Title
Objective Assessment of Motor Deficits in Patients with Degenerative Spinal Cord Disorders

Author
Lee, Sunghoon

Publication Date
2014

Peer reviewed|Thesis/dissertation

University of California

Los Angeles

Objective Assessment of Motor Deficits in Patients with Degenerative Spinal Cord Disorders

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Computer Science

by

Sunghoon Lee

2014

Abstract of the Dissertation

Objective Assessment of Motor Deficits in Patients with Degenerative Spinal Cord Disorders

by

Sunghoon Lee

Doctor of Philosophy in Computer Science

University of California, Los Angeles, 2014

Professor Majid Sarrafzadeh, Chair

Degenerative spinal disorders such as cervical spondylotic myelopathy (CSM) or lumbar spinal stenosis (LSS) are caused by many factors including prolonged and inappropriate sitting positions, heavy labor, and genetically inherited predisposition. These ailments are closely related to the aging process of the human body, and their populations are growing rapidly in developed nations. Thus, there is a great need for a simple, inexpensive, reliable, and objective assessment system.

This dissertation introduces a system for assessing hand motor dysfunction, which is a common symptom of CSM, using a mobile handgrip device with simple target tracking tests. Data collected from 30 CSM patients and 30 age-matched control subjects is used to validate the system's three fundamental quantification objectives using a reliable feature selection technique: detecting the severity of hand motor deficits, correlation to the patients' perceived motor deficits, and detecting the changes in motor conditions as a result of medical treatment (e.g., surgical operation). Then, this dissertation introduces a more in-depth algorithm that quantifies the level of hyperreflexia, which is an unique symptom of CSM patients closely related to their fine motor controllability. This thesis further discusses a prediction algorithm that estimates the postoperative functional

outcomes of CSM patients using their clinical, demographic, and preoperative functional (handgrip) information. The handgrip device has also been tested on patients with other types of neuromotor ailments such as cerebral vascular accident (CVA) and chronic inflammatory demyelinating polyneuropathy (CIDP) through a pilot study. A technical solution to quantification of motor function for patients suffering from lower back spinal disorders (e.g., LSS) is briefly discussed as a future work.

The dissertation of Sunghoon Lee is approved.

Wei Wang

Douglas Stott Parker, Jr.

Daniel C. Lu

Majid Sarrafzadeh, Committee Chair

University of California, Los Angeles

2014

To my parents . . .

who made this book possible through

their endless love and support (and money . . .)

and

To my wife . . .

who shared all the tears and laughter.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am most grateful to my committee members, Dr. Majid Sarrafzadeh, Dr. Daniel C. Lu, Dr. Stott D. Parker Jr., and Dr. Wei Wang for their support, encouragement, and guidance throughout my studies. I would like give my special thanks to Professor Majid Sarrafzadeh for his excellent supervision throughout my M.S. and Ph.D. studies. I would also like to thank Dr. Daniel C. Lu who has been always supportive towards my research. I would like to thank Ruth Getachew, Charles Li, Hadyn Hoffman, Derek C. Lu, and Andrew Campion for their collaborative works. Last but not least, I would like to thank Nima Ghalehsari, Mehrdad Razaghy , Brian H. Paak, Jordan H. Garst, Marie Espinal, Nima Jahanforouz, Amir Ali Ghavam, and Marwa Afridi for their voluntary dedications to data collection and patient interaction.

Vita

2008	B.A.Sc. (Computer Engineering), Simon Fraser University, Burnaby, British Columbia, Canada.
2010	M.S. (Electrical Engineering), UCLA, Los Angeles, California, USA.
2013	M.S. (Computer Science), UCLA, Los Angeles, California, USA.

Publications

Sunghoon Ivan Lee, Charles Li, Haydn A. Hoffman, Derek S. Lu, Ruth Getachew, Bobak Mortazavi, Jordan H. Garst, Marie Espinal, Mehrdad Razaghy, Nima Ghalehsari, Brian H. Paak, Amir A. Chavam, Marwa Afridi, Arsha Ostowari, Hassan Ghasemzadeh, Daniel C. Lu, Majid Sarrafzadeh, "Quantitative Assessment of Hand Motor Function in Cervical Spinal Disorder Patients Using Target Tracking Tests," IEEE Transactions on Neural Systems and Rehabilitation Engineering (IEEE TNSRE), submitted for review.

Sunghoon Ivan Lee, Bobak Mortazavi, Haydn A. Hoffman, Derek S. Lu, Charles Li, Brian H. Paak, Jordan H. Garst, Mehrdad Razaghy, Marie Espinal, Eunjeong park, Daniel C. Lu, Majid Sarrafzadeh, "A Prediction Model for Functional Outcomes in Spinal Cord Injury Patients using Gaussian Process Regression," IEEE Journal of Biomedical and Health Informatics (J-BHI), submitted for review.

Eunjeong Park, Sunghoon Ivan Lee, Sunghwan Oh, Andrew Campion, Nima Ghalehsari, Sangsoo Park, Daniel Lu, Majid Sarrafzadeh, "A Wearable System to Recognize Alcohol-Induced Gait Using Sensor-Equipped Smart Shoes," IEEE Transactions on Human-Machine Systems, submitted for review.

Ruth Getachew, Sunghoon Ivan Lee, Andrew Yew, Jon Kimball, Derek S. Lu, Jordan H. Garst, Nima Ghalehsari, Brian H. Paak, Mehrdad Razaghy, Marie Espinal, Arha Ostowari, Amir Ghavamrezaii, Sahar Pourtaheri, Majid Sarrafzadeh, Daniel C. Lu, Utilization of A Novel Digital Measurement Tool for Quantitative Assessment of Upper Extremity Motor Dexterity: A Controlled Pilot Study, J Neuroeng Rehabil, accepted for publication.

Sunghoon Ivan Lee, Hassan Ghasemzadeh, Bobak Jack Mortazavi, Majid Sarrafzadeh, "Pervasive Assessment of Motor Function: A Lightweight Grip Strength Tracking System," IEEE Journal of Biomedical and Health Informatics (J-BHI) (previously IEEE TITB, SCI-Indexed), vol. 17, no. 6, pp. 1023-1030, 2013.

Sunghoon Ivan Lee, Hassan Ghasemzadeh, Bobak Jack Mortazavi, Andrew Yew, Ruth Getachew, Mehrdad Razaghy, Nima Ghalehsari, Brian H. Paak, Jordan H. Garst, Marie Espinal, Jon Kimball, Daniel C. Lu, Majid Sarrafzadeh, "Objective Assessment of Overexcited Hand Movements using a Lightweight Sensory Device," 2013 IEEE Body Sensor Network Conference (IEEE BSN'13), MIT, USA, May, 2013.

Sunghoon Ivan Lee, Jonathan Woodbridge, Ani Nahapetian, Majid Sarrafzadeh, "MARHS: Mobility Assessment System with Remote Healthcare Functionality for Movement Disorders," ACM SIGHIT International Health Informatics Symposium 2012 (SIGHIT IHI 2012), Miami, USA, January, 2012.

CHAPTER 1

Introduction

Cervical spondylotic myelopathy (CSM) is a degenerative spinal disorder in the cervical (i.e., neck) region, which is the most common spinal cord dysfunction in adults who are over the age of fifty in North America [You00, TB10]. CSM is closely related to the aging process of the human body, which can be caused by many factors including inappropriate sitting postures, heavy labor or genetically inherited predisposition [You00,Kli10]. As a consequence of the trend towards aging societies and changes in working environments that involve prolonged sitting, the CSM population is expected to grow rapidly in developed nations [You00]. Chronic disk degeneration, inflammatory diseases or other soft tissue abnormality caused by CSM often result in significant pressure on the spinal cord or nerve roots. A major complaint of CSM patients is the impairment of hand motor function [DSH06], which produces symptoms such as loss of hand dexterity, numbness, stiffness, weakness, fatigue, and tremor. These symptoms may develop to severe weakness or complete paralysis [FJH12]. Virtually all CSM patients have radiographic evidence (e.g., X-ray, MRI, or CT images) of degenerative changes in their cervical spine by age 30, which often develops until 60 years of age [HT04, You00]. Thus, frequent monitoring of physical conditions in CSM patients is essential for the purposes of assessing the ailment progress, evaluating the results of a medical treatment, and preventing possible onset of the ailment.

Unfortunately, frequent radiographical testing is extremely costly, especially for weekly, monthly or yearly monitoring. Consequently, current methods of

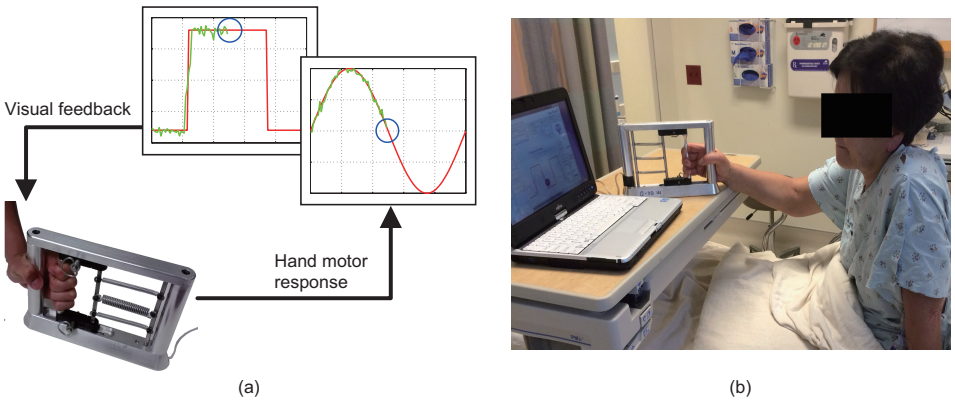

Figure 1.1: (a) The handgrip device and two tracking waveforms used to in this study (left: step, right: sinusoidal). (b) A picture of a CSM patient performing the test before her surgical operation.

clinically assessing the ailment progress or neuromotor deficits heavily rely on patient-reported outcomes such as Oswestry Disability Index [FP00], Japanese Orthopedic Association [FCK07], or American Spinal Injury Association motor scores [ETK96]. However, these methods suffer from variabilities among responders and most importantly, some of these methods are known to carry response shift. Response shift refers to changes in an individual's internal standard of perceived health status, which often occurs after treatment such as a surgical operation [Fin10]. These make the method especially difficult for longitudinal tracking of patient progress [SGC05,STC13]. Clinical maneuver (e.g., Hoffmann's sign [DM91]) is another frequently used method. However, it often has limited ordinal scales and suffers from variations in observer's standard for evaluating motor function, which makes them difficult to be used for longitudinal patient monitoring [CWC11]. As a consequence, a simple, objective, reliable, and accurate movement assessment system for quantifying motor deficits of patient CSM patients is in great need [SD07].

Handgrip motor function has recently received attention as a simple, accurate, and economical assessment methodology [JRK09] since the hand muscle control-

lability is closely related to the quality of life and the motor function required for activities of daily living (ADL) such as eating, writing, or picking up small objects [SD07]. For instance, authors in [SD07] have investigated various methods for utilizing handgrip strength to determine mobility and self-care ability. However, the work in [SD07] only considers the patient's ability to exert a certain level of grip forces rather than the fine controllability of their hand muscles, which certainly has greater correlation with ADL [KGG05, MI94, SN00].

Target tracking tests based on handgrip strength, which quantify fine hand motor controllability, have been well studied in other ailments carrying neuromotor deficits such as stroke [NPL11, KZB04, LGM13a], Parkinson's Disease (PD) [PBC10, KZB04], brain injury [KHM95, KGG05] and chronic inflammatory demyelinating polyneuropathy (CIDP) [LGM13a]. Target tracking tests visualize a predefined waveform that a subject must track by adjusting their hand muscle strength in order to minimize the error between the waveform and the subject's response. Quantification of fine motor function based on tracking error has shown to be significantly larger in patients with neuromuscular diseases (e.g., stroke or PD) compared to control subjects [KZB04]. Furthermore, in a previous study [LGM13a], it has been shown that the results of target tracking contain motor characteristics that are specific to ailments such as stroke or CIDP. However, to the best knowledge of the authors, use of target tracking tests to quantify hand motor deficits in chronic spinal cord disorder, as well as its correlation to the perceived motor deficits, has not yet been investigated.

This dissertation provides a comprehensive overview of the use of target tracking tests for objectively quantifying motor deficits in patients with degenerative spinal cord disorders. The system employs a portable and mobile handgrip device that can be easily deployed at home or hospital environments. The data analytic methodologies discussed in this work will allow to (i) measure the level of motor deficits in hand dexterity, (ii) identify motor features that have the greatest cor-

relation to the perceived motor dysfunction and quality of life, (iii) identify the changes in functional outcomes as a result of medical treatment, and (iv) predict the postoperative motor outcomes of patients based on their preoperative motor conditions and demographic/clinical information. The technical contributions of this paper are the following.

First, a novel system that measures the hand motor capacity using a lightweight and inexpensive handgrip device is introduced. This portable sensing platform provides various dimensions of motor characteristics such that the known symptoms of the associated ailments can be quantified.

Second, the reliability and validity of the target tracking test using the handgrip device are validated based on a metric that employs multi-dimensional motor features (which are extracted based on the known symptoms of CSM). Machine learning algorithms and feature selection techniques are utilized to investigate the system's (i) ability to detect the appearance of ailment compared to the age-matched controls, (ii) correlation to the perceived level of motor deficits, and (iii) ability to detect any changes in motor function as a result of medical treatment.

