain classification, while L_b^i and L_d^i denote the loss functions evaluated at the i-th training example.

As depicted in [19], the goal is to find the parameters $\hat{\theta}_f, \hat{\theta}_b, \hat{\theta}_d$ such that

$$(\hat{\theta}_f, \hat{\theta}_b) = \arg\min_{\theta_f,\theta_b} E(\theta_f, \theta_b, \hat{\theta}_d) \tag{4}$$

$$\hat{\theta}_d = \arg\max_{\theta_d} E(\hat{\theta}_f, \hat{\theta}_b, \theta_d). \tag{5}$$

The traditional method to solve Equation 4 and Equation 5 is to use the following gradient descent steps with a given learning rate μ.

$$\theta_f \longleftarrow \theta_f - \mu\left(\frac{\partial L_b^i}{\partial \theta_f} - \lambda \frac{\partial L_d^i}{\partial \theta_f}\right) \tag{6}$$

$$\theta_b \longleftarrow \theta_b - \mu \frac{\partial L_b^i}{\partial \theta_b} \tag{7}$$

$$\theta_d \longleftarrow \theta_d - \mu \frac{\partial L_d^i}{\partial \theta_d}. \tag{8}$$

In order to realize the above optimization task efficiently, a gradient reversal layer GRL (the purple block in Fig. 4) is

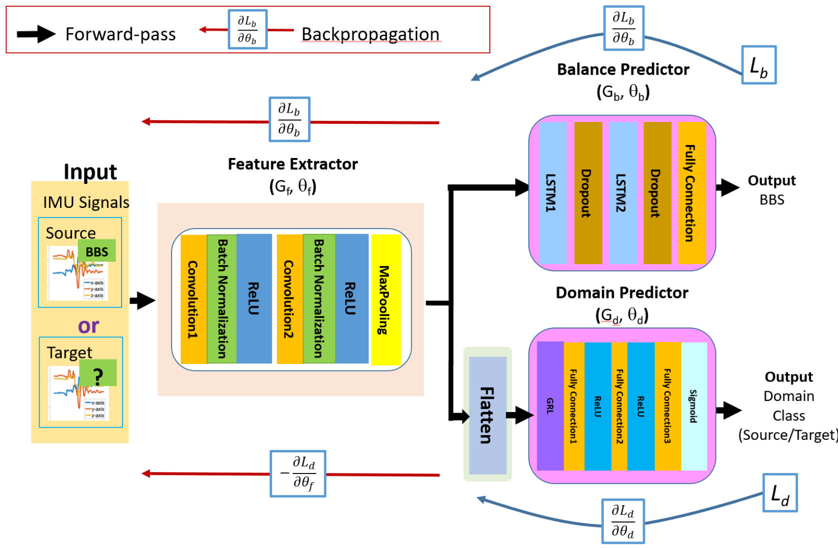

Fig. 4: The proposed domain adversarial training framework for the DARBS-G and DARBS-F.

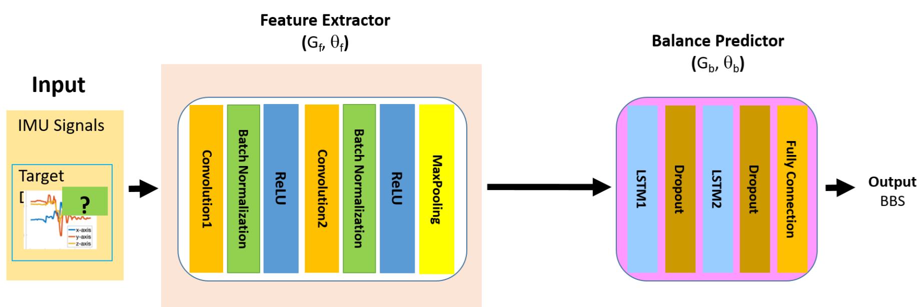

Fig. 5: The testing procedure for the DARBS-G and DARBS-F.

suggested to connect G_f and G_d [19] where I is the identity matrix and GRL is the function satisfying

$$\mathrm{GRL}(x) = x \tag{9}$$

$$\frac{d\mathrm{GRL}}{dx} = -I. \tag{10}$$

By using the function GRL, a new objective function $\hat{E}$ can be defined as follows.

$$\hat{E}(\theta_f, \theta_b, \theta_d) = \sum_{\substack{i=1 \\ d_i=0}}^{n_s} L_b(G_b(G_f(x_i;\theta_f);\theta_b), b_i) + \lambda \sum_{\substack{i=1 \\ d_i\in\{0,1\}}}^{n_t} L_d(G_d(\mathrm{GRL}(G_f(x_i;\theta_f));\theta_d), d_i). \tag{11}$$

Finally, instead of carrying out the optimization task (updates by running Equation 6, Equation 7, and Equation 8), we just do the stochastic gradient descent (SGD) for Equation 11. This completes the training phase.

