Automatic Detection of Compensation During Robotic Stroke Rehabilitation Therapy

YING XUAN ZHI1, MICHELLE LUkasik1, MICHAEL H. LI1, ELHAM DOLATABADI1, ROSALIE H. WANG1,2, AND BABAK TAATI1,3,4

1Toronto Rehabilitation Institute—University Health Network, Toronto, ON M5G 2A2, Canada
2Department of Computer Science, University of Toronto, Toronto, ON M5G 1V7, Canada
3Department of Computer Science, University of Toronto, Toronto, ON M5S 3G4, Canada
4Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada

CORRESPONDING AUTHOR: B. TAATI (babak.taati@uhn.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and in part by the Toronto Rehabilitation Institute—University Health Network.

ABSTRACT

Robotic stroke rehabilitation therapy can greatly increase the efficiency of therapy delivery. However, when left unsupervised, users often compensate for limitations in affected muscles and joints by recruiting unaffected muscles and joints, leading to undesirable rehabilitation outcomes. This paper aims to develop a computer vision system that augments robotic stroke rehabilitation therapy by automatically detecting such compensatory motions. Nine stroke survivors and ten healthy adults participated in this study. All participants completed scripted motions using a table-top rehabilitation robot. The healthy participants also simulated three types of compensatory motions. The 3-D trajectories of upper body joint positions tracked over time were used for multiclass classification of postures. A support vector machine (SVM) classifier detected lean-forward compensation from healthy participants with excellent accuracy (AUC = 0.98, F1 = 0.82), followed by trunk-rotation compensation (AUC = 0.77, F1 = 0.57). Shoulder-elevation compensation was not well detected (AUC = 0.66, F1 = 0.07). A recurrent neural network (RNN) classifier, which encodes the temporal dependency of video frames, obtained similar results. In contrast, F1-scores in stroke survivors were low for all three compensations while using RNN: lean-forward compensation (AUC = 0.77, F1 = 0.17), trunk-rotation compensation (AUC = 0.81, F1 = 0.27), and shoulder-elevation compensation (AUC = 0.27, F1 = 0.07). The result was similar while using SVM. To improve detection accuracy for stroke survivors, future work should focus on predefining the range of motion, direct camera placement, delivering exercise intensity tantamount to that of real stroke therapies, adjusting seat height, and recording full therapy sessions.

INDEX TERMS
Motion compensation, posture classification, rehabilitation robotics, stroke rehabilitation.

I. INTRODUCTION

Stroke is the result of a shortage of blood supply to the brain leading to cell death. In America, stroke has become the primary cause of disability [1]. Functional impairment in the upper extremities caused by stroke often leads to loss of independence in the performance of activities of daily living. Post-stroke rehabilitative interventions for improving upper limb function are effective in helping stroke survivors in early stages regain motor function and range of motion [2]. These interventions, which include constraint-induced movement therapy and repetitive task practice [3], are often performed under the supervision of a healthcare professional in an in-patient setting with prolonged sessions [4]. This can be costly to the healthcare system and may hinder patient access to rehabilitation, especially in remote regions.

A. ROBOTIC STROKE REHABILITATION THERAPY

Robotic stroke rehabilitation therapy has the potential to improve upper limb motor recovery for stroke survivors [5], [6]. Although robots are not designed to replace therapists, they are a potentially cost-effective complement to therapy assistance and can increase therapy time. For instance, with robotic rehabilitation devices, a single therapist could supervise the exercise sessions of multiple users simultaneously, rather than one patient at a time. This can greatly increase therapy delivery efficiency. In the past, our team at...
the Toronto Rehabilitation Institute (TRI) developed a haptic robot prototype with an end-effector that provides shoulder and elbow movement therapy. Although the potential clinical efficacy is promising [7], a major consideration when using this robotic rehabilitation system is whether users can correctly execute therapy exercises under minimal therapist supervision [8]. During therapy, common movement deficits are often present, including muscle weakness, abnormal postural adjustments, and loss of inter-joint coordination. When stroke survivors attempt to move and encounter these deficits, they naturally compensate with the available motor strategies by recruiting unaffected muscles and joints to overcome limitations in affected muscles and joints [9]. In the face of fatigue or unclear instructions, stroke survivors also tend to resort to compensatory motions during repetitive rehabilitation exercises [10]. Such compensations hinder recovery progress and can introduce additional orthopedic problems [11].