Third, a method specifically designed for analyzing the level of hyperexcitability in hand muscles, which is a unique symptom of CSM patients, is introduced. The analysis includes a series of four signal processing units: (i) the pre-processing unit, (ii) the abnormality (i.e. activation hypertonia) detection unit, (iii) the abnormality analytic unit, and (iv) the parameter extraction unit. The preprocessing unit performs a low-pass filter to reduce noise in the raw signals, and segments the signals into a number of subsignals. The detection unit statistically determines whether a resultant subsignal contains the outcome of the exaggerated muscle tone using machine learning algorithms. If activation hypertonia is noted, the analytic unit performs an in-depth analysis to locate important geometric points using dynamic time warping (DTW). The parameter extraction unit extracts important variables that characterize the severity of activation hypertonia.

Fourth, a method that predicts the postoperative functional outcomes of CSM patients based on their preoperative information is introduced. This is a typical example of a prediction problem that involves a finite set of independent variables (i.e., preoperative data of patients) and its associated noisy observation of a dependent variable (i.e., postoperative data of patients). The proposed method employs Gaussian Process Regression that predicts the observation variable using covariances among training data and their geometric positions within the feature dimension.

Fifth, a pilot study that applies the proposed system to other types of neuromotor ailments is discussed. In this study, various motor characteristics that are uniquely observed among Cerebral Vascular Accident (CVA or also known as stroke) and Chronic Inflammatory Demyelinating Polyneuropahty (CIDP) patients are analyzed. This pilot study enables new opportunities for accurate quantification of an individual's ailment specification, disease severity, and specific physiological symptoms.

The rest of this dissertation is organized as follows. In Chapter 2, the reliability and validity of the target tracking test in CSM patients are discussed. Chapter 3 and Chapter 4 discuss the in-depth analysis for hyperexcitability and the method for predicting postoperative functional outcomes, respectively. Chapter 5 elaborates the pilot study on CVA and CIDP patients. Future work and conclusions are discussed in Chapter 6.

CHAPTER 2

Validity and Reliability of the System

2.1 Objectives

This section investigates three different analytic objectives using a reliable feature selection technique: (i) analyzing the severity of motor deficits in preoperative CSM patients by comparing to age-matched control subjects, (ii) finding correlation of the motor characteristics of CSM patients to their perceived level of motor deficits, and (iii) detection of changes in motor functions as a result of medical treatment.

2.2 Materials

2.2.1 Participants

A total of 19 CSM patients (12 males) with a mean age of 62.8 and a standard deviation of 14.6 participated in the study. A total of 33 CSM patients were initially recruited from the UCLA Spine Center. 11 patients refused to participate in the study after their operations. Data belonging to 3 patients were corrupted due to system malfunction, and was removed from the study. As a consequence, this cohort study involves preoperative data of 19 patients, and 17 of them have returned to the clinic for follow-ups with minimal time span of 3 months after the surgery. All patients were verified for their anatomic evidence of CSM (e.g., location or the level of the cervical spinal injury) using conventional X-ray im-

ages. All patients received spinal cord decompression to alleviate pressure on the impacted nerve root, thereby improving motor function and associated pain. The surgical operation was performed by a single spine surgeon. A total of 19 age-matched control subjects (10 male) with ages of 59.1 ± 7.1 were recruited from the general population. The inclusion criteria were age and having no history of any neuromotor impairment.

The examination procedure was approved by the UCLA institutional review board. All participants provided consent after an explanation of the study protocol and the associated risks.

2.2.2 Sensing Platform

This work utilizes a handgrip device for quantifying fine hand motor capacity, which was introduced in [LGM13b] (Fig. 1.1). The major components of the handgrip device are the springs, the handle, and the displacement sensor in the body. The handle of the device is connected to the main frame by three springs, which provide physical resistance for grasping performance. These springs can be easily replaced to accommodate participants with different grip strength. This study used five springs with different tension forces (i.e., 0.38lbs/in, 0.88lbs/in, 1.94lbs/in, 5.10lbs/in, 10.7lbs/in), which allow a total of $5 \times 5 \times 5 = 125$ different spring combinations. The displacement sensor is embedded in the bottom of the frame that captures the absolute position of the handle at a sampling rate of 32Hz.

2.2.3 Examination Procedure

Participants start the examination procedure by measuring their maximum voluntary contraction (MVC), which represents the maximum grip force that participants can voluntarily exert. The measured MVC is used to normalize the maximum possible amplitude of the target tracking tests. The waveform within

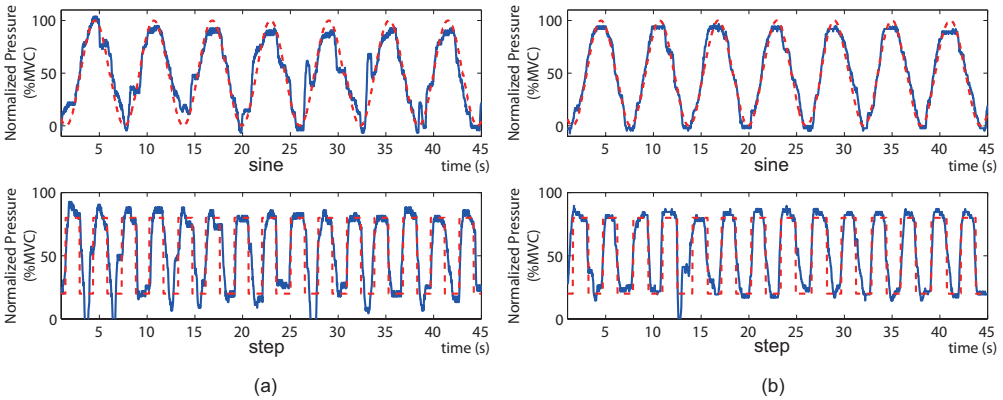

Figure 2.1: Sample test results collected from (a) CSM patients before operation, and (b) age-matched control subjects.

the screen moves to the left while the horizontal position of the blue circle is fixed in the middle of the x-axis as shown in Fig. 1.1 (a). The vertical position of the blue circle changes according to the grip force generated by participants. The screen also displays the trace history of the patient for visual feedback. The length of the test is 45 seconds.

Fine motor capacity of participants was tested using two different targets: sinusoidal and step waveforms. The sinusoidal waveform has a period of 6.17 seconds (i.e., 0.16Hz), which results in approximately seven sinusoidal cycles within a 45 seconds test. The amplitude of the waveform covers from 0 to 100% of the subject's MVC as illustrated in Fig. 2.1; the unit of the amplitude is in %MVC. The sinusoidal waveform investigates the motor-learning capability for repeated fine muscle control [Jon00]. The step waveform has a period of 3 seconds (i.e., 0.33Hz) with 50% duty cycle, which looks like a clock pulse as shown in Fig. 2.1. The higher amplitude (ceiling) is equal to 80%MVC and the lower amplitude (floor) is equal to 20%MVC. The step waveform investigates predictive tracking and ability to execute smooth constant velocity movements [Jon00]. Participants repeated each waveform three times per clinical visit. The first test (out of the three tests) is considered as a practice trial and was completely removed from

further analysis.

2.2.4 Patient Reported Functional Outcome

Participants reported the perceived level of motor impairments in performing ADL using two different patient reported outcomes: Oswestry Disability Index (ODI) [FP00] and modified Japanese Orthopaedic Association (mJOA) motor score [BLK91].

ODI is one of the well-known measures of perceived motor functions and the quality of life for patients with spinal injuries [FP00, TM94]. The survey contains ten six-multiple-choice questionnaires related to the motor function and the quality of life. The accumulated score ranges from 0 (completely disabled) to 50 (no dysfunction)[1], which is linearly scaled to $[0, 1]$ for simplicity.

mJOA is another well-known measure of motor deficits in CSM patients [BLK91]. This survey contains four multiple-choice questionnaires related to upper extremity and hand motor function. The accumulated score of mJOA ranges from 0 (completely disabled) to 18 (no dysfunction), which is again linearly scaled to $[0, 1]$.

2.3 Methods

Quantifying comprehensive motor function is a challenging task since there exists no ground truth labeling of the data [SC01]; true physical status of a patient is not known. As a consequence, researchers in the medical community emphasize reliability and validity of a new motor measure through various techniques including (i) test-retest reliability, (ii) detection of ailment, (iii) respon-

[1]The original ODI takes the measurement of 0 being completely healthy and 50 being completely disabled, but in this work we follow the *general systems performance theory* that all dimensions of human performance are in a form for which a higher numerical value represents superior performance [Jon00, Kon95]

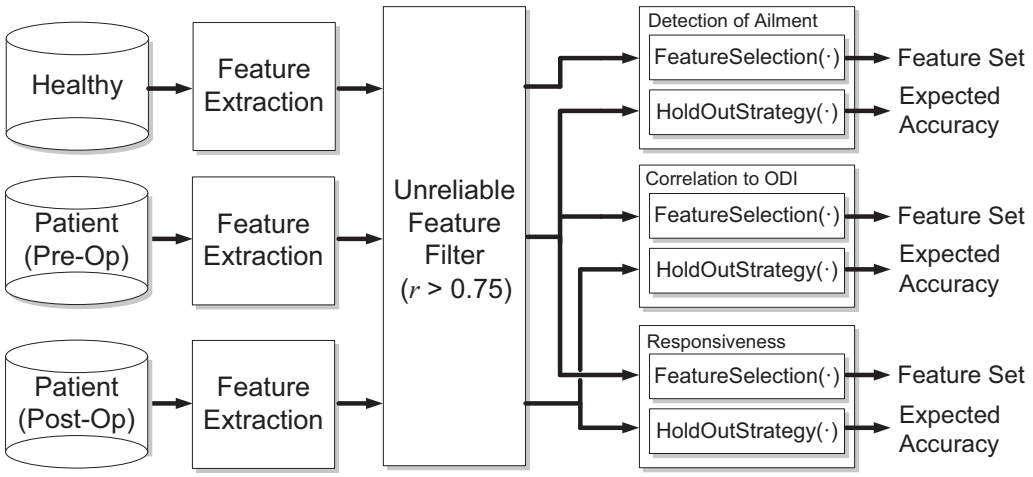

Figure 2.2: The data analysis quantifying various motor characteristics of CSM patients.

siveness to changes in physical conditions, and (iv) correlation to other existing methods [KKF97, MUC01, FP00]. Test-retest reliability examines the device's ability to produce similar results for subjects with similar physiological conditions [FCK07, GHW93]. Detection of ailment investigates the device's ability to detect the presence of ailment compared to healthy subjects [WW04]. Responsiveness defines the device's ability to detect changes in physical conditions of patients as a result of medical treatment [HWL09, DC86]. Correlation to other existing methods ensures that there exists a degree of correlation between the new medical device and existing methods [GHW93, GKW97, HN92, LBF97].

Interestingly, some of these methods have been frequently investigated in the engineering communities. For instance, detection of ailment is often formalized as a typical machine learning classification problem to discriminate patients from control subjects [PRM11, PMA13, LGM13a]. The major difference in addressing these objectives (ailment detection, responsiveness or correlation) between the medical and engineering communities is that the medical community usually incorporate a single feature to construct the metric [NPL11, PBC10] while the engineering community employs feature selection algorithms to construct the metric

in a multi-dimensional feature space. A recent work in [PMA13] proposed a motor function quantification method that employs the test-retest reliability criteria as a pre-filtering process in order to eliminate features that are not reliable. This provides a reliable motor quantification metric since the decision function of a classifier is constructed by polynomially (often linearly) combining those reliable features. The quantification method proposed in this chapter leverages and expands upon the method proposed in [PMA13].

The objectives of this chapter are to investigate (i) detection (quantification) of hand motor deficits of CSM patients, (ii) monitoring of ailment progress after medical treatment (i.e., responsiveness), and (iii) correlation of the hand motor function to the perceived level of motor deficits in performing ADL (i.e., correlation to other method). This work employs a feature selection technique that selects a subset of *reliable* features in order to address the aforementioned objectives as illustrated in Fig. 2.2. Addressing these three objectives separately hypothesizes that the motor characteristics uniquely observed among CSM patients compared to control subjects may not be necessarily identical to the motor characteristics that have the strongest correlation to the perceived quality of life nor to the motor characteristics that determine the effectiveness of a medical treatment. Section 2.3.1 investigates hand motor features considered in this work. Section 2.3.2 discusses the method for eliminating unreliable features and Section 2.3.3 elaborates the feature selection procedure. The detection of ailment, correlation, and responsiveness are discussed in detail in Section 2.3.4, Section 2.3.5, and Section 2.3.6, respectively.

2.3.1 Hand Motor Features

The motor features are extracted based on the known symptoms of CSM. The symptoms of CSM that are related to hand motor function include loss of hand dexterity due to numbness and stiffness, hyperreflexia, weakness, fatigue, and

tremor [You00]. Hyperreflexia is known as a unique symptom of CSM patients, which represents exaggerated muscle reflexes such as twitching [You00, LGM13b]. A total of 21 features (10 from sinusoidal and 11 from step waveforms) are extracted in this work.