Let $\tilde{\theta}_f$ and $\tilde{\theta}_b$ be the learned parameters during the learning stage. In the testing stage, the BBS score predictor defined by

$$b(x) \triangleq G_b(G_f(x;\tilde{\theta}_f);\tilde{\theta}_b) \tag{12}$$

can be used to predict BBS scores for samples from not only the source domain but also the target domain where the structure of the predictor is illustrated in Fig. 5.

In the last part of this section, we give detailed description for the feature extractor G_f, the balance predictor G_b, and the domain predictor G_d. The feature extractor G_f is a 1-D CNN model which can extract spatial features. G_f contains 2 convolutional layers and each convolutional layer is followed by a batch normalization layer and an activation layer. The balance predictor is an LSTM model which is suitable for extracting temporal features. This LSTM model contains two LSTM layers with 64 memory cells and each LSTM layer is followed by a dropout layer to prevent from overfitting. Its output layer is a fully connected network to predict BBS scores. The domain predictor is three-layer fully connected networks. Its last layer is an activation layer with the sigmoid function in order to output predicted value between 0 and 1.

Finally, the loss function L_b is the mean square error (MSE) defined by

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{b}_i - b_i)^2 \tag{13}$$

where n is the batch size, b_i is the actual normalized score of the BBS scale, $\hat{b}_i$ is the predicted score of the BBS predictor $G_b(G_f(x_i;\tilde{\theta}_f);\tilde{\theta}_b)$ for the test labelled sample (x_i, b_i). The loss function L_d is the binary cross-entropy error (BCE) defined by

$$\mathrm{BCE} = \frac{-1}{n_t}\sum_{i=1}^{n_t}\Big(d_i \log(\hat{d}_i) + (1-d_i)\log(1-\hat{d}_i)\Big) \quad (14)$$

where n_t is the number of training examples, d_i is the actual label of the domain class for the input x_i, $\hat{d}_i$ is the output of the domain predictor.

D. Model performance evaluation

To evaluate the prediction performance of the proposed model, this study uses five-fold cross-validation and randomly divide 120 participants into five equal parts. IMU data collected from 96 participants are used for training, while the remaining IMU data are used for testing. The process is repeated five iterations. During each iteration, the mean absolute error (MAE) between the predicted BBS scores and the actual BBS scores is calculated where MAE is defined by

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{B}_i - B_i| \quad (15)$$

where n is the number of test samples, B_i is the actual score of the BBS scale, $\hat{B}_i$ is the predicted score of the BBS predictor $G_b(G_f(x_i;\tilde{\theta}_f);\tilde{\theta}_b)$ for the test labelled sample (x_i, B_i).

IV. Experimental Results

The proposed models DARBS-F and DARBS-G are developed using Python 3.7.13 and Pytorch 2.0. The used optimizer is Adam and the learning rate is set to 0.0005. The training loss functions are MSE for the balance predictor and BCE for the domain predictor, respectively. The batch size is set to 64 and the number of epochs is set to 500. As mentioned in Section III-B.1, the window size is set to 3 seconds.

For the gait dataset and the F-dataset, we conduct three experiments which are the experiment of our domain-adaptive regression for BBS (DARBS), the source-to-source experiment (S2S), and the source-to-target experiment (S2T) for each dataset. First of all, we select the IMU data from a specific IMU location as the source domain data. In the experiment of DARBS, the remaining IMU data from other body parts serve as the target domain data as mentioned in Section III. In the S2S and S2T experiments, we directly train the composition of the BBS predictor and the feature extractor, that is $G_b(G_f(\cdot;\theta_f);\theta_b)$, on the source domain data. Then we test this trained composite generator on the source domain data and on each target domain data in the S2S and S2T experiments, respectively. The experimental results for the F-dataset and the gait data are shown in Table II and Table III, respectively.