B. COMPENSATION DETECTION

To take full advantage of robotic stroke rehabilitation therapy, measures should be taken to correct compensations. Posture correcting exoskeletons provide a solution by controlling joint movement [12]. However, the TRI robotic prototype contains an end-effector, allowing for many degrees of freedom at various joints. Controlling joint movement will inevitably hamper such movement freedom. Therefore, exoskeletons are not applicable in this scenario. Other robotic systems utilize physical constraint such as straps to keep users in the chair, which is non-ideal due to safety concerns and discomfort. Furthermore, visual marker-based and wearable-based kinematic assessments can detect compensation, and feedback from the system interface can be issued to prompt users to correct their postures. Yet, expensive or complex equipment are required [12], [13]. Compared to the other approaches, marker-free vision-based methods are more unobtrusive, convenient, economical, and adaptable [12], [13]. Marker-free methods also demonstrated sufficient accuracy and precision for many common clinical imaging applications [13]–[19].

A marker-free vision-based posture assessment prototype to detect and categorize compensations during robotic rehabilitation therapy was previously developed using the TRI robotic platform [20]. The system used a commercial depth camera (Microsoft Kinect) and its associated skeletal tracking algorithms to track upper body posture of healthy participants. Each participant sat in front of the robot and performed five repetitions of forwards-backwards movement of the robot end-effector. The exercise was repeated five times using either no compensation or one of four common compensatory strategies being simulated: shoulder elevation, trunk rotation, lean forward, or slouching. Compensatory postures were manually labeled and a multiclass classifier was trained based on the orientation of 3-D line segments connecting relevant skeletal points. The four common types of compensation were detected with high accuracy (86%) using a Hidden Markov Support Vector Machine classifier with a modified loss function [21]. However, the classifier was trained and tested on a relatively small sample size (n=8) of healthy participants and the clinical efficacy of the simulation was not validated. Therefore, expanding the dataset and evaluating the algorithms on stroke survivors are a natural extension of this work.

We recently released a more extensive clinical dataset – the Toronto Rehab Stroke Posture (TRSP) dataset [22]. This dataset includes the motions of 10 healthy volunteers and 9 stroke survivors while performing a series of upper limb exercises with the rehabilitation robot developed at TRI.* Two expert raters annotated the postures of stroke survivors. To assess how well the dataset discriminated between different types of compensation, a simple thresholding method was explored which relied on the angle of the spine, or the angle of the line segment connecting the left and right shoulders. The simple thresholding method did not yield satisfactory classification performance and the analysis concluded that a supervised machine learning algorithm considering multiple upper body joints should be explored as a possible way to improve compensation detection accuracy. In this study, we experimentally evaluate supervised classification algorithms to automatically detect compensations in the TRSP dataset.

C. OBJECTIVES

This study targets two objectives. First, we investigated if a classifier could detect common compensations simulated by healthy participants. Second, we examined if a classifier – based solely on patient data – could detect common compensations occurring in stroke survivors.

There are two advantages of using simulated data instead of data acquired from patients directly. First, post-stroke fatigue is a frequently reported and debilitating consequence of stroke [23]. Post-stroke fatigue deprives patients of the physical endurance required to consistently execute therapy exercises and the fatigue generally cannot be temporarily alleviated by rest [23]. Therefore, it is difficult to obtain sufficient training data exclusively from patients. Furthermore, due to varying degrees of physical impairment in stroke patients, some might be incapable of performing motions for a specific or challenging task [24]. The simulated motions, on the other hand, provides a valuable source of data to train the classifiers which can later be used to detect compensations in patients.

Lastly, validating the classifiers against patient data will inform the feasibility of a stand-alone system that detects compensation in real-time and prompts users to correct their posture when necessary.