The following symptoms are extracted from sinusoidal waveform. Mean absolute error (MAE-SINE) computes the average error between the target sinusoidal waveform and the patient's response over the length of the signals: $\sum_k |w_t[k] - w_r[k]|$, where w_t and w_r represent target waveform and patient's response, respectively. MAE-SINE is one of the most common metrics in quantifying comprehensive hand dexterity [Jon00]. Hyperreflexia in CSM patients is characterized as twitching or jerk of hand muscles that result in a sudden peak of the signal at the initialization or a sudden drop at the discharge of grip force as illustrated in Fig. 2.3. The characteristics of hyperreflexia are quantified using five features: Min2ZeroErr, Min2ZeroMaxErr, Max2ZeroErr, Max2ZeroMaxErr, and PhaseShift. Min2ZeroErr computes MAE between the target and the patients' response from the local minimum to the first zero point of the target waveform (Fig. 2.3), in order to quantify the error caused by twitching. Min2ZeroMaxErr computes the maximum error between the two waveforms from the local minimum and the first zero. Max2ZeroErr and Max2ZeroMaxErr are calculated in a similar manner to quantify the sudden drop from the local maximum. Phase shift (PhaseShift) represents the average time difference between the zero points of the target and those of the patient-generated waveforms in order to measure how quickly patients recover from the overexcited muscle activities. Since a single 45 second-long test contains seven sine cycles, each of these hyperreflexia features is computed as an average over the seven cycles. Fatigue (FTG-SINE) is quantified as a change in the MAE between the very first and the very last sinusoidal cycles since a patient with relatively high fatigue rate would lose their sensory-motor control over time, and as a consequence, the error rate increases. Tremor is quan-

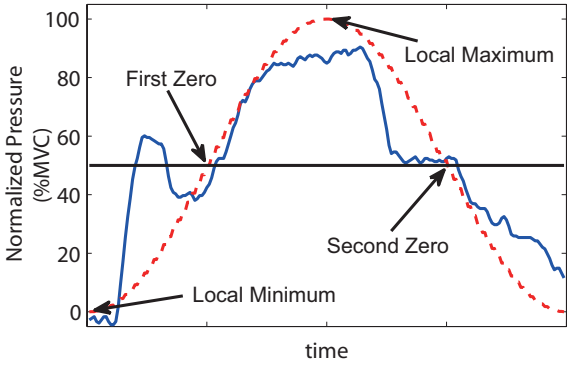

Figure 2.3: An example of hyperreflexia which is illustrated with a sudden peak at the initiation and a sudden drop at the discharge of grip force.

tified using the following three features, which are all computed in the frequency domain: 2ndFreq, ΔFreq, and ΔGain. Since the target waveform is a perfect sinusoidal waveform, it only contains a single peak in the frequency domain at its fundamental frequency. However, due to a high level of noise (e.g., twitching effect shown in Fig. 2.3 or tremor) produced by patients, there exists stronger energy in the higher frequency ranges. 2ndFreq computes the frequency with the second largest gain after the fundamental frequency of the patient generated waveform. Patients are expected to have higher second fundamental frequency compared to healthy subjects. ΔFreq computes the difference between the fundamental frequencies of the target waveform and the patient-generated waveform. ΔGain is the gain difference between the two frequencies.

A total of 12 features are extracted from step tests. MAE (MAE-STEP) is computed to represent the comprehensive motor control capacity under step waveform. Four additional features are computed to quantify the level of hyperreflexia: VEL-INC, AMP-INC, VEL-DEC, AMP-DEC. VEL-INC computes the velocity of the grip strength (%MVC per second) when the subject has to activate the strength from the floor (= 20%MVC) to the ceiling (= 80%MVC). AMP-INC is the maximum change in the amplitude at the instance of the muscle activation; VEL-INC

is computed as AMP-INC divided by the associated time. VEL-DEC and AMP-DEC are computed in a similar manner when the subject has to suddenly release the grip strength from the ceiling to the floor. Muscle weakness (or muscle endurance) is quantified using AVG-CEILING, STD-CEILING, AVG-FLOOR, and STD-FLOOR. AVG-CEILING and STD-CEILING respectively compute mean and the standard deviation of the amplitude of the patient's waveform while the subject is maintaining the strength at 80%MVC (ceiling). It is hypothesized that the value of AVG-CEILING would be low and the value of STD-CEILING would be high when a patient suffers from grip weakness. Similarly, AVG-FLOOR and AVG-CEILING computes those when the subject is maintaining the grip strength at 20%MVC (floor). Fatigue (FTG-STEP) computes the change in the MAE between the first two and the last two step cycles similarly to FTG-SINE. The hand motor features considered in this work are summarized in Table 2.1 with their mean values and standard deviations for control subjects, preoperative, and postoperative CSM patients.

2.3.2 Eliminating Unreliable Features

The proposed work employs a pre-processing that eliminates unreliable features based on test-retest criteria [PMA13]. In order to evaluate the test-retest reliability, the features of the last two (out of three) tests have been used to compute the intra-class correlation coefficient (ICC), denoted as r, according to [Ros06, GHW93]. The value of r is in the range of $[0, 1]$. Conventionally, $r < 0.4$ indicates poor, $0.4 \leq r < 0.75$ indicates fair to good, and $r \geq 0.75$ indicates excellent test-rest reliability [Ros06]. The r values are computed separately for the control, the preoperative, and the postoperative CSM groups. This work employs those features that achieve $r \geq 0.75$ for all datasets.

2.3.3 Feature Selection

The purpose of feature selection is a twofold. First, it identifies a subset of reliable features that best addresses the objective of the analysis for the purpose of minimizing the chances of overfitting. Second, it investigates the expected performance of classification/regression models, which are constructed based on the selected features.

This work utilizes a *wrapper* approach for feature selection, which evaluates all feature subsets within its feature searching space for their classification/regression performances, and selects the subset that produces the best performance [GE03]. First, the maximum cardinality of a feature subset is constrained based on the dataset-to-feature ratio as suggested in [PJF12]. For a linear classification model, the minimum dataset-to-feature ratio is limited to 10:1, and for a quadratic model, the cardinality of the selected feature set (denoted as F) should be such that $F \times (F+3)/4$ does not exceed the number of subjects of the smallest class [PJF12]. For example, the ailment detection involves data samples of 30 preoperative CSM patients and 30 controls. Thus, the maximum number of features in the selected subset should be limited by $\lfloor 60/10 \rceil = 6$ and 9 for a linear and a quadratic model, respectively. Then, this work employs a *forward selection* approach for constructing the feature searching space. The forward selection approach starts with an empty feature set and progressively add a feature that produces the best classification/regression performance until its size reaches the maximum cardinality. It is a greedy searching algorithm (i.e., it cannot retract its decision once a feature is added to the feature set), which does not guarantee providing an optimal solution but significantly reduces the computational complexity [GE03]. A leave-one-out cross validation (LOOCV) is employed to evaluate a feature subset.

The feature selection algorithm is described in detail in Fig. 2.4. The input D is an $N \times M$ matrix that represents the input dataset, where N is the number of

data samples, and M is the total number of features. y is an $N \times 1$ vector that represents the ground truth label, i.e. binary class label for ailment detection and ODI scores for correlation. F is the maximum allowable cardinality of the output feature set. The function BuildModel($\cdot$) represents an arbitrary machine learning (training) algorithm. This is the reason for the naming convention of *wrapper* that the searching is wrapped around a specific learning algorithm. The function Predict($\cdot$) computes the most likely label for the given testing data based on the model created by BuildModel($\cdot$). Evaluate($\cdot$) is a function that evaluates the classification/regression performance in a LOOCV, e.g. misclassification rate for a binary classification problem. In summary, this feature selection algorithm returns a feature subset that produces the best prediction performance given a dataset D.

Another important question is the expected classification/regression performance of the model based on the selected feature subset, when the model is indeed deployed and tested on an independent dataset. A *hold-out strategy*[2] is employed to address this important question [LGM13a, BZ08]. The hold-out strategy contains two layers of cross validation, of which are both LOOCV in this work. The outer LOOCV divides the entire dataset into a *validation set* (i.e., the left-out data) and a *learning set* (i.e., the remaining dataset). The *learning set* is then used as the input to the FeatureSelection($\cdot$) procedure (Fig. 2.4) in order to find out the best feature subset using another (inner) layer of LOOCV. The classification/regression model is constructed using the *learning set* that is projected on the selected feature subset. The performance of the constructed model is evaluated using the *validation dataset*. The expected performance is then computed as the average classification/regression accuracy evaluated on the validation datasets over the outer cross validation. It is noteworthy that a total of N feature subsets are selected as a result of the hold-out strategy (i.e., the outer LOOCV). In

[2]It is also known as a *nested cross validation* in [PMA13].

1: **procedure** FeatureSelection(D, y, F)
2: $P = [1 \cdots M]$ {initial index for the entire feature pool}
3: $S_0 = \emptyset$ {initial feature set is null}
4: **for** $i = 1$ **to** F **do**
5: **for** $j = 1$ **to** $|P|$ **do**
6: $D' = D[:, (S_{i-1} \cup P[j])]$; {adding a feature to evaluate}
 {starting LOOCV}
7: **for** $n = 1$ **to** N **do**
8: $D_{ts} = D'[n, :]$; {testing set}
9: $y_{ts} = y[n]$; {testing label}
10: $D_{tr} = D' - D_{ts}$; {training set}
11: $y_{tr} = y - y_{ts}$; {training label}
12: $\theta = \text{BuildModel}(D_{tr}, y_{tr})$; {building classification / regression model}
13: $y'[n] = \text{Predict}(D_{ts}, \theta)$; {make prediction}
14: **end for**
15: $R_P[j] = \text{Evaluate}(y, y')$; {evaluating current set}
16: **end for**
17: $S_i = S_{i-1} \cup P[\text{argmax}(R_P)]$ {adding the best feature}
18: $R_F[i] = \max(R_P)$ {performance at the cardinality i}
19: $P[\text{argmax}(R_P)] = \emptyset$ {removing the selected feature}
20: **end for**
21: **return** S_{i^*} where $i^* = \text{argmax}(R_F)$

Figure 2.4: The feature selection procedure for a given dataset D with its label y and the maximum cardinality F.

1: **procedure** HoldOutStrategy(D, y, F)
2: **for** $n = 1$ **to** N **do**
3: $D_v = D[n, :]$; {validation set}
4: $y_v = y[n]$; {validation label}
5: $D_l = D - D_v$; {learning set}
6: $y_l = y - y_v$; {learning label}
7: F = ComputeMaxCardinality($|D|$);
8: s = FeatureSelection(D_l, y_l, F);
9: θ = BuildModel($D_l[:, s], y_l$);
10: $y'[n]$ = Predict($D_v[:, s], \theta$);
11: **end for**
12: **return** Evaluate(y, y');

Figure 2.5: The hold-out strategy that computes the expected classification / regression performance accuracy.

theory, if the employed dataset is generic enough to represent the target population, the N feature subsets should be identical or share similar features to each other and to the feature subset selected when the entire dataset is inputted to FeatureSelection($\cdot$) of Fig. 2.4.

2.3.4 Detection of Ailment

In order to quantify the severity of hand motor deficits in CSM patients, the problem is formalized as a binary classification between the patient data obtained prior to the treatment and the age-matched control data. This would provide information regarding hand motor characteristics that are uniquely observed among CSM patients.

Three different classification algorithms are used as the training model (i.e., BuildModel($\cdot$) in Fig. 2.4): (i) Support Vector Machine with linear kernel (SVM),

(ii) Linear Discriminant Analysis (LDA), (iii) Quadratic Discriminant Analysis (QDA). The expected classification accuracy (Evaluate(·) in Fig. 2.5) is evaluated using the sum of the true positive rate and the true negative rate.

2.3.5 Correlation to Perceived Motor Deficits

The correlation of the handgrip function to the perceived motor deficits is formulated as a regression problem using both pre- and postoperative data obtained from 17 CSM patients (i.e., $N = 34$). ODI scores have been used to represent the perceived motor deficits. The correlation to mJOA scores was not investigated since the variation of the scores among CSM patients was not granular enough to form a regression; the minimum and the maximum scores were 15 and 18, respectively. Furthermore, ODI has a larger number of questionnaires that are closely related to motor functions required for ADL. This analysis would provide insights into the hand motor characteristics that have significant correlation to the perceived quality of life, which can be especially useful for clinicians, therapists, patients, and care-givers for various reasons including (i) applying stratified rehabilitation process, (ii) properly adjusting home and community appliances, and finally (iii) improving the quality of life [VKW11, KWK96].

The regression models tested in this work are (i) Support Vector Regression with linear kernel (SVR-L), (ii) Support Vector Regression with non-linear sigmoid kernel (SVR-R), and (iii) Multivariate Linear Regression (MLR). The expected regression accuracy (Evaluate$(\cdot)$) is evaluated using the mean error rate (MER) between the predicted value and the actual value.

2.3.6 Responsiveness

Responsiveness investigates the ability of the proposed system to detect any changes in the values of motor features depending on the patients' operation conditions. As a consequence, this problem is formulated as a binary classification problem between the patients whose motor functions have improved (i.e., functional patients) and the patients whose motor functions have deteriorated or no change (i.e., non-functional patients) as a consequence of surgery, similarly

Table 2.1: Hand motor features that are extracted based on the symptoms of CSM, and the patient reported outcomes.

