Patient Similarity and Joint Features for Rehabilitation Outcome Prediction

Gina L. Sprint
Diane J. Cook
Washington State University, Pullman, WA
{gsprint,cook}@eecs.wsu.edu

Douglas L. Weeks
St. Luke's Rehabilitation Institute
Spokane, WA
weeksdl@st-lukes.org

Abstract

Rehabilitation outcome prediction can be useful for clinicians providing therapy services for patients undergoing rehabilitation. Machine learning models trained with medical record information available at admission can predict rehabilitation outcomes fairly well. In our earlier work we found rehabilitation outcome prediction accuracy can be improved by also including sensor-based features that objectively quantify patient movement abilities during therapy tasks. In this paper, we hypothesize an even greater improvement in rehabilitation outcome prediction accuracy can be obtained with our proposed algorithm, called Joint Patient Prediction (JPP), which is based on patient similarity and joint prediction. To validate this hypothesis, we collected wearable inertial sensor data from 27 patients as they performed ambulatory tasks. We utilize features derived from the collected wearable sensor data and clinical information obtained from medical records at admission as inputs to our JPP algorithm. Results indicate we are able to achieve higher prediction accuracy when utilizing our proposed JPP approach over the baseline, non-JPP models. The proposed JPP method is a useful technique for mapping wearable sensor data collected during therapy tasks into rehabilitation outcome measures.

1 Introduction

Often when an individual suffers from an injury or illness, such as stroke, they undergo intense inpatient rehabilitation in order to regain everyday functioning. Patient functioning is typically assessed by therapists using standardized clinical rating scales, such as the Functional Independence Measure (FIM), to determine independence in activities of daily living [Hamilton *et al.*, 1987]. The FIM is administered at admission and discharge from inpatient rehabilitation by clinical staff who are credentialed to administer the instrument. Between the admission and discharge FIM assessments, clinical observations by therapists are typically used to characterize progress and make treatment decisions, including determining discharge dates. More precise, objective measurements of patient performance can be collected via pervasive technology, such as wearable inertial measurement units (IMUs). Features derived from wearable IMU sensor data can identify subtle performance changes during rehabilitation that are difficult to observe. In our earlier work [Sprint *et al.*, 2015b], we demonstrated that such IMU-based features can provide power for predicting future patient performance on clinical rating scales. Specifically, we utilized a combination of sensor-based features and admission medical record information to predict discharge FIM scores for 20 inpatient rehabilitation patients. Such predictions of discharge rehabilitation outcome measures can be used to help evaluate patient progress and prepare for discharge decisions.

In this paper, we utilize additional participants' wearable sensor data (N = 27) that we have collected since our previous analyses. We report updated discharge FIM motor score prediction accuracy for models trained with the additional wearable sensor data collected mid-stay of rehabilitation. We also propose a machine learning methodology called Joint Patient Prediction (JPP) that utilizes patient similarity and joint prediction techniques. We hypothesize JPP will improve the accuracy of discharge FIM motor score prediction. To validate our hypothesis, we utilize JPP to predict FIM motor scores at discharge and compare the results to baseline predictions made by models without JPP. Our approach provides insight into patient progress between admission and discharge by using movement data collected during therapy tasks, without the need to re-administer the entire FIM assessment.

2 Related Work

Several studies have investigated mapping technology-based measurements into clinical assessment scores [Zariffa *et al.*, 2012; Olesh *et al.*, 2014; Wang *et al.*, 2014; Simila *et al.*, 2014]. These studies have primarily examined the relationships between technology-based metrics and clinical rating scores measured close in time (e.g. concurrently). In contrast, Mostafavi *et al.* utilized data collected from a rehabilitation robotic device to project into the future and predict discharge assessment scores of FIM total score, FIM motor score, length of hospital stay, the Purdue Pegboard Test, and the Modified Ashworth score with statistically significant accuracy [Mostafavi *et al.*, 2013].