II. METHODOLOGY

A. EXPERIMENTAL SETUP

The robot used was a tabletop device designed to provide adaptive post-stroke rehabilitation exercises [8]. To operate the robot, the user sat in front of the robot and maneuvered

*The dataset is available at https://www.kaggle.com/derekdb/toronto-robot-stroke-posture-dataset
During rehabilitation exercises, the robot intelligently applies resistive or assistive forces to augment the tasks [8]. However, for simplicity, the robot was powered off during the study to remove the effect of resistive or assistive forces. A Microsoft Kinect v2 sensor was placed approximately 90 cm behind the robot and 30 cm above the tabletop, facing the participant in a slightly oblique direction. Fig. 1 illustrates the robot and a participant operating the end-effector of the robot in color and depth images. Color and depth streams were captured at 30 frames per second using Kinect Studio.

**B. DATASET**

This study was approved by the Research Ethics Board at the Toronto Rehabilitation Institute – University Health Network (TRI-UHN) University Centre. All participants provided written informed consent.

Data extracted from the TRSP dataset contains 3-D skeletal joint positions of 9 stroke survivors with varying degrees of mobility impairment, and 10 healthy participants with no mobility issues. The recruited stroke survivors had varying degrees of upper limb impairment: Chedoke-McMaster Stroke Assessment (CMSA) [25] arm stages 2-6. The healthy participants were matched approximately in height and weight with the stroke survivors.

Only the upper body joint positions were useful for classification because the patients remained seated throughout the exercises and the table occluded vision of the lower body. The joints used for classification included base of the spine, middle of the spine, spine at the shoulder level, shoulders, elbows, wrists, hips, and neck.

Both groups were instructed to perform a series of motions in a seated position using the robot. The patient group completed two basic types of motions defined as reach-forward-back and reach-side-to-side. They were each performed for five repetitions with the unaffected arm and the hemiparetic arm, at a comfortable pace and range of motion for the participant. The details of the motions are shown in Table I. The motions were intended to elicit three types of common post-stroke compensatory synergies: shoulder elevation, trunk rotation, and lean-forward. Participants were allowed to take breaks between repetitions and motions. Because the stroke survivors had varying physical impairment, some of them did not complete all of the motions. The healthy participants also performed the basic motions. Additionally, they simulated the three types of compensations during the basic motions with instructions and demonstration from the research team.

An expert rater with a background in kinesiology visually examined each frame in color streams, and labeled each frame for the presence of compensation and type of compensation.
if applicable. Because the healthy participants did not have compensations during basic motions, all the frames in those motions were labeled as no compensation (NC). For more details on the motions, compensation types, annotation, and participant demographics, please refer to [22].

C. DATA PREPROCESSING
To detect compensations, a supervised machine learning model needed to be trained based on the Kinect skeletal tracking information. Because four classes of labels were involved, a multiclass classifier was required. The Kinect tracking data included joint positions in 3D space (x-y-z) coordinates and joint orientations expressed as quaternions. Experimental results presented here incorporate only the joint positions as features, because preliminary analysis based on both joint positions and orientations did not significantly improve results.

The skeletal joint positions contain noise due to occasional unstable tracking. To reduce such effects, we used a Savitzky–Golay filter with a window size of 31 frames (1.03 s), and a polynomial order of 2. The resulting joint positions and line segments connecting them were animated and visually validated.

The tracked joint positions were expressed in the Kinect-centric coordinate system. Therefore, it was necessary to rotate and translate the coordinates to a participant-centric coordinate system, invariant to sensor placement and orientation [26], [27]. To do so, from the beginning of each recorded segment, a frame was manually selected at the neutral sitting position as the baseline frame. In this frame, the line segment connecting the neck to the head defined the up-and-down direction (y-axis); a component of the line segment connecting the left and right shoulders perpendicular to the y-axis defined the x-axis; and the cross product of the x and y axes defined the z-axis to form a right-handed coordinate frame. The positive directions along the x, y, and z axes respectively pointed toward the left, up, and from the person and toward the camera. By applying a rotation, joint positions were mapped to this coordinate system. Then, the joint coordinates were translated so that the origin (0,0,0) of the new coordinate system was located at the spine-shoulder joint.