Associated Symptom	Label	Waveform	Control Subjects		CSM Patients (Preop)		CSM Patients (Postop)	
			Mean (Std)	ICC, p-val	Mean (Std)	ICC, p-val	Mean (Std)	ICC, p-val
Comprehensive Measure	MAE-SINE	Sine	6.48 (0.90)	0.97, 0.49	11.65 (5.34)	0.97, 0.76	9.24 (5.39)	0.96, 1.00
	MAE-STEP	Step	7.49 (1.44)	0.94, 0.74	13.49 (9.05)	0.99, 0.97	10.24 (7.90)	0.99, 0.94
Hyperreflexia	Min2ZeroErr	Sine	6.13 (1.62)	0.86, 0.18	11.08 (4.61)	0.95, 0.62	8.60 (7.96)	0.99, 0.71
	Min2ZeroMaxErr	Sine	22.49 (6.59)	0.79, 0.44	37.06 (14.80)	0.93, 0.77	26.54 (17.70)	0.97, 0.94
	Max2ZeroErr	Sine	5.66 (1.86)	0.95, 0.92	9.74 (6.59)	0.94, 0.66	4.01 (9.82)	0.78, 0.95
	Max2ZeroMaxErr	Sine	1.83 (4.58)	1.00, 0.87	4.03 (1.60)	0.93, 0.41	1.42 (4.90)	1.00, 0.81
	PhaseShift	Sine	3.02 (1.42)	0.74, 0.50	12.61 (33.19)	0.99, 0.70	5.77 (5.31)	0.98, 0.98
	VEL-INC	Step	1.82 (0.39)	0.96, 0.95	1.95 (0.66)	0.96, 0.69	1.81 (0.29)	0.84, 0.89
	AMP-INC	Step	66.65 (2.95)	0.92, 0.87	67.37 (8.61)	0.99, 0.99	64.63 (6.14)	0.89, 0.88
	VEL-DEC	Step	-1.42 (0.21)	0.89, 0.51	-1.45 (0.47)	0.82, 0.77	-1.27 (0.16)	0.84, 0.48
	AMP-DEC	Step	-66.41 (3.11)	0.94, 0.89	-67.61 (9.36)	0.99, 0.96	-64.03 (5.77)	0.78, 0.55
Fatigue	FTG-SINE	Sine	-0.61 (1.81)	0.74, 0.89	-2.47 (6.95)	0.92, 0.93	0.87 (4.18)	0.45, 0.38
	FTG-STEP	Step	-0.22 (2.30)	0.79, 0.38	1.73 (9.83)	0.96, 0.77	-1.51 (9.18)	0.63, 0.40
Weakness	AVG-CEILING	Step	75.28 (1.56)	0.89, 0.99	69.54 (8.64)	0.99, 1.00	72.19 (8.96)	0.98, 0.92
	STD-CEILING	Step	10.85 (1.95)	0.90, 0.65	14.72 (7.63)	0.99, 0.91	12.04 (5.15)	0.99, 0.77
	AVG-FLOOR	Step	12.66 (1.85)	0.89, 0.83	29.84 (7.36)	0.98, 0.76	26.55 (4.30)	0.82, 0.28
	STD-FLOOR	Step	12.66 (1.85)	0.89, 0.72	17.04 (6.52)	0.99, 0.85	13.60 (4.10)	0.94, 0.82
Tremor	2ndFreq	Sine	0.33 (0.11)	0.80, 0.96	0.43 (0.16)	0.96, 0.79	0.39 (0.17)	0.93, 0.78
	ΔFreq	Sine	0.00 (0.00)	1.00, 1.00	0.00 (0.02)	0.99, 1.00	0.00 (0.00)	1.00, 1.00
	ΔGain	Sine	5.74 (4.31)	0.94, 0.42	7.63 (6.35)	0.98, 1.00	5.82 (5.85)	0.91, 0.91
Patient Reported Functional Outcome	ODI	-	-	-	0.65 (0.22)	-	0.76 (0.17)	-
	mJOA	-	-	-	0.85 (0.11)	-	0.95 (0.06)	-

to [DC86]. The functional patients are defined as those whose postoperative ODI and mJOA values have both improved compared to their preoperative values. In our study, 12 patients were categorized as functional and 5 patients as non-functional; this determines the input y for Fig. 2.4 and Fig. 2.5. Similarly, the difference between postoperative and preoperative values were computed for all motor features to be used as the input data: D for HoldOutStrategy($\cdot$). Three different classification algorithms were again used as the training model: SVM, LDA, and QDA. The accuracy evaluation function is also defined as the one in Section 2.3.4.

2.4 Results

2.4.1 Reliability of Features

Table 2.1 summarizes the ICC for all features, where unreliable features are indicated with shading. This table further provides p-values of analysis of variance (ANOVA) for intra-variation along with the r values.

2.4.2 Detection of Ailment

Table 2.2 summarizes the results of the feature selection that returns (i) the selected feature subset and (ii) the expected accuracy. QDA outperformed other classifiers in terms of its expected classification performance. The selected feature subset for QDA (FeatureSelection($\cdot$)) is composed of MAE-STEP, Max2ZeroErr, and 2ND-Freq.

The expected classification accuracy was as follows: true positive rate (TPR) of 0.737, true negative rate (TNR) of 0.842, and average classification accuracy (ACR) of 0.789. The selected feature subsets and their frequencies during the HoldOutStrategy($\cdot$) are summarized in Table 2.2. It is noteworthy that the most

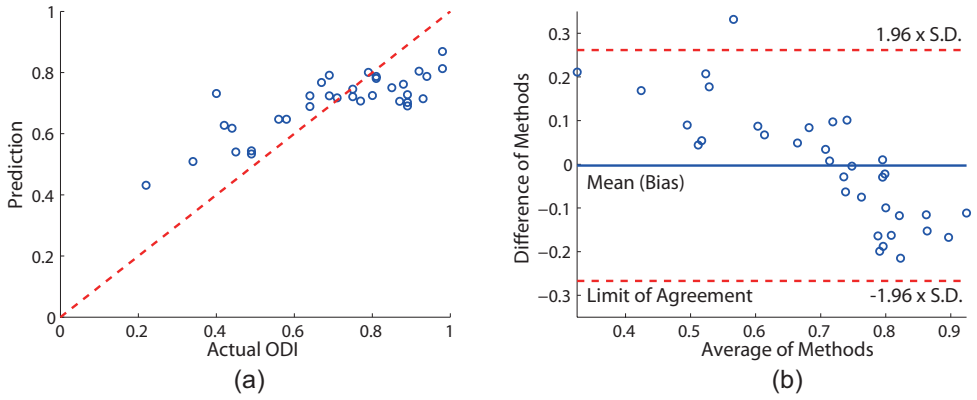

Figure 2.6: (a) The regression plot produced based on the hold-out strategy, and (b) its Bland-Altman plot.

frequently selected feature set during the hold-out strategy is identical to the selected feature set of the entire dataset.

2.4.3 Correlation to Perceived Motor Deficits

SVR with sigmoid kernel produced the best expected regression accuracy as it is summarized in Table 2.3. The selected feature subset for this non-linear SVR on the entire dataset contains Min2ZeroErr, ΔFreq, and VEL-INC.

The expected regression accuracy based on the hold-out strategy produced MER = 0.110 and $R^2 = 0.555$. The regression outcomes and the Bland-Altman plot are illustrated in Fig. 2.6. The selected feature subsets from the hold-out strategy are summarized in Table 2.3 (bottom). The most frequently selected feature subset on the outer layer of the LOOCV was identical to the feature subset selected from the entire dataset: <Min2ZeroErr, ΔFreq, VEL-INC>.

2.4.4 Responsiveness

The analytic results for responsiveness are summarized in Table 2.4. Liner SVM produced the best classification accuracy. The selected features from Fig. 2.4 are

Table 2.2: Expected classification performance and the best feature set for ailment detection (top), and the selected feature sets from the outer cross validation of the hold-out strategy (bottom).

	TPR	TNR	ACR	Selected Features
SVM	0.842	0.632	0.763	MAE-STEP, AVG-FLOOR, VEL-INC, PhaseShift
LDA	0.842	0.632	0.763	MAE-STEP, STD-FLOOR, AMP-INC
QDA	**0.737**	**0.842**	**0.789**	MAE-STEP, 2ND-Freq Max2ZeroErr

Selected Features	Freq
MAE-STEP, 2ND-Freq, Max2ZeroErr	18
MAE-STEP, 2ND-Freq, Max2ZeroErr, Min2ZeroErr	13
MAE-STEP, Max2ZeroErr, Max2ZeroMaxErr, STD-FLOOR	1
MAE-STEP, 2ND-Freq, AVG-CEIL, AVG-FLOOR	1
Max2ZeroMaxErr, STD-FLOOR, Min2ZeroErr, VEL-DEC	1
2ND-Freq, Max2ZeroErr, STD-FLOOR, VEL-DEC	2
Max2ZeroMaxErr, Min2ZeroErr, AMP-INC, VEL-DEC	2
Total	38

Table 2.3: Expected regression performance and the best feature set (top), and the selected feature sets from the outer cross validation of the hold-out strategy with Sigmoid SVR (bottom).

	MER	R^2	Selected Features
MLR	0.139	0.153	Max2ZeroErr ΔFreq, AVG-FLOOR, VEL-INC
SVR - Linear	0.133	0.302	MAE-SINE, 2ndFreq ΔFreq, VEL-INC
SVR - Sigmoid	**0.110**	**0.555**	Min2ZeroErr, ΔFreq VEL-INC

Selected Features	Freq
Min2ZeroErr, ΔFreq, VEL-INC	16
Max2ZeroMaxErr, VEL-INC, VEL-DEC	8
Max2ZeroMaxErr, ΔFreq, VEL-INC, VEL-DEC	5
Max2ZeroMaxErr, ΔFreq, 2NDFreq	2
Max2ZeroMaxErr, ΔFreq, 2NDFreq, VEL-DEC	1
Min2ZeroErr, Max2ZeroMaxErr, VEL-INC, VEL-DEC	1
Min2ZeroErr, Max2ZeroErr, VEL-INC, VEL-DEC	1
Total	34

Table 2.4: Expected classification performance and best feature set for responsiveness (top), and the selected feature sets from the outer cross validation of the hold-out strategy with SVM (bottom).

	TPR	TNR	ACR	Selected Features
SVM	**0.917**	**0.800**	**0.882**	Max2ZeroErr, AVG-FLOOR
LDA	0.917	0.600	0.824	Max2ZeroErr, STD-FLOOR
QDA	0.917	0.600	0.824	Max2ZeroErr, AVG-FLOOR

Selected Features	Freq
Max2ZeroErr, AVG-FLOOR	15
AVG-FLOOR, STD-FLOOR	1
Min2ZeroErr, Max2ZeroErr	1
Total	17

Max2ZeroErr and STD-FLOOR.

The expected accuracy is computed as TPR = 0.917, TNR = 0.800, and ACR = 0.882. The selected feature sets from the HoldOutStrategy(·) are summarized in Table 2.4. The most frequently selected feature set from the hold-out strategy was again identical to the selected feature of the entire dataset.

2.5 Discussion

2.5.1 Feature Reliability

As shows in Table 2.1, most features showed excellent reliability. Furthermore, none of the features showed statistical significance for intra-class variability. There exist three features with ICC less than 0.75 (i.e., PhaseShift, FTG-SINE, and FTG-STEP), which are removed from all the analyses.

2.5.2 Analytic Results

The feature set selected by FeatureSelection(·) for the ailment detection contains MAE-STEP, 2ndFreq, and Max2ZeroErr. Among the selected features, MAE-STEP is the most discriminant feature. The p-value of MAE-STEP ($p < 0.0071$) was significantly lower than 2ndFreq ($p < 0.041$) and Max2ZeroErr ($p < 0.013$). Furthermore, MAE-STEP was the only feature that was selected by all three tested classifiers (Table 2.1). The feature subset selected by FeaturesSelection(·) was identical to the feature subset that was most frequently selected during the hold-out strategy. Moreover, the most frequently selected subset was a close subset of the second most frequently selected subset as shown in Table 2.2. These two subsets were selected $31/38 = 81.6\%$ during the outer cross validation of the hold-out strategy, which indicates that <MAE-STEP, 2ndFreq, Max2ZeroErr> is the subset that well defines the motor characteristics of our CSM patients.

Table 2.3 shows that the non-linear regression model (sigmoid SVR) performs significantly better compared to the two linear models (MLR and linear SVR). This demonstrates that there exist non-linear relationships between the predictors and the ODI score. The selected features for sigmoid SVR are Min2ZeroErr, ΔFreq, and VEL-INC. It is noteworthy that all of these features belong to hyperreflexia category, which quantifies how accurately patients control their fine hand muscles at sub-maximal forces. Furthermore, most of the features selected throughout the hold-out strategy also belong to the hyperreflexia category (Table 2.3). This finding addresses similar conclusions made in [KGG05, MI94], and [SN00] that precise control of the sub-maximal grip strength is the most important motor function required in daily activities (rather than maximum voluntary contraction). Given its highly subjective nature of ODI [Fin10] and the fact that the regression results are generated based on two layers of cross validation, the results in Fig. 2.6 and Table 2.3 are comparable to other studies. ODI has been compared to other measures: Pearson's $r = 0.83$ for Pain Disability In-

dex [GHW93], $r = 0.62$ for Visual Analogue Scale [GHW93], $r = 0.62$ for McGill Pain Questionnaire [HN92], $r = 0.77$ for Short Form-36 [GKW97], and $r = 0.66$ for Roland-Morris (RM) questionnaire [LBF97]. The correlation coefficient of the regression was $r = 0.81$. Especially in [FP00], the Bland-Altman plot between ODI and RM shows that the magnitude of the limit of agreement is approximately equal to 0.32, which is comparable to this work's result: 0.26. Fig. 2.6 (a) and (b) show that the prediction results are overestimated when the ground truth ODI value is less than 0.7, and underestimated when greater than 0.7. It is because most of the ODI scores of the patients were established between 0.55 and 0.85, and as a consequence, the predicted results are forced to fit within the range; the mean and the std. dev. of the ODI scores of both pre- and postoperative CSM patients were 0.70 and 0.20, respectively.

The selected features that best discriminate functional from nonfunctional patients after decompression surgery are Max2ZeroErr and AVG-FLOOR. The improvement in Max2ZeroErr scores for the functional and the nonfunctional groups were -4.41 and 5.24 ($p < 0.0037$), respectively. This implies that the Max2ZeroErr for the functional group has been significantly reduced compared to the nonfunctional group. In a similar manner, the improvement in AVG-FLOOR was -3.03 for the functional and -0.77 for the nonfunctional group ($p < 0.080$). This feature set was also the most frequently selected feature subset from the HoldOutStrategy($\cdot$) with $15/17 = 88.2\%$. Other two features also share similar motor functions as shown in Table 2.4. It is especially noteworthy that Max2ZeroErr was selected for all three analyses performed in this chapter. This result highlights the medical finding that hyperreflexia is a common symptom of CSM patients [You00].