Research related to patient similarity and joint prediction represents a recent direction of research. With the advent of electronic health records, a notion of similarity between people for healthcare applications has emerged, called "patient similarity". Researchers have utilized patient similarity to propose novel electronic health record data mining techniques [Wang and Sun, 2015; Hielscher *et al.*, 2014; Klenk *et al.*, 2010; Chan *et al.*, 2010]. For joint prediction, work by Minor *et al.* applied joint prediction to a multi-output regression learner trained to forecast activity occurrences in smart home environments [Minor *et al.*, 2015]. For their forecasting approach, sensor-based features (X_{sensor}) were mapped into predictions of the time until each activity $a \in A$ would occur (the target variable Y), where $A = \{a_1, a_2, ..., a_T\}$ is a set of T activity labels. The researchers applied joint prediction to the activity forecasting problem by augmenting the sensor-based features (X_{sensor}) with previous occurrence predictions for each activity ($\hat{Y}$) as joint features: $X = X_{sensor} \oplus \hat{Y}$. Minor and colleagues reported a 85.11% decrease in prediction error when utilizing joint features over the baseline model (no joint prediction). Furthermore, the researchers also explored utilizing ground truth time values (Y) for the joint features in lieu of $\hat{Y}$ predictions. In this case, the associated predictor was called the *oracle predictor*.

3 Methods

We combine clinical and sensor-based features to form training data for machine learning models constructed to predict discharge clinical assessment scores. Specifically, we utilize clinical and sensor-based features to predict discharge FIM motor scores for rehabilitation patients. The FIM measures functional status on a 0-7 rating scale for 18 items representing 6 domains: self-care, sphincter control, transfers, locomotion, communication, and social cognition. In addition to a total FIM score, separate scores are developed from the motor function items and cognitive function items. To predict discharge FIM scores, we collect data from patients undergoing inpatient rehabilitation and train machine learning models. Our approach consists of the following four steps:

1. Collect wearable inertial sensor data from patients as they ambulate throughout an ecological environment (see Section 3.1).
2. Extract sensor-based features from the sensor data and medical record information (see Section 3.2).
3. Train and test machine learning models to predict discharge FIM scores for the patients (see Section 3.3).
4. Improve upon initial prediction results by applying patient similarity and joint prediction techniques (see Section 3.4).

The following sections provide details on each of these steps.

3.1 Ambulation Circuit Study

Participants (N = 27; Male = 19, Female = 8) were recruited from the inpatient rehabilitation population at a large inpatient rehabilitation facility. Recruited participants were mostly recovering from stroke (N = 20) and non-traumatic brain injuries (N = 3). As a group, the participants exhibited admission FIM motor scores of 36.15 ± 9.84 and discharge FIM motor scores of 64.96 ± 11.25. To collect movement

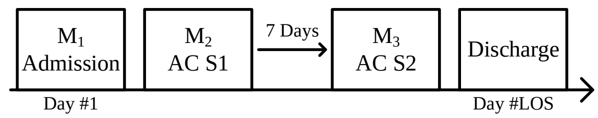

Figure 1: The ambulation circuit wearable sensor study timeline. LOS refers to length of stay, calculated as the number of days between admission and discharge inclusive. M_1, M_2, and M_3 refer to prediction models (see Section 3.3).

data, we attached three inertial sensors (Shimmer3, sampling frequency of 51.2 Hz) to the bodies of participants. One sensor was placed on the center of mass and one sensor on each shank. While wearing the sensors, participants performed an ambulation circuit (AC), a continuous sequence of activities in a simulated community environment at the rehabilitation facility. The AC includes rising from a seated position in a chair, moving with both linear and curvilinear gait, surface transitions, a transfer into and out of a sport utility vehicle, and sitting back down in the chair. We collected AC data at two different testing sessions (see Figure 1 for an overview of the AC study timeline). The first session (S1) occurred shortly after the participant became physically able to walk the distance required of the gait task (11.26 ± 5.51 days from admission). The second session occurred one week later, a date that was typically close to their discharge (2.82 ± 2.76 days before discharge). In summary, the AC is an extension of the common clinical assessment, the traditional Timed Up and Go (TUG) test [Sprint *et al.*, 2015a], including a greater range of functional tasks (e.g., car transfers) and situational challenges (e.g., different flooring surfaces; a curvilinear pathway). Although the AC environment is unique to the facility, the majority of the AC mobility features we report (see Table 2) can be computed from any assessment in any environment involving a chair transfer and walking (5 Times Sit-to-Stand, TUG, etc.).

We process the inertial movement data collected from the AC sensors with algorithms we designed for the AC data. First, we align the timestamps from the three different sensor platforms. Next, we correct for the orientation of the shank sensors by transforming the sensor local coordinate system to the body coordinate system [Chen, 2013]; a right handed system with the X-axis along the anterior-posterior body axis, the Y-axis along the vertical body axis, and the Z-axis along the medial-lateral body axis. We then filter acceleration data with a 4^{th} order zero-phase band pass Butterworth filter using cutoff frequencies of 0.1 Hz and 3 Hz for the COM accelerometer and 0.1 Hz and 10 Hz for the shanks. The gyroscope signals for all sensors are low passed filtered at 4 Hz.