Since the source IMU data and the target IMU data comes from different distributions, it is expected that the MAE of the composite generator trained on the source IMU data should have high MAE when testing on the target IMU data. On the other hand, its MAE should be low when testing on the source IMU data. The results of the S2T and S2S experiments demonstrate this speculation. For the composite generator trained on given source IMU data, Table II and Table III show that the average MAE tested on target IMU data is much higher than the MAE tested on the source IMU data.

The first step of the proposed DARBS is to transform the source and target IMU data space into a common feature space and then carry out BBS prediction. After transformation by the feature extractor G_f, the transformed source and target IMU data have similar distributions. Thus it is expected that, after trained on the source IMU data, the proposed DARBS outperforms the composite generator in the S2T experimental setting when tested on target IMU data. The results of the DARBS experiment demonstrate this speculation. In fact, Table II and Table III show that, when testing on target IMU data, the average MAE of our DARBS is lower than that of the composite generator trained on the source IMU data.

Moreover, overall, among all source IMUs, the DARBS-F model can achieve the lowest average MAE, which is 1.3744. Therefore, it is suggested to take IMU 4 (right thigh) as the source IMU for training and to evaluate BBS for the BBS task ”stand on one foot (right foot)”.

However, as mentioned in [15], it is more practical to ask the participant to carry out the walking test than to stand on one foot. Thus, let us focus on the prediction performance of DARBS-G. Among all source IMUs, the DARBS-G model can achieve the lowest average MAE, which is 2.3329. Therefore, we suggest taking IMU 7 (left calf) as the source IMU for training and to evaluate BBS for the walking test ”walk forward in a straight line for 15 m”.

In these experiments, we remark that the MAE results of S2S and S2T can be regarded as the upper and lower bounds, respectively, of prediction performance of the proposed DARBS framework. For the F-dataset and the gait dataset, the MAE gaps between DARBS and ”S2S” are 0.3092 and 0.919, respectively. For the BBS task ”stand on one foot (right foot)”BBS task”, the prediction performance of DARBS-F is close to the optimal performance in the ”S2S” sense. However, for the walking test, the MAE gap is slightly large. Therefore, it can be left as our future work to improve the performance of DARBS-G.

V. Conclusion

In this paper, we propose DARBS-F and DARBS-G based on the domain adversarial training method to deal with the cross-position balance assessment problem. The proposed methods are able to transfer knowledge learned from the source IMU domain data to any target IMU domain by minimizing the domain classification loss and the BBS prediction loss simultaneously. Experimental results not only show that the overall BBS prediction performance using the proposed DARBS approach outperforms that using the simple composite of the feature generator and the balance predictor when testing on the target IMU data but also demonstrate the feasibility and effectiveness of the proposed DARBS framework for dealing with the cross-position balance assessment problem.

References

[1] A. Viswanathan and L. Sudarsky, ”Balance and gait problems in the elderly” in *Handbook of Clinical Neurology,* vol. 103, pp.623–634, 2012.

TABLE II: The MAE of DARBS-F and the S2T-experimental results for the F-dataset

Test / Source	S2T								
	IMU 1	IMU 2	IMU 3	IMU 4	IMU 5	IMU 6	IMU 7	AVG	S2S
IMU 1		1.6057	1.7198	1.595	1.8052	1.7447	1.5684	1.6731	1.0821
IMU 2	1.7562		1.7868	1.6721	1.6853	1.6975	1.5361	1.689	1.1078
IMU 3	1.686	1.7399		1.6279	1.645	1.6596	1.5176	1.646	1.1491
IMU 4	1.9704	1.9669	2.0421		2.0849	1.712	1.7984	1.9291	1.0652
IMU 5	1.7064	1.7065	1.6947	1.6219		1.5972	1.5562	1.6471	1.0415
IMU 6	1.8765	1.8517	1.8573	1.4337	1.883		1.7357	1.7729	1.1343
IMU 7	1.6589	1.6469	1.6834	1.6383	1.7199	1.6196		1.6611	1.1883