2.5.3 Limitations and Future Works

All the classification and regression performances reported in this chapter are computed using the hold-out strategy, which produces more fair estimate rather than optimistic estimate [PMA13]. However, the small sample size makes it difficult to generalize our findings to the general CSM population. This was highlighted in the correlation analysis that we need CSM patients with more diversified motor functions in order to improve the (underestimated and overestimated) regression results (Fig. 2.6). Nonetheless, the selected feature subsets of all analyses (detection, correlation, and responsiveness) were identical to the most frequently selected feature subsets of the hold-out strategy. Furthermore, the selected subsets from the outer cross validation of the hold-out strategy shared similar features, which demonstrates that there exists a consistent motor pattern among the tested dataset.

The classification performances reported in this chapter (i.e., ACR of 78.9% for ailment detection and 88.2% for responsiveness) would not be ideal to be used as a diagnostic tool for CSM. This is because the proposed device quantifies hand motor function of patients rather than analyzing anatomic or neurologic markers. However, the short-duration tracking test (4 minutes and 30 seconds) can be effectively used as an assistive screening/monitoring tool, since CSM is closely related to the aging process, and is known to be genetically inherited [You00, Kli10].

Due to its simplicity of use, short testing time, and inexpensive cost, this device has a great potential for remote patient monitoring, which has been discussed in a previous study [LWN12]. For this purpose, the wireless data communication would be necessary to enhance the user interface. Furthermore, gamification of the system to improve patient adherence to the system (for the purpose of frequent monitoring) would be an interesting application, similarly to [FCZ11]. Future

works should also incorporate stratified analysis based on clinical variables such as the location of the herniated disk, the months after injury, or the values of ODI. It is expected that these would produce more sophisticated detection model based on features that are uniquely observed among the tested group.

2.6 Conclusion

This chapter introduced a quantification measurement of hand motor function in CSM patients based on target tracking tests using a handgrip device. Three different quantification objectives were addressed using a reliable feature selection technique: detection of ailment, correlation to the perceived motor deficit, and the responsiveness. The hold-out strategy is employed to compute the expected classification/regression performance, which achieved ACR = 78.9%, MER = 11.0%, and ACR = 88.2% for detection, correlation, and responsiveness, respectively. This pilot study validates the validity and reliability of the handgrip device for quantifying hand motor functions in CSM patients. We believe that this study enables new research opportunities in analyzing various types of motor characteristics in spondylotic myelopathy, which is a growing chronic condition.

CHAPTER 3

Assessment of Overexcited Hand Movements

3.1 Objectives

Patients who suffer from neuro-degenerative diseases (e.g., stroke and Parkinson's disease) or degenerative spinal cord disorders often carry movement deficits in upper extremities [Mor00b, LWN12]. Among many motor symptoms associated with these ailments, we are particularly interested in hyperexcitability in hand muscles, which is defined as a motor disorder characterized by exaggerated tendon jerk reflexes [FYK80] due to an excessive velocity increase in muscle tone [LHC02b]. Handgrip hyperexcitability creates involuntary forces during grasping performance, which intensely restrict daily activities requiring sophisticated hand muscle manipulation such as eating, clothing, and bathing.

Traditional assessment methodology for hyperexcitability relied on subjective observations of muscle behavior, and as a result, many attempts have been made to objectively quantify the level of hyperexcitability. Existing solutions to quantify hyperexcitability of muscle movements have concentrated on techniques such as clinical scales, Electromyographic (EMG), and biomechanics. However, these techniques are often highly complicated to be deployed at clinical (or in-home) settings, large in size, and extremely expensive. As a consequence, it was not economically feasible to deploy these techniques for a large patient population, and this creates a need for an accurate and affordable assessment system [KZB04].

Sensing platforms that can be easily deployed on the body have been actively

researched and are considered as alternative approaches to diagnose, to quantify, and to rehabilitate patients with motor deficits such as in [HRS12]. Body sensing systems utilize accurate, simple, and inexpensive sensors to collect physiological data in order to quantify motor performance [LGa13a, LGa13b]. These characteristics allow (i) easy ways to collect sensory data either pervasively or from a simple motor task, (ii) economic deployment of the system for a large patient population, and (iii) improvement in clinical benefits for patients. Clinical benefits of body sensing systems for assessing motor abnormalities include (i) economic benefits [HBe05], (ii) frequent and continuous measurement of motor function progress over time, (iii) quantifying the effectiveness of medical treatments, such as surgical operations or medications, and (iv) early diagnosis of motor function for potential patients.

In this chapter, a low-cost system that objectively quantifies the level of hyperexcitability in hand dexterity is introduced. A term *activation hypertonia* is used to describe the hyperexcitability during voluntary grip contraction (details are provided in Section 3.4). The proposed system utilizes a lightweight handgrip sensory device to assess the level of activation hypertonia, which makes the system highly portable. The system provides a simple target tracking task to examine fine hand motor skills for patients with cervical spinal cord disorders [Jon00]. The collected body signals are then analyzed by a series of four signal processing units: (i) the pre-processing unit, (ii) the abnormality (i.e. activation hypertonia) detection unit, (iii) the abnormality analytic unit, and (iv) the parameter extraction unit. The preprocessing unit performs a low-pass filter to reduce noise in the raw signals, and segments the signals into a number of subsignals. The detection unit statistically determines whether a resultant subsignal contains the outcome of the exaggerated muscle tone using machine learning algorithms. If activation hypertonia is noted, the analytic unit performs an in-depth analysis to locate important geometric points using dynamic time warping (DTW). The parameter extraction

unit extracts important variables that characterize the severity of activation hypertonia. The system has been clinically tried in cohort study under collaboration with the UCLA Department of Neurosurgery in order to evaluate its performance.

The rest of the chapter is organized as follows. Related works are discussed in Section 3.2. Section 3.3 provides the overview of the proposed system in detail. In Section 3.4, the definition of activation hypertonia and its physiological characteristics are discussed. Section 3.5 elaborates the signal processing techniques that extract important information about activation hypertonia. The experimental results are presented in 3.6, followed by concluding remarks in Section 3.7.

3.2 Related Works

Exaggerated muscle movements such as muscle overactivity and spasticity in upper limbs have been actively researched. Existing methods to quantify the level of hyperexcitability in hand muscle usually involve passive resistance (external torque) against the applied patient-generated force. In [LHC02a] and [LBK99], torque based devices are used to quantify spastic movements in elbow flexors. Especially in [LHC02a], the authors introduced a new measurement metric based on a second order linear model of the spastic velocity in order quantify the viscous component of hypertonia. Some works combined torque-based resistance devices with EMG in order to analyze the changes in muscle tone (i.e. electric signal generated by muscles) during the spastic movement [PMD00, dPP03].

Although the aforementioned works focus on measuring exaggerated muscle performance, the fundamental research objective is different from what has been discussed in this work. That is, the aforementioned works focus on muscle excitability during passive motor functions, where as the proposed system examines the exaggerated muscle behavior during voluntary grip contraction. Moreover, the aforementioned works involve apparatuses that are extremely large, expensive, and

complicated to use in an in home setting, which makes the systems unsuitable for portability and which would not be scalable to a large patient population.

A recent study reports that very few measurement systems exist to quantify the level of spastic muscle during functional activity (i.e. during voluntary grip contraction or relaxation) [Azz12]. In [Azz12], the author uses the maximal grip strength in order to measure the level of spasticity, although the grip strength reflects the comprehensive motor performance of patients. Similarly, in [NH05], a case of spasticity during voluntary grip contraction is reported, and the overall grip strength has been used to generically represent the overall motor function.

Clinical scales are also frequently used to assess fine motor performance (including hyperexcitability), which are often constructed based on patient-reported surveys or observation of simple muscle performance. For example, the Modified Ashworth Scale (MAS) is the most common measurement for muscle spasticity in clinics nowadays [BNH12, BS87]. Furthermore, functional measures of hand performance such as Wolf Motor Functional Test (WMFT) have been automated to generically assess the spasticity [BNH12, HRS12, WPM10].

However, the above methods often rely on subjective measurements (e.g., clinical scales) or on quantitative methods that represent comprehensive hand muscle movement (e.g., grip strength). On the other hand, the proposed method quantifies the degree of hyperexcitability in hand muscles based on physiological motor function observed during voluntary hand contraction. As a result, parameters from various dimensions of motor functions (i.e. grip force, time, and velocity) can be accurately analyzed. This work is an extension of an abstract [LGa13b] in which the medical significance of hyperexcitability during voluntary contraction is highlighted. This chapter focuses on the signal processing techniques of the acquired body signals in order to extract information reflecting the overexcited muscle behavior.

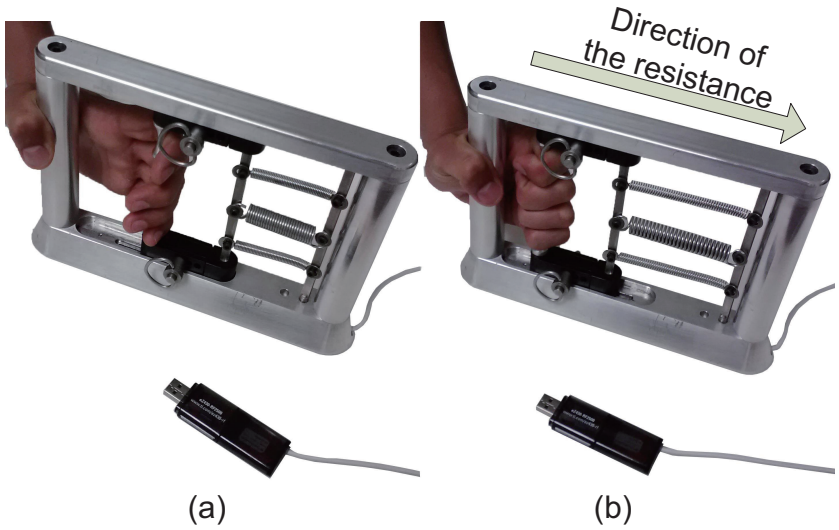

Figure 3.1: (a) The digital handgrip device used in its resting position. (b) The handgrip when the handle is moved to its maximum displacement. The direction of the spring resistance is illustrated in arrows.

3.3 System Overview

The proposed system consists of the digital handgrip device, the software system, and the signal analytic framework. The handgrip device collects sensory data from the participating patients and delivers it to the software system. The functional objective of the software system is to guide the patient to follow the examination procedure, to provide visualized feedback, and finally to store the captured data. The captured data is then processed by the signal analytic framework in order to extract information related to activation hypertonia in hand dexterity. The digital handgrip device and the software system are discussed in detail in the following two subsections.

3.3.1 Sensory Device

The handgrip device is illustrated in Fig. 3.1, which is composed of three major components: the handle, the springs, and the displacement sensor.

The handle is connected to the main body of the device by three springs, which allow the patients to make voluntary grasping performance. The three replaceable springs with known constants create resistance against the direction of the grip force as shown in Fig. 3.1 (b). Furthermore, these springs allow the system to be calibrated to individuals with different ranges of maximum voluntary contraction (MVC)[1]. The handle is also connected to the displacement sensor embedded in the bottom of the main body to locate the position of the handle. The spring constants and the displacement information can be combined to measure the grip strength in standard units such as *Newton* using Hooke's law: $F = -k \cdot x$, where k is the spring constant and x is the displacement.

3.3.2 Software Framework

The software starts the examination process by performing another calibration that measures the MVC. The reason behind this additional calibration is to perform the examination that is maximally accommodated to individuals with various levels of MVC since changing springs may not provide sufficient granularity.

Upon completion of the calibration process, subjects are tasked to track a moving sinusoidal waveform by adjusting their grip strength. Fig. 3.2 illustrates an examination provided by the software. The red sinusoidal waveform is the target waveform that moves to the left at a constant speed. The maximum amplitude of the waveform is equal to the subject's MVC as a result of the calibration process. The blue circle located in the middle of the x-axis moves freely in y-axis according to the grip strength applied to the sensory device. The green waveform appearing in the left half of the screen is real-time feedback of the subject's past performance. The examination is 45 seconds long and it contains seven sinusoidal cycles (i.e. the frequency of the sinusoidal waveform is approximately 15.6Hz), and the data is stored for post-processing.

[1]MVC is defined as the maximum voluntary grip force of an individual

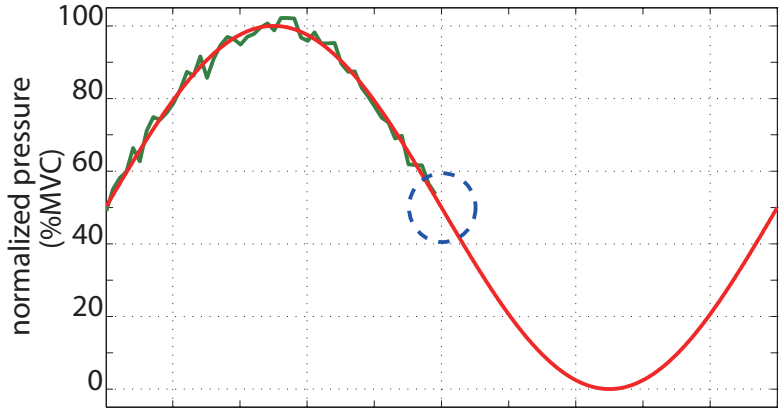

Figure 3.2: Exemplary illustration of the target tracking task used in the proposed system.