3.2 Feature Extraction

As AC participants underwent rehabilitation, data became available at four points displaced in time: admission, AC S1, AC S2, and discharge. Metrics computed from data collected from admission, AC S1, and AC S2 serve as features to machine learning models, which are trained to predict FIM motor scores at discharge (see Figure 1 for an overview of the

Table 1: Features extracted from medical records available at admission.

Feature	Description
Age	Age in years
Gender	Male or female
RIC	Rehabilitation impairment category
Comorbidity tier	No relevant comorbidities, tier 1 (most severe/expensive), tier 2 (medium severe/expensive), or tier 3 (least severe/expensive)
Case mix group relative weight	Modifier determined by comorbidities and complications
FIM_A motor score	Sum of the 13 FIM motor task scores
FIM_A cognitive score	Sum of the 5 FIM cognitive task scores
Reciprocal FIM_A motor score	Reciprocal of admission FIM score [4]
17 FIM_A task scores	17 total scores, one score for each FIM task

A = admission, FIM = functional independence measure.

AC study timeline).

Features derived from admission medical record information include patient characteristics such as age, gender, and rehabilitation impairment category, as well as individual FIM task scores (see Table 1 for all admission features). In addition, we include the reciprocal of the FIM motor score as suggested by Sonoda and colleagues [Sonoda *et al.*, 2005]. Although additional data from medical records are available, we only include features that apply to all populations. For example, the number of days since stroke onset is only applicable to stroke populations and is not included as a predictor. Sensor-based metrics of AC performance include commonly-used approaches for assessing mobility in a clinical setting, such as the duration of a task; metrics computed from the sensor placed on the COM; and gait features, which refer to quantifications of steps and strides while walking. Gait features are computed from gait cycles identified by gait cycle event detection algorithms we apply to the shank sensor angular velocity signals [Greene *et al.*, 2010]. Table 2 summarizes the sensor-based features. At AC S2, we compute additional features to quantify the changes exhibited over one week of therapy from S1 to S2 for each AC performance feature, including the percentage change and the standardized mean difference effect size for repeated measures [Wolff Smith and Beretvas, 2009].

3.3 Baseline Prediction Models

We express discharge FIM motor score prediction as a supervised learning task that maps the admission and sensor-based features to discharge FIM motor score predictions. For generating predictions, we make use of epsilon support vector machines (ϵ-SVM). SVMs utilize a subset of the training data, called support vectors, to identify boundaries of maximal distance from the support vectors. In the case of regression, the SVM learns a function $F(x) \rightarrow w{\cdot}x{-}b$ to approximate a target variable y_i under ϵ precision for each feature vector x_i. The vector w is the learned weights, or coefficients, representing the relative importance of each feature for the SVM.

Table 2: Features computed from wearable inertial sensor data.

Feature	Description
Duration	Time to complete an AC task
Floor surface speed ratio	Ratio of walking speed on different floor surfaces
Walking speed	Distance divided by time
COM peak angular velocity	Maximum rotational velocity of the COM around the Z-axis
Root mean square (RMS) (normalized by time)	Square root of the mean of the squares of each COM acceleration signal
Smoothness index (harmonic ratio)	Ratio of even to odd Y-axis COM acceleration harmonics
Smoothness of RMS (normalized by time)	RMS of COM acceleration signal derivatives
Cadence	Steps taken per minute
Double support percent	Percentage of the gait cycle that both feet are on the ground
Gait cycle time	Stride duration
Number of gait cycles	Number of complete gait cycles (strides)
Shank peak angular velocity	Maximum Z-axis gait cycle angular velocity of the shank
Shank range of motion	Range of integrated Z-axis gait cycle angular velocity
Step length	Distance between steps
Step regularity	Regularity of the acceleration of sequential steps
Stride regularity	Regularity of the acceleration of sequential strides
Step symmetry	Ratio of step to stride regularity

AC = ambulation circuit, COM = center of mass, RMS = root mean square.

