Feature Selection and Classification for Assessment of Chronic Stroke Impairment

Jae-Yoon Jung, Janice I. Glasgow, and Stephen H. Scott

***Abstract*— Recent advances of robotic/mechanical devices enable us to measure a subject's performance in an objective and precise manner. The main issue of using such devices is how to represent huge experimental data compactly in order to analyze and compare them with clinical data efficiently. In this paper, we choose a subset of features from real-time experimental data and build a classifier model to assess stroke patients' upper limb functionality. We compare our model with combinations of different classifiers and ensemble schemes, showing that it outperforms competitors. We also demonstrate that our results from experimental data are consistent with clinical information, and can capture changes of upper-limb functionality over time.**

I. Introduction

Stroke (cerebrovascular accident) is defined as damage to brain tissue caused by the interruption of blood flow to the brain that lasts more than 24 hours [1]. It is the most common cause of disability and affects approximately 700,000 people each year in the United States [2]. Only about ten percent of stroke survivors can fully recover, and impairment often includes upper-limb hemiparesis resulting in a substantive reduction in the quality of life post-stroke [3]. Thus the main medical efforts for stroke patients are focused on rehabilitation and assessment [4].

Recent advances of robotic/mechanical devices enable us to measure a subject's performance in an objective and precise manner [5], [6], whereas most current clinical assessment measures require trained physicians who have specialized knowledge on how to perform the various assessment techniques, and still may suffer from reliability problems or from poor responsiveness [7]. The major issue of using such devices is how to represent huge experimental data compactly in order to analyze and compare them with clinical data efficiently (e.g., see [8]).

In this work, we extract distinctive features from experimental task data, calculate the outlier boundaries for each feature and trial, and build a performance data set which represents upper-limb functionality of chronic stroke patients compared with control subjects group. Based on this data, we implement a hierarchical ensemble network to generate estimation of stroke impairment, and to classify stroke patients' data. The reaching assessment score built from this model is demonstrated with the clinical information, and we show that this model can measure performance changes between repeated experiments, while Chedoke-McMaster scores [4], one of the major outcome measure for stroke impairment and recovery assessment, did not. Finally, we compare our classification results with five other classifier models with/without ensemble schemes, and show that our model outperforms competitors.

Jae-Yoon Jung and Janice I. Glasgow are with the School of Computing, Queen's University, Kingston, Ontario, Canada (email: {jung, janice}@cs.queensu.ca).

Stephen H. Scott is with the Department of Anatomy and Cell Biology, Canadian Institute of Health Research Group in Sensory-Motor Systems, Centre for Neuroscience Studies, Queen's University, Kingston, Ontario, Canada (email: steve@biomed.queensu.ca)

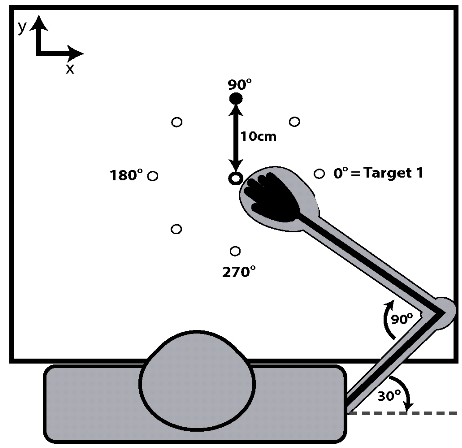

Fig. 1. An experimental setup for the unloaded reaching task. Subjects are asked to reach one of eight targets from the center. The movement cue is given by the illumination of a target. A trial is finished when the subject's hand reaches to the target and stays there. Note that the visual feedback is provided through a projected monitor screen, which prevents subjects from looking their arm movement directly.

II. Methods

A. Participants

Forty six hemiplegic stroke patients and 77 control subjects who report no previous neurological disorders were included in this work. Patients and control subjects were selected with the inclusion criteria of age over twenty and right-dominant handed. As we repeat the same experiment on both of the groups over time, the actual data set we used here consists of 52 left-arm affected stroke patients' experiment data, 53 right-arm affected, and 84 control subjects' data.