3.4 Background

An example of an instance of activation hypertonia and a normal performance is provided in Fig. 3.3 (a) and (b), respectively. As illustrated, the example in Fig. 3.3 (a) displays exaggerated muscle movement during voluntary initiation of the muscle contraction, and the example in Fig. 3.3 (b) shows a smooth curve without any notable shooting effect. This shooting effect is only observed among patients with hand movement deficits[2] (e.g., patients with cervical spondylotic myelopathy in this study). The term *activation hypertonia* is used to describe this physiological phenomenon, which is voluntarily initiated but not effectively controlled.

The activation hypertonia shows motor mechanism similar to hyperexcitability of the stretch reflexes in elbow or knee [LHC02b]. The exaggerated muscle movement is induced by sufficiently fast contraction velocity. Then, reactive muscle response starts to decrease the contraction velocity. For example, in Fig. 3.3 (a),

[2]The shooting effect has been sometimes observed among healthy subjects with age greater than 80. However, the amplitude of the shooting was minor and not comparable to the ones generated by patients.

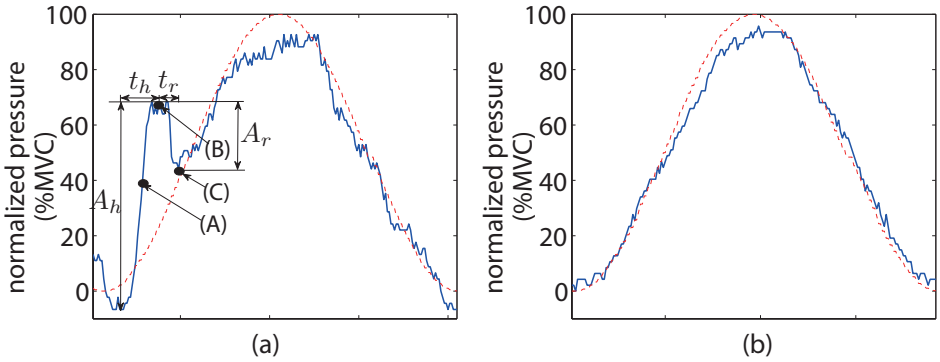

Figure 3.3: (a) An example of the exaggerated muscle movement during voluntary contraction. (b) An example of a typical normal muscle behavior.

annotation (A) represents a point where the contraction velocity is at its maximum, and this implies that the reactive muscle response is initiated to decrease the contraction velocity. At (B), the reactive response dominates the muscle movement and the spike starts to decrease. Finally at (C), the motor performance is adjusted to the target waveform.

Important parameters that characterize the exaggerated muscle behavior can be computed if geometric annotations (B) and (C) are accurately located. Thus, the proposed signal processing focuses on detecting the appearance of such an abnormal hand movement and accurately locating these important geometric annotations. The signal processing framework is composed of a series of four sub-units: (i) the pre-processing unit, (ii) the abnormality detection unit, (iii) the in-depth analytic unit that annotates the important geometric points, and (iv) the parameter extraction unit. A graphical summary of the signal processing framework is provided in Fig. 3.4.

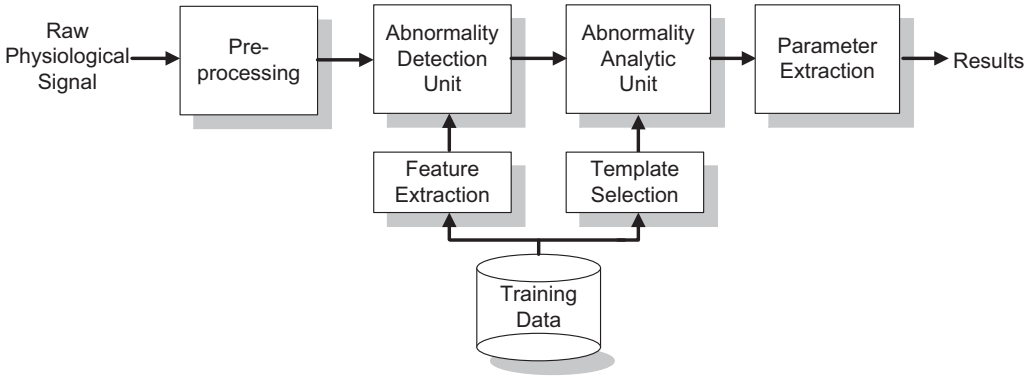

Figure 3.4: Graphical overview of the signal processing framework.

3.5 Quantification of Activation Hypertonia

3.5.1 Pre-processing Unit

The pre-processing unit performs a fifth-order butterworth low-pass filter in order to smooth out the signal. This process allows the geometric shape of the signal to be visualized more clearly and improves the accuracy of annotating important geometric points within the signal. The cut-off frequency is set to 18Hz given that the sample frequency is 32Hz and the frequency of the sinusoidal waveform is 15.6Hz. Then, the filtered time-series signal, which contains seven sinusoidal cycles, is partitioned into seven subsignals that represent muscle contraction (i.e. rising parts of the sinusoidal waveform where its derivative is greater than zero). An example of the results of the pre-processing unit is illustrated in Fig. 3.5 in order to help visualization.

3.5.2 Abnormality Detection Unit

The detection unit utilizes a machine learning algorithm to detect the appearance of the exaggerated muscle performance. For example, in Fig. 3.5, only the third to sixth segments contain instances of activation hypertonia, and thus the analysis should be performed limited to these segments.

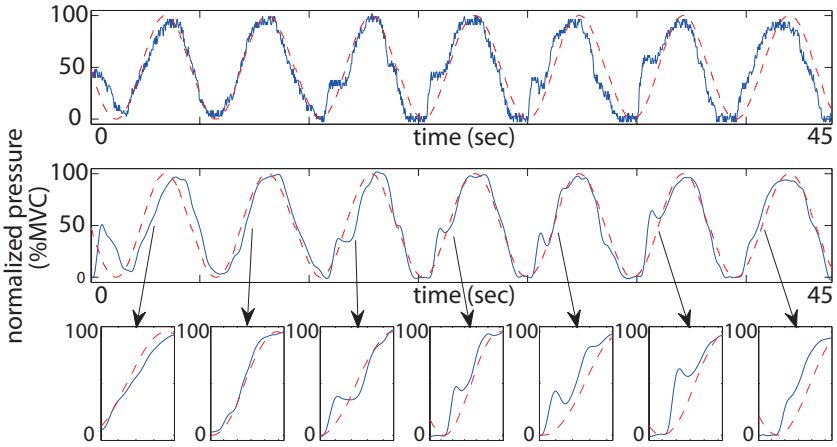

Figure 3.5: An example of the results of the pre-processing unit. (Top) The raw input time-series. (Middle) The low-pass filtered time series. (Bottom) The partitioned subsignals representing muscle contraction.

A total of six features are extracted to be used in the classification process. These features represent variables that generically differentiate the geometric shape of the signals with and without the exaggerated muscle behavior. The first feature is the maximum velocity of the signal (i.e. maximum amplitude of the derivative). The second feature computes the difference between the maximum and the minimum velocity. The third feature considers the number of local maxima found in the patient's response. The fourth and fifth features compute the maximum amplitude and the relative location (to the length of the segment) of the local maxima, respectively. The sixth feature computes the mean absolute difference between the patient's response and the target sine waveform since the segments containing exaggerated muscle performance usually have higher error rate. The proposed system employs a binary Support Vector Machine (SVM) for the classification algorithm. The classification of a new signal detects whether the signal contains the exaggerated muscle performance or not.

3.5.3 Abnormality Analytic Unit

The abnormality analytic unit employs a dynamic time warping (DTW) algorithm in order to extract the information about the geometric annotations discussed in Section 3.4 (e.g., (A), (B), and (C) in Fig 3.3 (a)). The subsignal containing an instance of activation hypertonia is compared against a template that best represents the exaggerated muscle performance in the training data[3]. A template selection technique, which is similar to the work introduced in [GJ10] and [H 10], is performed to choose a representative template.

The template selection technique is constructed as follows. The output template is denoted as T. Further, D represents a subset of the subsignals within the training data that contain instances of the activation hypertonia. A single subsignal in D is represented as τ_i, where $1 \leq i \leq |D|$. Then, the template selection starts its process by constructing a matrix M that represents the unnormalized distance between all segments in D:

$$M_{i,j} = dist(\tau_i^{'}, \tau_j^{'}),$$

where $\tau^{'}$ is the derivative of the subsignal τ, and $dist(\tau_i^{'}, \tau_j^{'})$ is the unnormalized distance between the warped subsignals $\tau_i^{'}$ and $\tau_j^{'}$. The DTW is performed on the derivative of a signal because hyperexcitability of contraction is known to depend on velocity, and the geometric annotations are also highly relevant to the derivative. Note that M is a symmetric matrix ($M_{i,j} = M_{j,i}$) and its diagonals are equal to zero since the distance between two identical signals is zero ($M_{i,i} = 0$).

Then, the template T is selected to be the subsignal that has the minimum average distance to all other signals:

$$T = \underset{i \in D}{\operatorname{argmin}} \frac{1}{|D| - 1} \sum_{j \neq i \in D} M_{i,j}, \tag{3.1}$$

where $|D|$ is the total number of subsignals in D.

[3]Note that this training data is identical to the training data discussed in Section 3.5.2

Given that the training signals are all annotated for their important geometric points, the result of the DTW between the template and the testing subsignal can easily locate these points by warping. Note that all signals are height-normalized (to the one with shorter height) before the DTW.

3.5.4 Parameter Extraction Unit

The parameter extraction unit computes the parameters that characterize activation hypertonia as discussed in Section 3.4. A total of six parameters are extracted using the geometric annotations computed as a result of the abnormality analytic unit.

The six parameters are labeled as A_h, t_h, v_h, A_r, t_r, and v_r. A_h and t_h represent the maximum amplitude and the time required to reach the peak of the hyperexcitability, respectively. Since the hyperexcitability of contraction or stretch reflexes are known to depend on velocity, the velocity to reach the peak is also an interesting parameter ($v_h = A_h/t_h$). Similarly, A_r, t_r, and v_r represent the reactive response amplitude, the time required to reach the local minimum, and the associated velocity ($v_r = A_r/t_r$), respectively. A graphical example of these parameters is provided in Fig. 3.3 (a).

3.6 Experimental Results

Table 3.1: Classification results of the leave-one-patient-out cross validation.

Patient ID	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	Average
TP Rate	0.98	0.97	0.97	1.00	1.00	1.00	1.00	1.00	0.98	0.99
TN Rate	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

3.6.1 Clinical Cohort Study

The examination procedure has been approved by the local institutional review board. The trial has been conducted for 12 months on 9 patients (mean age of 58.2 with standard deviation of 13.5) with Cervical Spondylotic Myelopathy (CSM). CSM compresses the spinal cord in the cervical area and causes hand movement deficits such as loss of hand dexterity, weakness, and coordination problems. All participated patients received a surgical spinal decompression, which decompresses the pressure applied on the pinched nerves.

Patients have participated in the study at least once prior to the operation and at one week, one month, and three months following the operation. At each clinical visit, patients performed the tracking examination exactly three times, which resulted in a total of 141 examinations. As discussed in 3.5.1, each tracking result produces seven subsignals. However, the very first subsignal for all examination results have been discarded from the analysis since some patients started the examination while holding the blue circle in the middle and some just left the circle at zero, which produced unnecessary diversity in its geometric shape. As a result, a total of $846 = 141 \cdot 6$ subsignals were considered in this analysis. Among 846 signal subsignals, $|D| = 186$ ($\approx 22.0\%$) signals showed the exaggerated muscle movement, and they all have been annotated for the important geometric points.

3.6.2 Results of the Analytic Framework

This section presents the experimental results of the technique discussed in Section 3.5. A leave-one-patient-out cross validation is used to evaluate the performance of the proposed method without polluting the results in (i) detecting the appearance of activation hypertonia and (ii) locating the important geometric annotations.

The classification results for detecting the appearance of activation hypertonia are summarized in Table 3.1. In this table, *TP Rate* and *TN Rate* represent true

positive and true negative rate, respectively, and the positive class is defined as the signals that contains the exaggerated muscle behavior and the negative class as the signals with normal muscle behavior.

When the system detects the appearance of the exaggerated muscle behavior in the signal, it performs the DTW using a template computed by Eq. (3.1). It is interesting to note that a signal belonging to P_6 is selected as a template for all cross-validations except for P_6. When P_6 was examined as the left-out patient, a signal that belongs to P_1 was selected as its template. This shows the robustness of the results of the proposed system in terms of its consistency in its geometric shape that the same template was chosen for all data (except one case for the left-out patient P_6). The DTW annotated the important geometric points with 99.5% (= 185/186) accuracy for all results. Only one signal has been mis-annotated for its unusual geometric shape. Some of the correctly annotated as well as the one incorrectly annotated result are illustrated in Fig. 3.6.

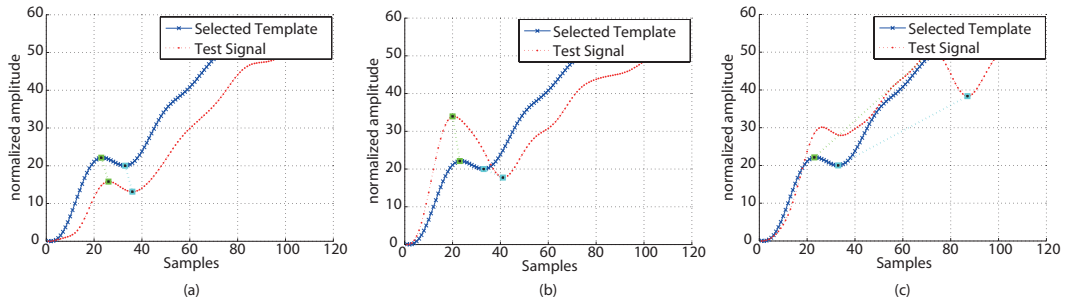

Figure 3.6: Examples of the results of the analytic framework. (a) Correctly annotated result of the test signal that is has similar dynamic geometric shape as the template (b) Correctly annotated result of the test signal that has more dynamic placement of the annotation compared to the template. (c) Incorrectly annotated signal due to an additional peak

3.6.3 Significance of the Analysis

In the previous section, we demonstrated the accuracy of the method to detect the appearance of activation hypertonia and to annotate important geometric points. In this section, the medical significance of the computed results is validated by comparing them to the physical condition of the patients.