We construct three different SVM models to predict rehabilitation outcomes for AC participants, one model for each of the three different points in time when data were collected. M_1 is a model trained with data available upon admission, M_2 is a model trained with data available at AC S1, and M_3 is a model trained with data available at AC S2. Figure 1 depicts this timeline and the associated models. Each model M_1, M_2, and M_3 produces a prediction (P_1, P_2, and P_3) for the same clinical outcome (see Figure 2), i.e. discharge FIM motor scores. These predictions represent the change in model prediction accuracy over time as additional data are collected.

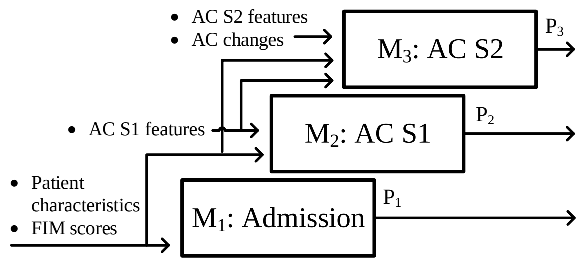

Figure 2: Ambulation circuit (AC) model construction approach for M_1 (admission), M_2 (AC session 1), and M_3 (AC session 2).

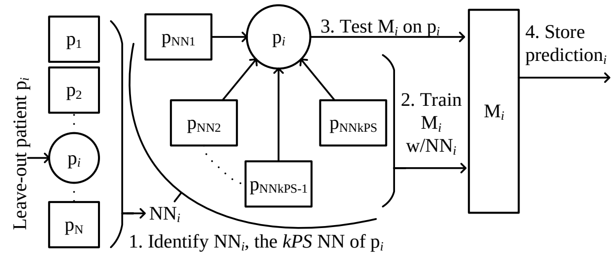

Figure 3: Using patient similarity to select training samples for a leave-out patient p_i in leave-one-out cross validation. NN stands for nearest neighbors.

In order to remove noisy or redundant features, we apply feature selection techniques. First, we correlate individual features with discharge FIM motor scores to investigate their individual predictive ability. Features with a Pearson correlation coefficient $r < 0.1$ are determined to be noisy and not used during model training. Next, we apply a wrapper-based recursive feature elimination algorithm with cross validation (RFECV) to identify the optimal set of features [Kohavi and John, 1997]. The model we used in RFECV is a linear SVM trained with 10-fold cross validation with mean squared error scoring. The top-ranked RFECV features are then selected as the inputs to the prediction models.

3.4 Joint Patient Prediction

To achieve higher prediction accuracy, we propose model enhancements based on patient similarity (PS) and joint prediction (JP). We hypothesize a combination of PS and JP techniques, called Joint Patient Prediction, will improve accuracy of discharge FIM motor score prediction for the AC participants. While the techniques we propose here are presented in the context of inpatient rehabilitation, our patient similarity and joint prediction approaches are general and suitable for other application domains.

Patient Similarity

We propose to utilize patient similarity to improve our prediction approach. In machine learning, k nearest neighbors (NN), or k-NN [Cover and Hart, 1967], is a supervised learning algorithm that identifies the k labeled training data points that are closest to an unlabeled data point. To measure proximity, a distance function computes a dissimilarity measure between two samples. The unlabeled point is then labeled using the label that appears most frequently among the k closest training data points (nearest neighbors). In the case of patient similarity, a distance function, like those used in k-NN, computes a dissimilarity measure between two patients. For the distance function in our PS approach, we use the heterogeneous Euclidean overlap metric (HEOM) [Wilson and Martinez, 1997] that was utilized in previous patient similarity work by Hielscher *et al.* [Hielscher *et al.*, 2014]. We compute a pairwise distance matrix, *dist*, relating each AC participant to each other using HEOM. Let the current test patient held out during each fold of leave-one-out cross validation (LOOCV) be denoted as $p_i, 1 \leq i \leq N$. For our proposed PS algorithm, *dist* is queried to identify the kPS ($1 \leq kPS < N$) nearest neighbors of the test participant p_i. The set of kPS nearest neighbors, NN_i, forms the training set for the current fold. This approach trains a model using only patients that are most similar to the current test patient. The diagram presented in Figure 3 outlines the patient similarity algorithm.