B. Experimental Device

We used a robotic exoskeleton platform for this experiment, called KINARM (Kinensiological Instrument for Normal and Altered Reaching Movements, BKIN Technologies, Kingston, ON) [9]. This device enables: 1) to facilitate a subject's flexion and extension movements of the shoulder and elbow with the arm projected on the horizontal plane;

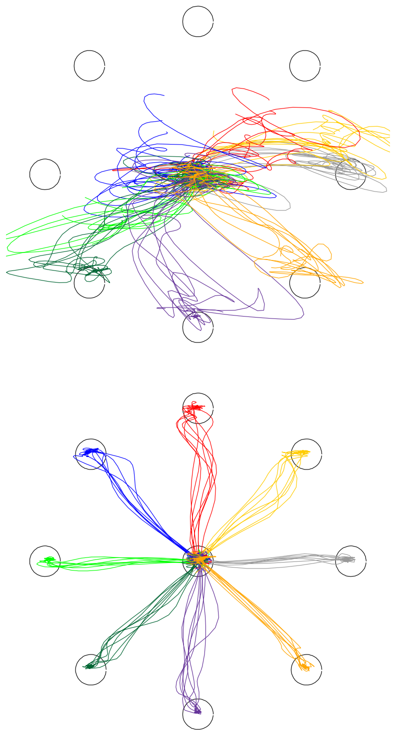

Fig. 2. Examples of the raw hand trajectory data. The upper figure shows a typical stroke subject's hand movements during reaching tasks from the center point to eight different targets, and the lower figure depicts a typical control subject's hand trajectory. Color coding is used in order to designate trials towards the same target direction.

2) to minimize effects of gravity during movements by attaching braces to the upper and lower segments of each arm; 3) to provide a visual feedback through a projected monitor screen, preventing a subject from looking his/her arm movement directly; and 4) to measure and record various aspects of upper-limb motor performance including hand position, tangential hand velocity, shoulder angles, and elbow positions.

C. *Task*

In an experimental session, a subject performed two sets (left and right arm) of an unloaded, center-out reaching task [10], [11] which is illustrated in Fig. 1. Subjects were instructed to reach out from a given center point to one of the eight fixed peripheral targets ($0^\circ = L_1/R_1, 45^\circ, ..., 315^\circ = L_8/R_8$) when a target light is turned on. No restriction on the minimum/maximum velocity was given and subjects were instructed to keep their hand at the target until the target light was turned off. The order of illuminated targets are selected in a random manner, but with a configuration that the total number of repeated trials per each direction would be the same. Fig. 2 and fig. 3 show an example of the hand trajectory and velocity selected from a typical stroke patient's data and a control data.

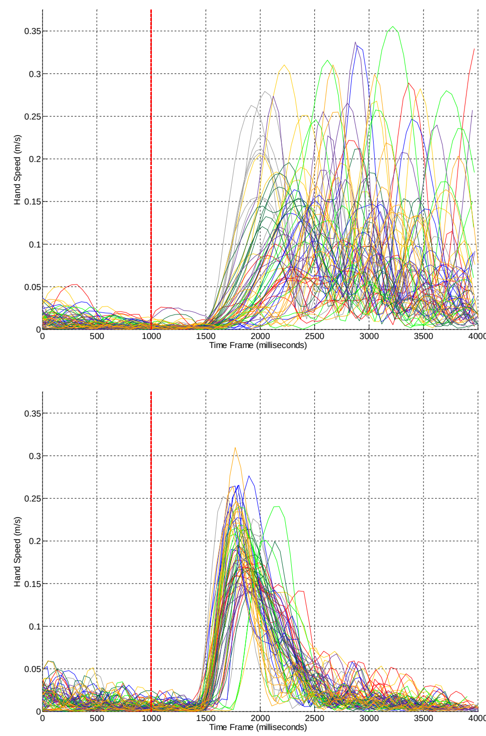

Fig. 3. Examples of the raw tangential hand speed data. For each trial, velocity data are illustrated as if the target light is turned on at time frame 1000. Upper and lower graph show changes of hand speed during reaching task from a stroke patient (upper) and a control subject (lower), respectively. Most of control subjects end trials within three to four second period, while many of stroke patients cannot finish a trial within this period, as shown in the upper graph.