First of all, the six parameters discussed in Section 3.5.4 are computed for all subsignals that contain an instance of the exaggerated muscle behavior. Then, the parameter values are averaged among examination results produced per clinical visit. For example, three examinations are performed per clinical visit, and six subsignals are produced per examination. Thus, each parameter can be computed by averaging the values of at most 18 subsignals containing the exaggerated muscle performance. Addition to the six parameters, one more parameter, which counts the number of subsignals that contain the exaggerated muscle behavior, is considered.

Then, parameter values obtained before the surgical operation and the parameter values collected three-month or later after the operation are compared in terms of percentage of improvement. Suppose that ρ_i^{pre} and ρ_i^{post} represent the preoperative and postoperative values of the parameter i ($1 \leq i \leq 7$), respectively. The percentage of improvement is computed as $PI_i = \left(\rho_i^{post} - \rho_i^{pre}\right) / \rho_i^{pre}$.

Patient's also have evaluated their improvement in motor function using Oswestry Disability Index (ODI), a validated motor functional scale [FP00]. According to [STH95], the patients with ODI ≥ 0.6 are considered to in the *functional* group, and the patients with ODI < 0.6 to be in the *non-functional* group. The functional group is defined as those subjects whose comprehensive motor performance is close to that of healthy subjects, and the non-functional group is defined as whose motor performance is relatively disabled. All patients were categorized as non-functional patients prior to the surgical operation. After the operation, seven

Table 3.2: A summary of the relationship between the reactive response velocity and the patient reported postoperative ODI.

	Functional							Non-Functional	
Patient ID	P1	P2	P3	P4	P5	P6	P7	P8	P9
ODI	0.81	0.84	0.75	0.52	0.84	0.84	0.89	0.36	0.31
P.I of v_r	0.31	0.26	0.39	0.49	0.54	0.17	0.78	-0.27	-0.33

out of the nine patients showed improvement in their motor functions and were categorized as functional patients according to their ODI. Furthermore, qualitative interviews also supported that the surgical operation was successful for those seven functional patients and thus improved their motor functions.

The changes in values of the seven parameter computed by the proposed analytic framework before and after the operation have been compared against the postoperative patient's motor condition (i.e. ODI). Out of the seven parameters, v_r, which represents the muscle response velocity to recover from the spastic movement, showed the strongest correlation to the postoperative ODI values (p-value < 0.029) as summarized in Table 3.2. Intuitively, this result shows that the reactive velocity against the spastic movement has been increased for the patients whose surgical operation has successfully improved their overall muscle performance, and the velocity has been decreased for the patients whose surgical operation was not successful. This is strong evidence that the proposed system (i.e. the handgrip device and the analytic framework) can successfully quantify the functional improvement or degradation as a result of medical treatment.

3.7 Future Work and Conclusions

This chapter introduces a highly portable system that accurately quantifies the level of overexcited hand movement during voluntary hand contraction. The sys-

tem utilizes a lightweight sensing platform and a signal processing framework composed of a series of four sub-processing units. A clinical cohort study has been conducted to validate the system on nine patients with degenerative cervical spinal cord disorders who have hand movement deficits, and the effectiveness of the system has been validated through an in-depth analysis. The results show that the proposed system can be useful for quantifying the level of activation hypertonia and measuring the functional progress at a low cost. Further, frequent and continuously tracking of motor performance over time may be used in the clinic or home settings to assess the need for clinical intervention or to predict the surgical success.

There exist many potential research directions to be pursued in the future. For example, analyzing voluntary reflexion in addition to contraction may provide more dimensions in motor characteristics of patients. Furthermore, utilizing a waveform that requires faster contraction or reflexion velocity, such as a step function, may be useful to investigate the response of patients against the excited movement.

CHAPTER 4

A Prediction Model for Functional Outcomes in Degenerative Spinal Cord Disorders Patients using Gaussian Process Regression

4.1 Objectives

There are approximately 400,000 patients suffering from spinal cord disorders in the United States, with nearly 15,000 new patients each year [AFM09,SF01]. The chronic or traumatic degeneration in the spine results in reduction of the spinal canal diameter and compresses the spinal cord, which is the central pathway for electrical transmission between the central nervous system and the peripheral nervous system. The reduction of the spinal diameter impairs the transmission of electrical signals, and thus results in loss of sensory and/or motor function [AFM09]. The physiological symptoms of SCI that are associated with hand movements include the loss of hand dexterity, numbness, stiffness, weakness, fatigue, and termor. More specifically, characteristic symptoms of SCI at cervical regions include hyperreflexia that shows exaggerated reflexes of the muscles such as twitching [You00]. As a consequence, patients with SCI often have problems coordinating fine movements using hand muscles, which restrict various daily activities such as eating, bathing, or lifting small objects [LLM05].

There exist various methods that quantify physical conditions and/or the level of motor deficits of SCI patients such as radiological imaging (e.g., X-ray, MRI,

and CT) [TB10], clinicians observations (e.g., finger-to-nose or heel-to-shin examinations) [LGD98, Ma03], and patient-reported functional outcomes. Among these techniques, patients' self-ratings of perceived level of motor function and quality of life, such as the Oswestry Disability Index (ODI) [Kop00] and Short Form 36 (SF-36) [Bom00], have been widely used as primary measures for clinical effectiveness [Fin10, TS96, PA00] since the fundamental objective of medical treatment is to improve the well-being of patients [AKE98].

Predicting the patient-reported functional level after medical treatments (e.g., surgical operations, rehabilitation, or medication) has always been of great interest [PWA98, Dij99, LPH11, THC08]. Accurate post-treatment prediction allows (i) defining realistic and achievable clinical goals to patients and their loved ones, (ii) stratifying patients for appropriate rehabilitation processes, (iii) planning for proper discharge schedules, (iv) anticipating and preparing for home and community adjustment, and (v) optimizing medical costs by supporting appropriate medical services [KWK96, VKW11]. Regression based on various demographic and clinical variables has been the most commonly used prediction platform since patient-reported outcomes are often real values. Most existing works employ regression models that assume a predefined relationship (i.e., often linear) between the predictors and the outcome. Although these methods can be implemented easily and provide clear interpretability, they have two major shortcomings. First, the assumption of a predefined relationship (e.g., linear, log-linear, polynomial, or exponential) may not necessarily be true because one form of measure may be more sensitive at a certain range of physical conditions than others. Second, these types of regression methods focus on fitting the data points to minimize the prediction error, and as a result, they produce a single best value rather than providing a probabilistic prediction (i.e., predictive distribution); predictive distribution is especially important since it provides a comprehensive summary about the prediction. For instance, the variability of the prediction distribution can be

used as a reasonable measure of the confidence of a prediction [HSL13].

This chapter introduces a prediction method for post-surgical functional outcomes using Gaussian Process Regression (GPR), which specifically addresses the aforementioned shortcomings of the existing prediction methods. The proposed prediction method is performed in assistance with a simple target tracking examination using a lightweight and inexpensive handgrip device, which contribute significantly to prediction performance. Target tracking based on grip strength is known to effectively quantify the motor condition of patients with hand movement deficits [PBC10, NPL11, LGM13a]. A clinical cohort trial has been conducted in collaboration with the UCLA Department of Neurosurgery and the Department of Orthopedic Surgery in order to validate the proposed prediction method. Note that the proposed method is not only limited to the SCI population but can also be applied to other ailments carrying movement deficits such as Parkinson's disease or stroke, as this chapter provides the design and parametrization details of applying GPR for motor function prediction.

4.2 Related Works

Predicting perceived motor function or quality of life in SCI patients has been of great interest [LPH11, THC08, PWA98, Dij99]. The work in [LPH11] finds the predictors for distinguishing various life satisfaction trajectories since the onset of SCI rehabilitation. Logistic regression has been used to find the predictors for low and high life satisfaction trajectories. Authors in [PWA98] used multivariate linear regression to find correlation to the self-reported motor function: Sickness Impact Profile (SIP68). This study reports that the severity of the injury is the best predictor. Similar studies have been performed in [THC08] and [Dij99], where the authors have used hierarchical multiple linear regression and step-wise linear regression as the prediction models, respectively. Both studies compared the life

satisfaction to various demographic and clinical parameters including gender, age, and number of rehospitalizations.

Prediction of perceived health status is also an active research topic in other ailments carrying motor deficits. For example, [VKW11] reviewed 48 articles on prediction of functional outcomes in patients with stroke. All studies reviewed by [VKW11] employed either linear or logistic regression for prediction. Similarly, [SMM11] reviewed 29 articles that predict health-related quality of life in Parkinson's disease. This paper reports that multivariate regression was the most frequently used prediction platform followed by step-wise linear regression and hierarchical multivariate linear regression. [BWB05] utilizes linear and logistic regressions in order to predict motor function related quality of life in patients with multiple sclerosis.

Although the aforementioned works examine various variables for predicting self-reported functional outcomes, their prediction models have two major shortcomings; (i) they assume predefined relationships between predictors and the target variables (i.e., often a linear relationship) and (ii) they return a single value that best fits the predefined models instead of providing a probability distribution of the prediction. As a consequence, this chapter focuses on introducing a prediction model that overcomes these shortcomings by a novel use of GPR. GPR has recently received much attention for predictions in various fields including biomedicine [SFM08], robotics [XCO11], and communications [JXC13]. GPR also has been used to predict speech behavioral outcomes in stroke patients based on parameters obtained from MRI image processing and demographics [HSL13]. However, this method considers GPR as a black box without providing significant interpretation about the value of hyperparameters and associated prediction results. On the other hand, this work provides estimation methods for some of the hyperparameters by introducing a number of assumptions that allow for more systematic uses of GPR and minimizes the chances of overfitting. Furthermore,

this work provides an in-depth interpretation about the prediction results by investigating the covariance matrix.

4.3 Materials

4.3.1 Cohort Clinical Trial

This chapter validates the prediction methods, which will be discussed in detail in Section 4.4, through a dataset collected from a 24 month-long cohort trial. A total of 34 cervical spinal cord injury patients with hand movement deficits participated in the study. 7 of these patients decided to drop out of the study, and 9 patients were new patients whose three months post-operative data has not yet been collected. Data collected from 3 patients were corrupted and removed from the study due to either mistakes during the data collection process or malfunction of the system itself. As a consequence, this study is validated through a dataset collected from 15 patients (mean age of 62.3 and a std. dev. of 13.1). All patients have received a surgical decompression operation that alleviates the nerve pressure on the spine. The examination procedure was approved by the UCLA institutional review board, and all patients provided consent to participate in the study.

4.3.2 Measure of Self-Reported Motor Function

There exist various forms of self-reported functional outcomes. The Oswestry Disability Index (ODI) has been used as one of the primary condition-specific assessment tools for general SCI patients [Kop00] as well as patients with upper limb deficits [FP00]. The ODI is a survey consisting of ten questionnaires regarding the level of pain in the affected area and the degree of disabilities in everyday activities such as sleeping, self-care, sex life, social life, and traveling [FP00]. Each question in the questionnaire has five or six answer choices, and patients must

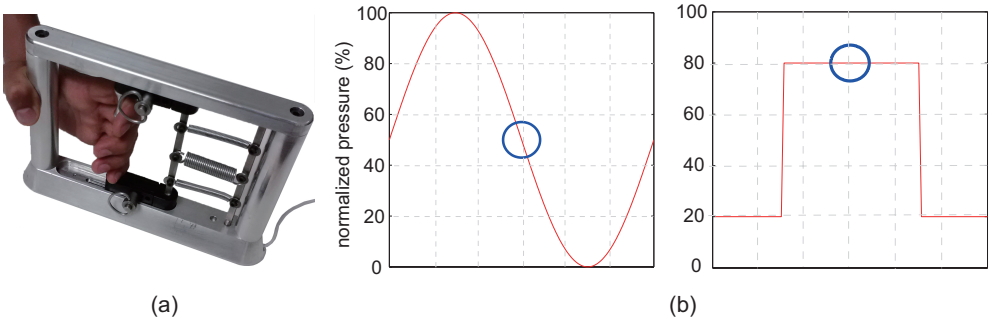

Figure 4.1: (a) The handgrip device and (b) the two target functions used in this study.

select the one that best describes the effect of pain in performing these activities. The answer is scaled to [0, 1] based on the number of available choices, where zero indicates completely disabled conditions in performing the specified activity, and one indicates completely healthy condition. The overall ODI score is computed as the mean value over the ten questions.

4.3.3 Demographic and Clinical Variables

The objective of this work is to introduce a prediction method for post-surgical functional outcomes based on a patient's pre-surgical information. Various clinical and demographic variables are used to predict the changes in motor function after the surgery. Variables considered in this work include age, narrowest diameter of the spinal canal before surgery, and months after injury. The narrowest diameter of the spinal canal is measured using conventional X-ray imaging; the narrowest diameter at the level of greatest herniation in the mid-sagittal cut is measured in centimeters. The number of months after injury is reported by patients.

4.3.4 Measure of Subjective Fine-Motor Function

As previously stated, the prediction is performed in assistance with a specially designed handgrip device and target tracking tests in order to better describe the functional conditions of patients prior to surgery. The target tracking examination based on grip force has been widely used to quantify the level of fine-motor function in various types of ailments including Parkinson's disease [PBC10], stroke [NPL11], and chronic inflammatory demyelinating polyneuropathy (CIDP) [LGM13a].