There are several options for features to utilize when comparing two patients during generation of the distance matrix (*dist*). Lane and colleagues proposed different types of similarity networks for building communities of smartphone users for activity inference [Lane *et al.*, 2011]. The researchers utilized physical, lifestyle, and sensor-data similarity networks. We adapt the concept of multiple similarity networks proposed by Lane and colleagues by exploring the following patient similarity features used for computing *dist*:

1. $dist_{demo}$: Using only admission demographic features (see Section 3.2 for a description of the admission features).
2. $dist_{FIM-A}$: Using only admission FIM scores.
3. $dist_{M1}$, $dist_{M2}$, $dist_{M3}$: Using only the top ranked features as determined by RFECV (see Section 3.3 for details about feature selection) for each model M_1, M_2, and M_3, respectfully.

In the next section, we discuss our proposed joint prediction algorithm that makes use of the pairwise patient distance matrices.

Joint Prediction

Joint prediction represents a recent direction of machine learning research. To explain joint prediction, consider a model M trained to learn a mapping from a feature space X to a target variable Y, $M : X \to Y$. For joint prediction, each sample $x_j \in X$ is augmented with additional "joint" features. The joint features depend on the machine learning problem formulation, but are typically features of the same nature as the target variable Y. To apply the concept of joint prediction to rehabilitation outcome prediction, we include discharge FIM motor score predictions ($\hat{Y}$) from other participants in the training set as joint features. The predictions from other patients can be achieved by one of the following methods, where p_i is the leave-out patient in LOOCV:

Algorithm 1 JointPatientPrediction(X, Y, kPS, kJP)

Input: X = N × M matrix of N patients' feature vectors x, $|x| = \mathrm{M}$
Input: Y = vector of ground truth values for each patient in X
Input: kPS = the number of nearest neighbors to include in the training set
Input: kPS = the number of joint features to include
Output: $\hat{Y}$ = a vector of N patients' predictions
Compute $dist$, a N × N matrix of pairwise distances between patients in X
for each patient feature vector x_j in X:
  $NN_j = dist[: kJP]$ // kJP nearest neighbors of x_j
  $NNJP_j = Y[NN_j]$ // the joint features
  $x_j = x_j \oplus NNJP_j$ // adding the joint features
end for
Perform feature selection
for each patient feature vector x_i in X:
  Initialize: prediction model M_i
  $NN_i = dist[: kPS]$ // kPS nearest neighbors of p_i
  Train M_i with NN_i
  $\hat{Y}_i$ = Test M_i on x_i
  Store $\hat{Y}_i$ at $\hat{Y}[i]$
end for
return prediction vector $\hat{Y}$

- Oracle: Utilize the ground truth values (Y) from N-1 other patients in the training set as features to predict the ground truth value for the current patient p_i.
- Single pass: Train a model with N-1 patients (holding out p_i). Then, using the same model that was just trained, make N-1 predictions ($\hat{Y}$) for the same N-1 patients used to train the model. Use these N-1 predictions as joint features to predict the ground truth value for the current patient p_i.
- Double pass: Using a second layer of LOOCV, generate predictions for N-1 patients ($\hat{Y}$). Use these N-1 predictions as joint features to predict the ground truth value for the current patient p_i.

The differences between the above joint prediction methods lies in which "predictions" are included as joint features. The choice for which of the aforementioned approaches to use depends on the application context. If ground truth information for all training patients is available, then the oracle approach can be utilized. However, if ground truth information is not available for all patients, then predictions from a previously trained model need to be substituted for ground truth information. Both of the non-oracle predictors (single and double pass) will produce predictions representative of "true" predictions; however, the double pass predictor approach should produce less accurate predictions than the single pass predictor since the model has not previously been trained on the test samples in LOOCV. In this paper, we consider oracle and single pass joint features. Future work involves investigating the effects of double pass joint prediction for larger datasets.

In our joint prediction approach, we utilize concepts from patient similarity that were presented in Section 3.4. Specifically, we use the pairwise distance matrices to identify similar participants to include their discharge FIM motor scores as joint features. We also use patient similarity to pare down the size of the training set each fold as described previously. We name the combined patient similarity and joint prediction approach Joint Patient Prediction. For JPP, the kJP ($1 \leq kJP < N$) nearest neighbors of each patient p_j are identified as the set NN_j. NN_j can be generated by querying a distance matrix (see Section 3.4 for details regarding the computation of distance matrices). For JPP, we fix kJP to 3 joint features for M_1, 4 joint features for M_2, and 5 joint features for M_3. We compute these joint features by querying the nearest neighbor of p_j in each of the distance matrices $dist_{demo}$, $dist_{FIM-A}$, $dist_{M1}$, $dist_{M2}$, and $dist_{M3}$. Consequently, the models M_1, M_2, and M_3 will each have 3, 4, and 5 joint features, respectively. For example, NN_j for M_2 is generated as $NN_j = \{NN_{demo}, NN_{FIM-A}, NN_{M1}, NN_{M2}\}$. The target Y values of the samples in NN_j (discharge FIM scores for rehabilitation outcome prediction) are extracted as the set of joint features of j, $NNJP_j$. This approach to generate $NNJP_j$ produces kJP joint features that are the nearest neighbors to p_j based on different categories of similarity. Finally, the feature vector of p_j, x_j, is augmented with the joint features: $x_j = x_j \oplus NNJP_j$. Algorithm 1, Joint Patient Prediction, presents a general version of the JPP algorithm with a single distance matrix $dist$ and an example LOOCV model M_i.