The joint coordinates were then normalized to keep the shoulder width at unit length, in order to account for variations in body size and height. Finally, during each task, only one arm was used to manipulate the end-effector and the other one was kept at rest. Because the arm at rest did not contribute to the motions, the x coordinates were inverted so that the actuating arm was always on one side – in our case, the right side.

D. MULTICLASS CLASSIFICATION
A multiclass linear Support Vector Machine (SVM) classifier and a Recurrent Neural Network (RNN) classifier were trained to classify the posture of the participants. The RNN classifier contains a hidden Long Short-Term Memory (LSTM) layer, which captures dynamic temporal behavior of the data. Leave-one-participant-out cross-validation (LOPOCV) was used to assess classifier performance. In each fold of cross-validation, an inner loop of LOPOCV was used for hyperparameter tuning. We also experimented with a multiclass Random Forest classifier and a Softmax classifier, but only the result from the SVM and the RNN was reported because the other classifiers did not achieve superior results.

To replicate the previously reported results in [20] on the simulated compensations, we first trained and tested the classifiers on the posture of healthy participants. Afterwards, we repeated the same method on the stroke survivors. The Receiver Operating Characteristic (ROC) and F1-score for each class were used as performance metrics. F1-score – calculated as the harmonic mean of precision and recall – is more useful than accuracy for unbalanced class distribution.

III. RESULTS
The rater visually examined 115 motions (49,910 frames) of the healthy participants and 35 motions (15,895 frames) of the stroke patients, excluding 5 motions where skeletal tracking failed.

The posture labels are extracted and their distributions are shown in Fig. 2. Compensations are scarcer in patients since the healthy participants were each asked to simulate all compensation while the stroke survivors completed the motions at their own comfort and ability level. Consequently, the label distribution is more unbalanced in the patient group than in the healthy group.

The ROC curves of postures in healthy participants and patients are presented in Fig. 3 using both classifiers. A separate ROC curve is plotted for detecting each type of posture (LF, TR, SE, and NC). Micro-average and macro-average ROC curves are also shown. The figures also report the area under the ROC curve (AUC). Larger AUC values (closer to 1) indicate better classification. Details of the classification results are presented in Table II.

![FIGURE 2. The scatterplot of the #Frames for each participant, sorted by the posture types. NC: no compensation; SE: shoulder-elevation; TR: trunk-rotation; LF: learn-forward.](image-url)
IV. DISCUSSION AND LIMITATIONS

In this study, we aimed to detect compensations during robotic stroke rehabilitation therapy. To our knowledge, this is the first study to investigate the efficacy of marker-free vision-based motion assessment in experiments with stroke survivors. Results demonstrated that supervised machine learning based on the Kinect tracked joint positions can discern certain compensations that occurred in motions simulated by healthy participants. Using the SVM classifier, LF compensations were detected with excellent performance (AUC = 0.98, F1 = 0.82), followed by TR compensation (AUC = 0.77, F1 = 0.57). SE compensations were not well detected (AUC = 0.66, F1 = 0.07). The performance of the RNN was similar. Results from the motion captured from stroke survivors could not be used to train an accurate posture detection classifier. Despite moderate AUC values with both SVM and RNN, F1-scores were low for all three compensations (Table II). Because the dataset contains unbalanced class distribution, F1-scores are a better indicator of classification performance.

There are multiple protocol-related issues that may explain the undesirable performance of the classifier when trained and tested on the patient group. First, the number of compensations in the patient group was limited, which made training a classifier to detect those compensations difficult. This might be due to the fact that the stroke survivors only repeated each motion five times. In contrast, real therapy sessions typically last significantly longer and compensations become more frequent with fatigue. Moreover, the robotic arm was powered off to keep the resistive force consistent for all participants, but this failed to deliver the intensity of real stroke exercises. Additionally, stroke survivors with milder impairments may have recovered enough to regain the motor functions necessary for the completion of the motions in this study without compensation. It is expected that inclusion of more complex tasks involved in a real robotic therapy session will elicit more compensations. Furthermore, there was a large variation in the range of motions, i.e., the distance between the two end-effector stop points during the reach-forward-back and reach-side-to-side motions. As no instructions were given for...
TABLE 2. Classification result using SVM and RNN. LF: learn-forward; TR: trunk-rotation; SE: shoulder-elevation; NC: no compensation; SVM: support vector machine; RNN: recurrent neural network.