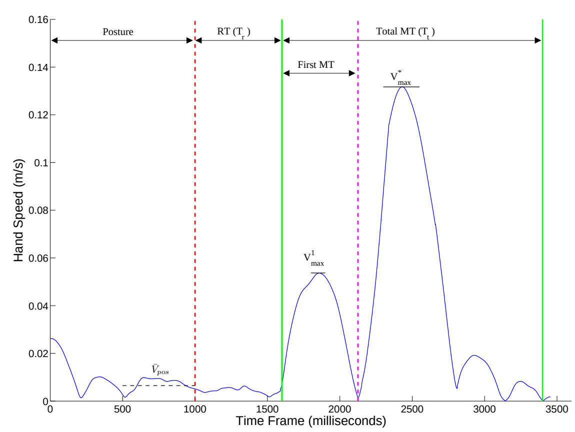

Fig. 4. An illustrative example of tangential hand velocity profile of a single trial. The leftmost (red) vertical line designates the time frame on which the target light is turned on. The following two vertical lines (green) show points when the subject starts moving (onset) and stops moving (offset), respectively. The magenta line between onset and offset specifies the end frame of the first movement in this trial, which is defined by the frame on which the first local minimum velocity occurs after onset.

D. Feature Selection and Data Preprocessing

We selected eight features that capture the main characteristics of the original data [12], [13], [14]. Posture period is defined as the interval between the end of the previous trial and the current target light on, and we calculated the mean speed of 500 milliseconds before target on time ($\bar{V}_{pos}$). Reaction time (T_r) is the time interval until the subject starts to move after target on. First peak velocity (V^1_{max}) is the first local maximum velocity, and the maximum velocity (V^*_{max}) is the global maximum velocity. For most of control cases, V^1_{max} is equal to V^*_{max}, which is not necessarily true for stroke subjects. First movement distance error (E_{dist}) and directional error (E_{dir}) are defined as the error in distance and angular direction relative to the optimal path length and direction during the first movement, which is illustrated in Fig. 4. Total movement time (T_t) is the time between the subject starts to move and stops, and path length ratio (P) is defined as the actual path length during T_t over the optimal distance between the center and the target.

Next, we chose outlier boundaries based on the control data, as specified below. For each feature, all control data was collected and sorted per direction. Assuming each row designates one session data and each column means a set of sorted trials (e.g., $P^{L_1}_{(1)}$ specifies the column of the smallest path length ratio, in the direction L_1), the outlier boundary values were selected either to be $mean \pm 2.58\ \sigma$ if the current column passed Lilliefors' normality test [15] with $\alpha = 0.05$, where σ means the standard deviation of this column. Otherwise, the outlier values were chosen to be the maximum (minimum) value(s) of the current column.

With this boundary set, we calculated the number of outlier trials per each feature and the side of arm, for both of the patient and control subject groups. The training/testing data after this preprocessing consist of 16 attributes (8 features x two sides) per each session.

E. Hierarchical Ensemble Networks

We used a neural network ensemble model as our classifier, as it has shown good performance in our previous works [8], [16]. First, the current training data $T = \{[\bar{V}_{pos}, T_r, T_t, V^1_{max}, V^*_{max}, E_{dist}, E_{dir}, P,\ C]^i \mid i = 1, ...|T|\}$, where the classification label $C \in \{0, 1\}$, was partitioned into 16 subgroups according to the arm tested ($s = L \mid R$) and three randomly partitioned features ($f = 1, ..., 8$). Each subnetwork s_f trained to examine if the stroke patient group and the control group can be separated by these features. A feedforward network with five hidden nodes was used for each subnetwork, trained by a resilient backpropagation algorithm. Next, the intermediate output data from these 16 subnetworks O_{s_f} were fed into the main neural classifier, in order to produce an estimation of this session being a control subject (= 1.0) or not (= 0.0). Finally, this whole procedure is repeated with ten times of 10-fold cross validation [17] to obtain a generalized performance expectation for this ensemble classifier.