4 Results

Prior to training, admission and AC data are standardized by subtracting the mean and scaling to unit variance. SVMs with a linear kernel are trained and evaluated using LOOCV. The regression models are evaluated using root mean square error (RMSE), normalized RMSE (NRMSE), and Pearson correlation coefficients (r). To JPP, we compare our JPP results for each model (M_1, M_2, M_3) to the prediction results obtained for each model without JPP (baseline). For further explanation, we compare M_1 without JPP to M_1 with JPP, M_2 without JPP to M_2 with JPP, and M_3 without JPP to M_3 with JPP. We test the JPP models with oracle joint features, and again with single pass joint features. For each JPP and non-JPP set of models, we also test multiple parameter values for patient similarity. Specifically we run experiments with $kPS = \{1, 2, ..., N-1\}$. Experimental results reveal single pass joint features yield higher prediction accuracy over oracle joint features. Consequently, we report JPP results based on single pass joint features.

For M_1, there are 8 JPP configurations with improved prediction accuracy over the baseline M_1 predictor. The top 5 M_1 configuration results, sorted by ascending RMSE, are shown in Table 3. For M_2, there are 26 JPP configurations with improved prediction accuracy over the baseline M_2 model. Table 4 shows the top 5 prediction results for M_2. For M_3, 3 JPP configurations yield improved predictions and these results are shown in Table 5.

Table 3: M_1 joint patient prediction results for discharge functional independence measure (FIM) motor score. Only the top 5 experiment results demonstrating improvement over the baseline M_1 predictor are shown.

kPS	Joint Features?	RMSE	NRMSE	*r*
25	Yes	6.70	16.75%	0.74**
24	Yes	6.85	17.12%	0.73**
26	Yes	6.91	17.28%	0.72**
22	Yes	7.14	17.86%	0.70**
25	No	7.14	17.86%	0.70**
26 (baseline)	No	7.52	18.80%	0.66**

NRMSE = normalized root mean square error, r = Pearson correlation coefficient, RMSE = root mean square error, ** = $p < 0.01$.

Table 4: M_2 joint patient prediction results for discharge functional independence measure (FIM) motor score. Only the top 5 experiment results demonstrating improvement over the baseline M_2 predictor are shown.

kPS	Joint Features?	RMSE	NRMSE	*r*
18	Yes	5.05	12.62%	0.88**†
17	Yes	5.20	12.99%	0.88**†
19	Yes	5.27	13.17%	0.86**†
16	Yes	5.30	13.24%	0.87**
23	Yes	5.40	13.49%	0.85**
26 (baseline)	No	6.69	16.72%	0.77**

NRMSE = normalized root mean square error, r = Pearson correlation coefficient, RMSE = root mean square error, ** = $p < 0.01$, † = significantly ($p < 0.10$) improved results from baseline M_2.

5 Discussion

In this paper, we train regression models to predict discharge FIM motor scores for 27 inpatient rehabilitation participants who performed two testing sessions of the ambulation circuit. We hypothesize that patient similarity and joint prediction would improve accuracy for predicting discharge FIM motor scores. Results indicate our JPP algorithm does improve prediction accuracy for M_1 (see Table 3), M_2 (see Table 4), and M_3 (see Table 5). JPP results for M_1 and M_3 indicate higher values of kPS (kPS = 22, 23, 24, 25, and 26) yield the highest prediction accuracy. This result suggests larger training sets could be more useful for training admission and AC S2 models, but not necessarily for training the AC S1 model. For M_1, there are also two experiments (kPS = 24, 25) without joint features that yielded improved prediction results over the baseline M_1 predictor. This observation provides evidence that patient similarity techniques can improve prediction accuracy on their own.