<table>
<thead>
<tr>
<th></th>
<th>Healthy Group</th>
<th>Patient Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>SVM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>TR</td>
<td>0.66</td>
<td>0.50</td>
</tr>
<tr>
<td>SE</td>
<td>0.19</td>
<td>0.04</td>
</tr>
<tr>
<td>NC</td>
<td>0.8</td>
<td>0.94</td>
</tr>
<tr>
<td>Weighted Avg</td>
<td>0.73</td>
<td>0.78</td>
</tr>
<tr>
<td>RNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LF</td>
<td>0.84</td>
<td>0.77</td>
</tr>
<tr>
<td>TR</td>
<td>0.66</td>
<td>0.45</td>
</tr>
<tr>
<td>SE</td>
<td>0.31</td>
<td>0.24</td>
</tr>
<tr>
<td>NC</td>
<td>0.81</td>
<td>0.90</td>
</tr>
<tr>
<td>Weighted Avg</td>
<td>0.74</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Stop points, participants whose range of motion was small may have only moved within a comfortable range, thereby not presenting any compensation. Compared to the healthy participants, the stroke survivors generally had smaller range of motion and fewer compensations. This is illustrated in Fig. 4 that compares the postures of a stroke patient and a healthy participant at the rightmost point of their reach-side-to-side motion. It can be clearly seen that at the critical points where TR is most likely to occur, the compensation of the stroke survivor is much less conspicuous compared to that of the healthy participant with an exaggerated simulation of TR. Lastly, the Kinect sensor was pointing at the participant in a slightly oblique direction. As a result, part of the upper body was occluded. When the occluded limb was moving the end-effector, inaccurate tracking occurred, which impacted detection of TR and SE compensations. As such, the aforementioned protocol-related issues may amount to insufficiently amount of compensations in stroke survivors.

FIGURE 4. A comparison between the compensations of a patient (left) and a healthy participant (right). Both were labeled as TR.

Raters can also be subject to visual bias [28] and visual bias that lead to inconsistent labeling. When a rater watches a motion in which compensations don’t occur frequently, a slight deviation in the posture might appear to be more pronounced than it is, and subsequently be labeled as compensation. However, when the rater watches a motion fully of serious compensations, a frame identical to the previous one may be labeled as no compensation, because its less obvious compared to other compensations. Another issue was associated with the structure of the robotic arm. During movements, the user rests their elbow on a support attached to the robotic arm (See Fig. 1b). The end-effector of the robotic arm was raised roughly 2 cm from the table, causing the participant to raise the shoulder of the active arm during arm retraction. From the annotator’s point of view, the shoulder was clearly elevated, although the shoulder elevation might not be due to the compensation, per se. The annotator, therefore, had to mentally account for this contributing factor and assume compensations when they are most likely to occur.

V. CONCLUSIONS

Automated compensation detection using Kinect-tracked joint positions of healthy participants during robotic-assisted rehabilitation achieved excellent performance for LF, followed by TR and SE. However, the same classifiers fared poorly when trained on the data of stroke survivors. We provide potential explanations for low F1-scores. In future research, we recommend the following improvements to the experimental design that should improve compensation detection. Stop points for motions should be pre-defined to elicit compensation, the Kinect sensor should directly face the participant, and the seat should be adjusted so that the shoulder is not elevated at natural sitting position. Further, the robotic arm should be programmed to provide a constant baseline resistive force to match the intensity of real stroke exercises. It will also be useful to record and analyze compensations during a full therapy session when fatigue gradually increases the rate of compensations. The results of our study provide useful information towards development of a posture monitoring system capable of prompting users to correct erroneous postures in real-time.

REFERENCES