III. Results

A. Reaching Assessment Score

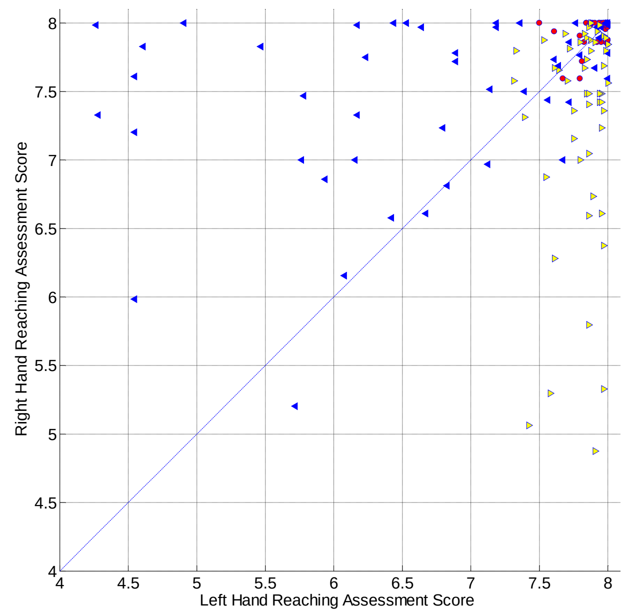

Fig. 5. Left and right reaching assessment scores are shown in two-dimensional plane. Each marker represents one session data. Red circles, blue left triangles, yellow right triangles specify control subjects, left arm affected stroke patients, and right arm affected stroke patients respectively. The affected arm information is independently reported through clinical assessments, not estimated from the experimental data or our ensemble network results.

Based on the final sub-classifier outputs O_{s_f} for each session, we constructed an outcome index that measures motor performance of each arm. More specifically, the reaching assessment score of session i was determined as a sum of O_{s_f} values,

$$score^i = [\sum_{f=1}^{8} O^i_{Lf} \sum_{f=1}^{8} O^i_{Rf}]$$

The maximum score of $[8.0\ 8.0]$ means that the subject showed no upper limb deficits during the experimental task, whereas the minimum score of $[0.0\ 0.0]$ implies the opposite. Fig. 5 illustrates the result combined with the affected arm information, independently collected through clinical examinations. Many of the stroke session scores appear on [7.5 8.0] range, but it is not surprising in a sense that about 45 percent of stroke session data (47/105 sessions) were diagnosed clinically intact by Chedoke-McMaster measure.

B. Responsiveness

Next, we check the responsiveness (i.e., sensitivity to changes within patients over time [4]) of our measure by comparing the reaching scores from the same subject with Chedoke-McMaster scores, which are measured by physicians. Fig. 6 plots the changes of the reaching score while the corresponding Chedoke-McMaster scores remain the same during various time intervals. Each line segment depicts the result of two sessions done by the same subject, and the x-axis represents the time interval between two sessions.

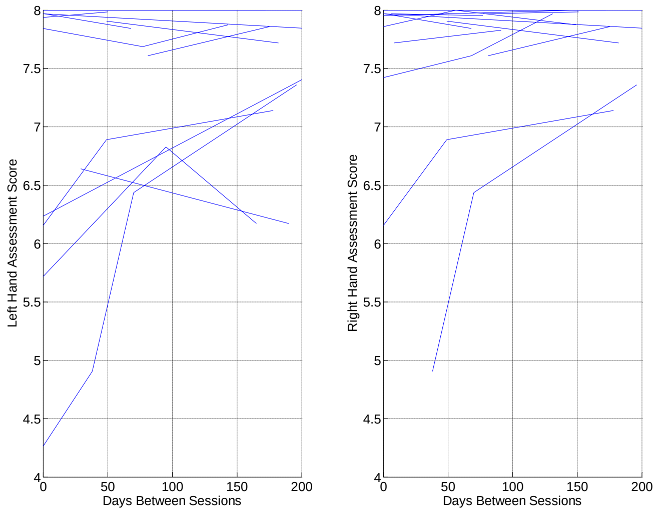

Fig. 6. Plots of left and right reaching assessment score changes over time. Each line segment represents changes of our assessment score between two sessions done by the same subject, over the given time intervals specified in x-axis. Note that only changes while the corresponding clinical-based Chedoke-McMaster scores remain fixed are shown in this figure. The first session date is set as day zero for all sessions.