For M_2, three configurations of JPP yield results that are significantly ($p < 0.10$) more accurate than the baseline M_2 model (see Table 4). Furthermore, M_2 models with joint features are more accurate (kPS = 18, RMSE = 5.05) than the best M_3 predictor with JPP (kPS = 26, RMSE = 6.08). This

Table 5: M_3 joint patient prediction results for discharge functional independence measure (FIM) motor score. Only results demonstrating improvement over the baseline M_3 predictor are shown.

kPS	Joint Features?	RMSE	NRMSE	*r*
26	Yes	6.08	15.21%	0.79**
25	Yes	6.21	15.52%	0.78**
24	Yes	6.31	15.78%	0.77**
26 (baseline)	No	6.35	15.87%	0.77**

NRMSE = normalized root mean square error, r = Pearson correlation coefficient, RMSE = root mean square error, ** = $p < 0.01$.

result suggests that with JPP, only one AC session is sufficient to produce highly accurate FIM motor score predictions from wearable sensor data collected mid-stay of inpatient rehabilitation. Essentially, JPP can eliminate the need for a second AC data collection session.

Limitations of this study include the low sample size and all data are collected from the same inpatient hospital. A wide variety of patients attending other rehabilitation facilities would be more representative of the population and potential clinical utility of the models. Finally, the AC participant population is primarily recovering from a stroke (74.07%). A wider variety of patient impairments would be more representative of all types of patients admitted to inpatient rehabilitation. Future work aims to further improve prediction results by investigating additional patient similarity techniques and JPP configurations, particularly for M_1 and M_3 models. We also aim to explore the effects of double pass joint features and apply the Joint Patient Prediction algorithm to larger datasets, such as data collected from electronic health records.

6 Conclusions

We investigated the predictive abilities of features derived from wearable inertial sensor data to predict rehabilitation outcomes, as measured by the discharge Functional Independence Measure, without the need to re-administer the FIM assessment battery. While baseline models trained with clinical and sensor-based features performed well, we were able to achieve an even higher level of prediction accuracy when utilizing our proposed Joint Patient Prediction algorithm. Correlations as high as $r = 0.88$ (RMSE = 5.05, NRMSE = 12.62%) for LOOCV predicting discharge FIM motor score were obtained. A wearable sensor system and associated rehabilitation outcome prediction algorithms, like JPP, are potentially useful tools to inform therapists and help them better provide services to their patients during recovery.

Acknowledgments

We wish to thank our therapist collaborators at St. Luke's Rehabilitation Institute and Vladimir Borisov for their insights and help with data collection. This material is based upon work supported by the National Science Foundation under Grant No. 0900781.

References

[Chan *et al.*, 2010] L.W.C. Chan, T. Chan, L.F. Cheng, and W.S. Mak. Machine Learning of Patient Similarity: A Case Study on Predicting Survival in Cancer Patient After Locoregional Chemotherapy. In *2010 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, pages 467–470, December 2010.

[Chen, 2013] Shanshan Chen. *Gait Feature Extraction from Inertial Body Sensor Networks for Medical Applications*. PhD thesis, University of Virginia, 2013.

[Cover and Hart, 1967] T. Cover and P. Hart. Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13(1):21–27, January 1967.

[Greene *et al.*, 2010] Barry R. Greene, Alan O. Donovan, Roman Romero-Ortuno, Lisa Cogan, Cliodhna Ni Scanaill, and Rose A. Kenny. Quantitative Falls Risk Assessment Using the Timed Up and Go Test. *IEEE Transactions on Biomedical Engineering*, 57(12):2918–2926, December 2010.

[Hamilton *et al.*, 1987] B. B. Hamilton, C.V. Granger, F.S. Sherwin, M Zielezny, and J.S. Tashman. A Uniform National Data System for Medical Rehabilitation. In *Rehabilitation Outcomes: Analysis and Measurement*, pages 137–150. Paul H. Brookes Publishing Company, Baltimore, MD, 1987.

[Hielscher *et al.*, 2014] T. Hielscher, M. Spiliopoulou, H. Volzke, and J.P. Kuhn. Using Participant Similarity for the Classification of Epidemiological Data on Hepatic Steatosis. In *2014 IEEE 27th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1–7, May 2014.