Test / Source	DARBS-F								
	IMU 1	IMU 2	IMU 3	IMU 4	IMU 5	IMU 6	IMU 7	AVG	S2S
IMU 1		1.40051	1.4991	1.3943	1.407	1.3481	1.4651	1.419	1.0821
IMU 2	1.5058		1.5027	1.4153	1.4552	1.399	1.3769	1.4424	1.1078
IMU 3	1.4912	1.5076		1.4742	1.4787	1.5155	1.4143	1.4802	1.1491
IMU 4	1.4534	1.3525	1.4154		1.3948	1.2262	1.4042	1.3744	1.0652
IMU 5	1.4957	1.4574	1.4711	1.2979		1.3909	1.3703	1.4138	1.0415
IMU 6	1.5184	1.34133	1.5091	1.3222	1.3404		1.4439	1.4125	1.1343
IMU 7	1.5489	1.4935	1.5217	1.5298	1.5794	1.4812		1.5257	1.1883

TABLE III: The MAE of DARBS-G and the S2T-experimental results for the gait dataset

Test / Source	S2T								
	IMU 1	IMU 2	IMU 3	IMU 4	IMU 5	IMU 6	IMU 7	AVG	S2S
IMU 1		2.5718	2.554	2.6865	2.5906	2.6768	2.5662	2.6076	1.5203
IMU 2	2.6796		2.5601	2.4281	2.5736	2.6172	2.4236	2.547	1.5807
IMU 3	2.4861	2.4429		2.5738	2.5001	2.475	2.4347	2.4854	1.4369
IMU 4	2.5718	2.6513	2.8058		2.7411	2.6133	2.8134	2.6994	1.6073
IMU 5	2.4651	2.4565	2.472	2.629		2.4547	2.5751	2.5087	1.3605
IMU 6	2.4829	2.4802	2.5015	2.4589	2.5174		2.4911	2.4886	1.5697
IMU 7	2.4294	2.4325	2.4124	2.4056	2.7623	2.4005		2.4737	1.4139

Test / Source	DARBS-G								
	IMU 1	IMU 2	IMU 3	IMU 4	IMU 5	IMU 6	IMU 7	AVG	S2S
IMU 1		2.4136	2.4562	2.3695	2.3149	2.3687	2.3539	2.3794	1.5203
IMU 2	2.353		2.3564	2.3525	2.386	2.3481	2.3544	2.3584	1.5807
IMU 3	2.3406	2.3461		2.3445	2.3349	2.3552	2.3462	2.3445	1.4369
IMU 4	2.4886	2.3356	2.4691		2.419	2.4264	2.4969	2.4392	1.6073
IMU 5	2.3453	2.3456	2.3415	2.3386		2.3451	2.3392	2.3425	1.3605
IMU 6	2.3518	2.3513	2.3668	2.3592	2.349		2.3506	2.3547	1.5697
IMU 7	2.3395	2.3374	2.3286	2.3267	2.3333	2.3321		2.3329	1.4139

[2] K. Berg, S. Wood-Dauphine, J. I. Williams and D. Gayton, ”Measuring balance in the elderly: Preliminary development of an instrument”, *Physiotherapy Canada,* vol. 41, no. 6, pp.304–311, Nov. 1989.

[3] D. Podsiadlo and S. Richardson, ”The timed ‘up & go’: A test of basic functional mobility for frail elderly persons,” *J. Amer. Geriatrics Soc.,* vol. 39, no. 2, pp.142–148, Feb. 1991.

[4] E. Jonsson, Å. Seiger and H. Hirschfeld, ”One-leg stance in healthy young and elderly adults: A measure of postural steadiness?,” *Clin. Biomech.,* vol. 19, no. 7, pp.688–694, Aug. 2004.

[5] T. M. Steffen, T. A. Hacker and L. Mollinger, ”Age- and gender-related test performance in community-dwelling elderly people: Six-minute walk test Berg balance scale timed up & go test and gait speeds,” *Phys. Therapy,* vol. 82, no. 2, pp.128–137, Feb. 2002.

[6] A. Yingyongyudha, V. Saengsirisuwan, W. Panichaporn and R. Boonsinsukh, ”The mini-balance evaluation systems test (mini-BESTest) demonstrates higher accuracy in identifying older adult participants with history of falls than do the BESTest Berg balance scale or timed up and go test,” *J. Geriatric Phys. Therapy,* vol. 39, no. 2, pp.64–70, Apr. 2016.