TABLE I

CLASSIFICATION PERFORMANCE COMPARISON

Classifier Type	SS	CC	CS	SC	Err (%)
k-NN	79.7	86.3	18.7	4.3	12.2
k-NN + Bagging	80.7	85.0	20.0	3.3	12.3
NaiveBayes	80.3	87.7	17.3	3.7	11.1
NaiveBayes + Bagging	80.7	87.0	18.0	3.3	11.3
NaiveBayes + Boosting	80.3	84.7	20.3	3.7	12.7
Log. Regression	74.0	94.4	10.6	10.0	10.9
Log. Regression + Bagging	76.7	95.0	10.0	7.3	9.2
SVM	79.7	89.7	15.3	4.3	10.4
SVM + Bagging	80.0	92.0	13.0	4.0	9.0
SVM + Boosting	78.7	93.0	12.0	5.3	9.2
Decision Tree	76.3	92.0	13.0	7.7	11.0
Decision Tree + Bagging	76.2	96.1	8.9	7.8	8.8
Decision Tree + Boosting	78.9	93.5	11.5	5.1	8.8
Ensemble Networks	79.7	98.0	7.0	4.3	6.0

Most lines move towards performing better than the previous assessment, showing that our measure are responsive in both short term (e.g., a few days) or relatively long term (6 months or more) periods in the rehabilitation process.

C. *Classification Performance*

Five different classification algorithms were considered for performance comparison: k-nearest neighbor [18], Naive-Bayes [19], Logistic Regression Models [20], Support Vector Machines [21], and Decision Tree (C4.5) [22]. Bagging [23] and boosting (AdaBoost.M1) [24] are combined with above classifiers in order to build an ensemble, and the option of ten iterations and a resampling ratio of 1.0 was applied to both schemes.

Table I summarizes the results for the different classification algorithms. Column SS and CC correspond to the average number of session data set that were correctly classified as stroke and control. Column CS and SC correspond to the average number of incorrect classifications as stroke to control and vice versa. The error column shows the misclassification rate of each classifier in percentage, averaged over ten iterations of 10-fold cross validation procedures. The maximum possible number of SS and CC are 84 and 105 respectively, and the lowest error possible from blind estimation is 44.4 percent for this data.

We tried 1 to 10 nearest neighbor options and Table I shows the best classification rate among them. Ensemble version of classifiers typically performed about two percent better than the original version, as shown in regression, SVM, and decision tree models, but nearest neighbor and naive Bayes algorithm did not get such performance boost in this work. We also tried different type of kernels and its parameters including linear, polynomial, radial basis function and sigmoid kernels for SVM, and the best results are shown above. Our ensemble classifier outperforms all competitors, but the performance difference is not statistically significant. However, the error rate is significantly lower in general compared with our previous result [16], in which similar feature data and ensemble networks were used.

IV. CONCLUSION

In this work, we extracted characteristic features from real-time, experimental task data, reduced data set by summarizing features into a compact outlier sets, and trained a hierarchical ensemble network model in order to identify stroke patients and assessing their upper limb functionality. The reaching assessment score proposed here is calculated from the partial output of sub-network classifiers, and we showed that this outcome measure coincides with clinical assessment information of affected arm. We also showed that this score can capture the changes of functionality over time, whereas Chedoke-McMaster score remained the same. The classification performance was compared with other algorithms and ensemble schemes including nearest neighbor, SVM, logistic regression, naive-bayes, and decision tree models, and our network outperformed all compared classifier models.

Our future goal would be to discover relationship between the experimental data and more subtle clinical assessment information, including type of strokes, lesion location, and CT/MRI data.

V. ACKNOWLEDGEMENT

We thank K. Moore and H. Bretzke for technical help. This work was supported by NSERC and Canadian Institutes of Health Research (CHIR) MOP 81366.

REFERENCES

[1] S. B. O'Sullivan and T. J. Schmitz, *Physical Rehabilitation*, F. A. Davis Company, PA, USA, 2007.