[Klenk *et al.*, 2010] Sebastian Klenk, Jurgen Dippon, Peter Fritz, and Gunther Heidemann. Determining Patient Similarity in Medical Social Networks. In *Proceedings of the First International Workshop on Web Science and Information Exchange in the Medical Web*, pages 6–14, 2010.

[Kohavi and John, 1997] Ron Kohavi and George H. John. Wrappers for Feature Subset Selection. *Artificial Intelligence*, 97(12):273–324, 1997.

[Lane *et al.*, 2011] Nicholas D. Lane, Ye Xu, Hong Lu, Shaohan Hu, Tanzeem Choudhury, Andrew T. Campbell, and Feng Zhao. Enabling Large-Scale Human Activity Inference on Smartphones Using Community Similarity Networks (CSN). In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 355–364. ACM, 2011.

[Minor *et al.*, 2015] Bryan Minor, Janardhan Rao Doppa, and Diane J. Cook. Data-Driven Activity Prediction: Algorithms, Evaluation Methodology, and Applications. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 805–814, New York, NY, USA, 2015. ACM.

[Mostafavi *et al.*, 2013] S.M. Mostafavi, J.I Glasgow, S.P. Dukelow, S.H. Scott, and P. Mousavi. Prediction of Stroke-Related Diagnostic and Prognostic Measures Using Robot-Based Evaluation. In *2013 IEEE International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6, June 2013.

[Olesh *et al.*, 2014] Erienne V. Olesh, Sergiy Yakovenko, and Valeriya Gritsenko. Automated Assessment of Upper Extremity Movement Impairment due to Stroke. *PLoS ONE*, 9(8), 2014.

[Simila *et al.*, 2014] Heidi Simila, Jani Mantyjarvi, Juho Merilahti, Mikko Lindholm, and Miikka Ermes. Accelerometry-Based Berg Balance Scale Score Estimation. *IEEE Journal of Biomedical and Health Informatics*, 18(4):1114–1121, 2014.

[Sonoda *et al.*, 2005] Shigeru Sonoda, Eiichi Saitoh, Shota Nagai, Yuko Okuyama, Toru Suzuki, and Miho Suzuki. Stroke Outcome Prediction Using Reciprocal Number of Initial Activities of Daily Living Status. *Journal of Stroke and Cerebrovascular Diseases: The Official Journal of National Stroke Association*, 14(1):8–11, 2005.

[Sprint *et al.*, 2015a] G. Sprint, D.J. Cook, and D.L. Weeks. Toward Automating Clinical Assessments: A Survey of the Timed Up and Go. *Biomedical Engineering, IEEE Reviews in*, 8:64–77, 2015.

[Sprint *et al.*, 2015b] G. Sprint, D.J. Cook, D.L. Weeks, and V. Borisov. Predicting Functional Independence Measure Scores During Rehabilitation With Wearable Inertial Sensors. *IEEE Access*, 3:1350–1366, 2015.

[Wang and Sun, 2015] Fei Wang and Jimeng Sun. PSF: A Unified Patient Similarity Evaluation Framework Through Metric Learning With Weak Supervision. *IEEE Journal of Biomedical and Health Informatics*, 19(3):1053–1060, May 2015.

[Wang *et al.*, 2014] Jingli Wang, Lei Yu, Jiping Wang, Liquan Guo, Xudong Gu, and Qiang Fang. Automated Fugl-Meyer Assessment using SVR model. In *Bioelectronics and Bioinformatics (ISBB), 2014 IEEE International Symposium on*, pages 1–4. IEEE, 2014.

[Wilson and Martinez, 1997] D. Randall Wilson and Tony R. Martinez. Improved Heterogeneous Distance Functions. *Journal of artificial intelligence research*, pages 1–34, 1997.

[Wolff Smith and Beretvas, 2009] Lindsey Wolff Smith and S. Natasha Beretvas. Estimation of the Standardized Mean Difference for Repeated Measures Designs. *Journal of Modern Applied Statistical Methods*, 8(2):600–609, 2009. 00000.

[Zariffa *et al.*, 2012] J. Zariffa, N. Kapadia, J. L. K. Kramer, P. Taylor, M. Alizadeh-Meghrazi, V. Zivanovic, U. Albisser, R. Willms, A. Townson, A. Curt, M. R. Popovic, and J. D. Steeves. Relationship Between Clinical Assessments of Function and Measurements From an Upper-Limb Robotic Rehabilitation Device in Cervical Spinal Cord Injury. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(3):341–350, 2012.