[7] G. Fahlström, K. Kamwendo, J. Forsberg and L. Bodin, ”Fall prevention by nursing assistants among community-living elderly people. A randomised controlled trial,” *Scandin. J. Caring Sci.,* vol. 32, no. 2, pp.575–585, Jun. 2018.

[8] J. L. Zimbelman, S. P. Juraschek, X. Zhang and V. W. H. Lin, ”Physical therapy workforce in the United States: Forecasting nationwide shortages,” *PM&R,* vol. 2, no. 11, pp.1021–1029, Nov. 2010.

[9] D. Bacciu et al., ”A learning system for automatic Berg balance scale score estimation,” *Eng. Appl. Artif. Intell.,* vol. 66, pp.60–74, Nov. 2017.

[10] L. Johnson et al., ”An automated electronic assessment tool can accurately classify older adult postural stability,” *J. Biomech.,* vol. 93, no. 27, pp.6–10, Aug. 2019.

[11] P. Badura and E. Pietka, ”Automatic Berg balance scale assessment system based on accelerometric signals,” *Biomed. Signal Process. Control,* vol. 24, pp.114–119, Feb. 2016.

[12] A. Shahzad, S. Ko, S. Lee, J.-A. Lee and K. Kim, ”Quantitative assessment of balance impairment for fall-risk estimation using wearable triaxial accelerometer,” *IEEE Sensors J.,* vol. 17, no. 20, pp.6743–6751, Oct. 2017.

[13] B.-S. Lin, Z. Zhang, C.-W. Peng, C.-C. Lin, C.-H. Lin and B.-S. Lin, ”Automatic Assessment System Based on IMUs and Machine Learning for Predicting Berg Balance Scale,” *IEEE Sensors Journal,* vol. 22, no. 20, pp.19919–19930, 15 Oct.15, 2022.

[14] H. Similä, J. Mäntyjärvi, J. Merilahti, M. Lindholm and M. Ermes, ”Accelerometry-based Berg balance scale score estimation,” *IEEE J. Biomed. Health Informat.,* vol. 18, no. 4, pp.1114–1121, Jul. 2014.

[15] C.-H. Lin, C.-W. Peng, I.-J. Lee, C.-M. Wu and B.-S. Lin, ”Analysis of Gait Data Based on Deep Learning to Predict Multiple Balance Scale Scores,” *IEEE Sensors Journal,* vol. 24, no. 1, pp.920–930, 1 Jan.1, 2024.

[16] J. Choi, S. M. Parker, B. A. Knarr, Y. Gwon and J.-H. Youn, ”Wearable sensor-based prediction model of timed up and go test in older adults,” *Sensors,* vol. 21, no. 2, pp.6831–6847, Oct. 2021.

[17] E. Barry, R. Galvin, C. Keogh, F. Horgan and T. Fahey, ”Is the timed up and go test a useful predictor of risk of falls in community dwelling older adults: A systematic review and meta- analysis,” *BMC Geriatrics,* vol. 14, no. 1, pp.1–14, Feb. 2014.
[18] X. Gu et al., ”Beyond Supervised Learning for Pervasive Healthcare,” *IEEE Reviews in Biomedical Engineering,* vol. 17, pp.42–62, 2024.
[19] Y. Ganin et al., “Domain-adversarial training of neural networks,” *J. Mach. Learn. Res.,* vol. 17, no. 1, pp.2030-–2096, May 2015.

Hsin-Lung Wu received his Ph.D. degrees in Computer Science and Information Engineering from National Chiao Tung University, Taiwan in 2008. He is currently a Professor in the Department of Computer Science and Information Engineering at National Taipei University, Taiwan. His main research interests include design and analysis of algorithms, computational complexity, theory of machine learning, and deep learning.

Bor-Shing Lin (M’01–SM’20) received the B.S. degree in electrical engineering from National Cheng Kung University, Taiwan, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taiwan, in 1999 and 2006, respectively. He is currently a Distinguished Professor with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan. His research interests include smart medicine, embedded systems, wearable systems, biomedical signal processing, biomedical image processing, and portable biomedical electronic system design. He is a fellow of the Institution of Engineering and Technology (IET), U.K., and the British Computer Society (BCS), U.K.

Han-Yuan Chang received his master degree in Computer Science and Information Engineering from National Taipei University, Taiwan in 2024. His research interests include machine learning and deep learning.