[2] W. Rosamond *et al*, "Heart Disease and Stroke Statistics—2007 Update", *Journal of The American Heart Association*, Circulation vol. 115, e69-e171, 2007.

[3] M. Vestling, B. Tufvesson, and S. Iwarsson, "Indicators for return to work after stroke and the importance of work for subjective well-being and life satisfaction", *Journal of Rehabilitation Medicine*, 35:127-131, 2003.

[4] K. Salter, J. Jutai, L. Zettler, M. Moses, N. Foley, and R. Teasell, *Evidence-Based Review of Stroke Rehabilitation (EBRSR)*, Heart and Stroke Foundation of Ontario and Ministry of Health and Long-Term Care of Ontario, 2007.

[5] S. H. Scott, "Apparatus for measuring and perturbing shoulder and elbow joint positions and torques during reaching", *Journal of Neuroscience Methods*, 89:119-127, 1999.

[6] D. Nozaki, I. Kurtzer, and S. H. Scott, "Limited transfer of learning between unimanual and bimanual skills within the same limb", *Nature Neuroscience*, 9 (11):1364-1366, 2006.

[7] D. J. Gladstone, C. J. Danells and S. E. Black, "The fugl-meyer assessment of motor recovery after stroke: a critical review of its measurement properties", *Neurorehabilitation and Neural Repair*, 16 (3):232-240, 2002.

[8] J. Y. Jung, J. I. Glasgow, and S. H. Scott, " Trial Map: A Visualization Approach for Verification of Stroke Impairment Assessment Database", In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 4113-4116, 2008.

[9] K. Singh and S. H. Scott, "A motor learning strategy reflects neural circuitry for limb control", *Nature Neuroscience*, 6 (4):399-403, 2003.

[10] A. P. Georgopoulos, J. F. Kalaska, R. Caminiti, and J. T. Massey, "On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex", *Journal of Neuroscience*, 2:1527-1537, 1982.

[11] K. M. Graham, K. D. Moore, D. W. Cabel, P. L. Gribble, P. Cisek, and S. H. Scott, "Kinematics and kinetics of multi-joint reaching in non-human primates", *Journal of Neurophysiology*, 89:2667-2777, 2003.

[12] K. M. Zackowski, A. W. Dromerick, S. A. Sahrman, W. T. Thach, and A. J. Bastain, "How do strength, sensation, spasticity and joint individuation relate to reaching deficits of people with chronic hemiparesis?", *Brain*, 127:1035-1046, 2004.

[13] R. F. Beer, J. P. A. Dewald, and W. Z. Rymer, "Deficits in the coordination of multijoint arm movements in patients with hemiparesis: evidence for disturbed control of limb dynamics", *Experimental Brain Research*, 131:305-319, 2000.

[14] P. Archambault, P. Pideon, A. G. Feldman, and M. F. Levin, "Recruitment and sequencing of different degrees of freedom during pointing movements involving the trunk in healthy and hemiparetic subjects" *Experimental Brain Research*, 126:55-67, 1999.

[15] H. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown", *Journal of the American Statistical Association*, 62:399-402, 1967.

[16] J. Y. Jung, J. I. Glasgow, and S. H. Scott, " A Hierarchical Ensemble Model for Automated Assessment of Stroke Impairment", In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 3186-3190, 2008.

[17] R. Kohabi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 2:1137-1145, 1995.

[18] G. Shakhnarovich, T. Darrell, and P, Indyk (Eds.), *Nearest-Neighbor Methods in Learning and Vision*, MIT Press, 2006.

[19] P. Domingos and M. J. Pazzani, "Beyond independence: conditions for the optimality of the simple bayesian classifier", In *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 105-112, 1996.

[20] A. Agresti, *Categorical Data Analysis*, Wiley-Interscience, 2002.

[21] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press Cambridge, MA, USA, 2001.

[22] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann San Mateo, CA, USA, 1992.

[23] L. Breiman, "Bagging predictors", *Machine Learning*, 24 (2):123-140, 1996.

[24] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm", In *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 148-156, 1996.