The handgrip device, which also has been used in [LGM13b], is illustrated in Fig. 4.1 (a). All tracking tests are normalized to the patient's maximum voluntary contraction (i.e., the maximum grip strength that a patient can voluntarily exert) in order to accommodate patients with different grip strengths. To do so, a set of three springs is chosen such that the combined tension force is greater than the patient's MVC with minimal difference. Then, the software measures the actual MVC and normalizes the maximum amplitude of the target to the MVC.

The test is performed on two different target functions as shown in Fig. 4.1 (b): sinusoid and step functions. The sinusoid function is known to examine motor-learning capability for fine muscle control, and the step function is known to investigate a patient's ability to predict and execute relatively rapid hand movement [Jon00]. The same examination procedure is applied to both functions. As the test begins, the target within the screen moves towards the left at a constant speed. The blue circle is always located in the middle of the horizontal axis, but its vertical position varies according to the applied grip strength. The length of the test is 45 seconds. The objective of the test is to minimize the error between the target and the patient's response. Mean Absolute Error (MAE) is the most frequently used motor measure [Jon00], which is computed as the mean value of the absolute differences between the target and the patient's response over the

period of the test. In order to follow the *general system performance theory* that all dimensions of human performance are in a form for which a higher numerical value represents superior performance [Kon95], this work uses Mean Absolute Accuracy (MAA), which is computed as $(1 - \text{MAE})$, as the primary metric for quantifying motor level. Patients were asked to repeat the tracking tests three times, and the average MAA of the three tests was used as the final measure (for both sinusoid and step functions). Note that the reliability of MAA for sinusoid and step functions has been validated on a dataset collected from 15 SCI patients and 40 age-match control subjects (mean age of 57.3 and a std. dev. of 9.71). The Test-Retest score has been computed by incorporating the last two scores of the three trials when a subject performed the test for the very first time (e.g., pre-surgical trial for patients). The R^2 values for the sinusoid and step functions are computed respectively as 0.922 and 0.936, which are highly reliable scores. Furthermore, the internal consistency has been computed using Cronbach's alpha based on the average sinusoid and step function scores, and the value was computed as 0.894, which also supports that these tracking scores are internally consistent.

4.3.5 Longitudinal Study

Patients were asked to visit the clinic at least once prior to the operation. Then, patients were scheduled to have a follow-up visit at least three months after the surgery, since a three month period is a clinically meaningful time for recovery [Fin10]. Both self-reported functional outcomes and the target tracking results were collected at each clinical visit.

Table 4.1: Summary of the important notations used in this chapter.

Symbol	Description
n	Number of data points (patients)
d	Size of feature dimension
$\boldsymbol{X}$	$n \times d$ training design matrix
$\boldsymbol{y}$	$n \times 1$ training observation vector
ρ, ρ'	ODI score at pre-, post-surgery
α, α'	Sinusoid MAA at pre-, post-surgery
β, β'	Step MAA at pre-, post-surgery
g	Age
r	Narrowest diameter of spinal canal
h	Months after injury

4.4 Prediction Method

This section discusses the prediction method in detail, which estimates the post-surgical conditions of patients given their pre-surgical information. The important notations used in this chapter are summarized in Table 4.1.

4.4.1 Background

Prediction problems often involve a finite set of independent variables $\boldsymbol{X}$ (i.e., pre-surgical data of patients) and the associated noisy observation of a dependent variable $\boldsymbol{y}$ (i.e., post-surgical data of patients). $\boldsymbol{X}$ is an $n \times d$ matrix where n represents the number of data points (i.e. the number of patients in this context) and d represents the size of variable dimension. $\boldsymbol{X}$ is also often called the design matrix, and can be written as $\boldsymbol{X} = \{\boldsymbol{x}_i | i = 1, ..., n\}$ where $\boldsymbol{x}_i$ is a vector of dimension d. $\boldsymbol{y}$ is the target (or observation) vector of size n. The relationship

between the input vector and the output variable can be written as

$$y_i = f(\boldsymbol{x}_i) + \varepsilon, \tag{4.1}$$

where $f_i = f(\boldsymbol{x}_i)$ is the latent (hidden) variable that represents the true physiological condition of the patient i, and ε represents the noise added to the observed variables. Then, the prediction problem can be stated as the following: given a training dataset $D = \{\boldsymbol{X}, \boldsymbol{y}\}$, find the best estimate of the dependent variable $\boldsymbol{y}_*$ of a new testing set $\boldsymbol{X}_*$.

Since the dependent variables are real values, the prediction problem can be formalized in a regression setting. The simplest solution to this problem may include the linear regression model. Despite of its simple implementation and interpretability, linear regression assumes a linear relationship between the independent variables (e.g., pre-surgical MAA under sinus track) and the target variable (e.g., post-surgical ODI), which may not necessarily be true since one form of the metric may be more sensitive at a certain range of physical conditions than the other. For instance, the sinusoid MAA may provide finer granularity in quantifying physical condition for the patients with relatively healthier conditions compared to the patients with severe conditions. This problem is not only limited to linear regression but also to all other types of regression models that assume a particular relationship between the dependent and independent variables: e.g., polynomial regression model.

On the other hand, GPR takes a less parametric approach. GPR constructs a relationship between $\boldsymbol{X}$ and $\boldsymbol{y}$ based on the geometric positions of $\boldsymbol{x}_i, \forall i$ within the feature space. The dependent variable $\boldsymbol{y}$ is considered as a collection of samples from an n-variate Gaussian distribution, and it allows GPR to provide not only the expected value of $\boldsymbol{f}_* = f(\boldsymbol{X}_*)$, but also the confidence range of $\boldsymbol{f}_*$ by incorporating the known variances (error ranges) of the variables (e.g., ODI and MAAs).

4.4.2 Gaussian Process Regression

The formal definition of a *Gaussian Process* is a collection of a finite number of random variables which has a joint Gaussian distribution [RW06], and it can be completely defined by its mean function $\mu_i = E[f(\boldsymbol{x}_i)]$ and covariance function (also known as kernel) $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ where $1 \leq i, j \leq n$: $f(\boldsymbol{x}_i) \sim GP\left(\mu_i, k(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)$. In this work, the *squared exponential* (SE) covariance function is employed to define the relationship between the observations:

$$\begin{aligned} \operatorname{cov}\Big(f(\boldsymbol{x}_i), f(\boldsymbol{x}_j)\Big) &= k(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &= \sigma_f^2 \exp\left(-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{x}_j)^T M(\boldsymbol{x}_i - \boldsymbol{x}_j)\right) + \sigma_n^2 \delta(\boldsymbol{x}_i, \boldsymbol{x}_j), \end{aligned} \tag{4.2}$$

where $M = \operatorname{diag}(\boldsymbol{\ell})^{-2}$ and $\boldsymbol{\ell} = \{\ell_k | k = 1, ..., d\}$. ℓ_k represent *characteristic length-scale* for each input dimension, and as a consequence M forms a $d \times d$ matrix with its diagonal consisting of ℓ_k and zero elsewhere. These parameters define the relationship between each of the independent variables and the target variable; it allows different relationships for different independent variables. σ_f^2 represents the maximum allowable covariance between the input variables in the d-dimensional feature space. σ_n^2 represents the variance in the observed noise ε in (4.1) assuming that ε is an independently and identically distributed Gaussian distribution: $\varepsilon \sim N(0, \sigma_n^2)$. Last, $\delta(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a Kornecker delta whose value is one when $\boldsymbol{x}_i = \boldsymbol{x}_j$, and zero otherwise.

4.4.3 Post-surgical ODI

This section, without loss of generality, considers predicting the post-surgical ODI based on the patient's pre-surgical data. Note that the proposed model can also be applied to any other movement disorders with its own measure of motor capacity. The equation (4.2) indicates that the variance of the dependent vector is defined based on the geometric positions of the input vectors within the feature space, which are eventually used to predict the expected value and the variance

of the target variable $\boldsymbol{y}_*$ (i.e., (4.3)). The underlying hypothesis is that clinical and demographic information, in combination with the metrics obtained from the handgrip device prior to the surgery, would construct an effective feature space for the prediction.

As discussed in Section 4.3, a total of six features are considered: (i) pre-surgical ODI (denoted as ρ_i), (ii) pre-surgical MAA under the sinusoid track (denoted as α_i), (iii) pre-surgical MAA under the step track (denoted as β_i), (iv) the patient's age (denoted as g_i), (v) the narrowest diameter of the spinal canal prior to the surgical treatment (denoted as r_i), and (vi) months after injury (denoted as h_i). However, a subset of three features are selected to construct the feature space instead of using all six features in order to avoid overfitting to the relatively small size of the dataset. Advanced feature selection algorithms such as those proposed in [LGM13a, PRM11, PMA13] could be employed in order to select the best performing feature set. However, since the scope of this work is to introduce a new prediction platform based on GPR, a rather simple feature selection technique based on an exhaustive search is used. That is, all possible combinations of feature set of size three (i.e., $\binom{6}{3}$) are evaluated in terms of mean absolute difference (MAD) between the prediction and the ground truth. Then, the feature set that produces the best prediction result is reported. The same feature selection procedure has been taken for the benchmarking models (e.g., multivariate linear model and support vector regression) such that their best prediction results are compared against the results of the proposed method.

The input and the target variables can be mathematically expressed as $y_i = \rho_i'$ and $\boldsymbol{x_i} \subset_3 \{\rho_i, \alpha_i, \beta_i, g_i, r_i, h_i\}$ where i represents the patient index and $\subset_3$ is a symbol for a subset of size three. Then, given the training data set $D = \{\boldsymbol{X}, \boldsymbol{y}\}$ and the testing input vector $\boldsymbol{X}_*$, the Gaussian process can be written in matrix

form as

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_* \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}_* \end{bmatrix} \begin{bmatrix} K(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2 I & K(\boldsymbol{X},\boldsymbol{X}_*) \\ K(\boldsymbol{X}_*,\boldsymbol{X}) & K(\boldsymbol{X}_*,\boldsymbol{X}_*) \end{bmatrix} \right),$$

where $K(\boldsymbol{X},\boldsymbol{X}_*)$ is a $n \times n_*$ matrix which represents covariance between each data point in $\boldsymbol{X}$ and each point in $\boldsymbol{X}_*$. $K(\boldsymbol{X},\boldsymbol{X})$, $K(\boldsymbol{X}_*,\boldsymbol{X})$, and $K(\boldsymbol{X}_*,\boldsymbol{X}_*)$ are also defined in a similar manner. Then the conditional distribution of the prediction variable $\boldsymbol{y}_*$ is also a Gaussian: $\overline{\boldsymbol{y}_*} = \boldsymbol{y}_* | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_* \sim N\Big(E(\overline{\boldsymbol{y}_*}), \mathrm{cov}(\overline{\boldsymbol{y}_*})\Big)$. Note that $E(\overline{\boldsymbol{y}_*})$ represents the expected value of the prediction variable and $\mathrm{cov}(\overline{\boldsymbol{y}_*})$ can be used to compute the confidence range. These two values can be computed based on the *Multivariate Gaussian Theorem* as

$$\begin{aligned} E(\overline{\boldsymbol{y}_*}) &= E(\boldsymbol{y}_* | \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*) \\ &= \boldsymbol{\mu}_* + K(\boldsymbol{X}_*,\boldsymbol{X})\Big[K(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2 I\Big]^{-1}(\boldsymbol{y} - \boldsymbol{u}) \\ \mathrm{cov}(\overline{\boldsymbol{y}_*}) &= K(\boldsymbol{X}_*,\boldsymbol{X}_*) \\ &\quad - K(\boldsymbol{X}_*,\boldsymbol{X})\Big[K(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2 I\Big]^{-1} K(\boldsymbol{X},\boldsymbol{X}_*). \end{aligned} \tag{4.3}$$

4.4.4 Hyperparameters

Intuitive meanings of hyperparameters (σ_f^2, σ_n^2, and $\boldsymbol{\ell}$) in (4.2) are as follows. As discussed earlier, σ_f^2 represents the maximum covariance between the two input variables. The maximum covariance is achieved when the two inputs are equal (i.e., $k(\boldsymbol{x}_i, \boldsymbol{x}_i)$), and thus σ_f^2 is closely related to the covariance of input vector $\boldsymbol{x}_i$ in each of the three feature dimensions. It is especially difficult to come up with the theoretical value for σ_f^2 since the variance in each dimension of ρ, α, β, g, r, and h is different. Furthermore, the inter-relationships between these variables are not verified. $\boldsymbol{\ell}$ represents a vector of characteristic length scales that determines sensitivity to changes in covariance according to the distances of two input vectors in each dimension. Both σ_f^2 and $\boldsymbol{\ell}$ are analytically intractable, and thus these parameters are selected by maximizing *a posteriori* estimates based on the training data: $p(\sigma_f^2, \boldsymbol{\ell} | \boldsymbol{X}, \boldsymbol{y})$, which is equivalent (based on the Bayes'

Theorem) to maximizing log marginal likelihood:

$$\log p(\boldsymbol{y}|\boldsymbol{X}, \sigma_f^2, \boldsymbol{\ell}) = -\frac{1}{2}\boldsymbol{y}^T K^{-1}\boldsymbol{y} - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi. \tag{4.4}$$

On the other hand, σ_n^2 can be estimated based on previous studies performed on ODI. Assigning a near-theoretical value to σ_n^2 (instead of computing the value based on maximum likelihood) is especially important since it reduces the chances of overfitting by minimizing the number of hyperparameters that need to be tuned to the training data. As shown in (4.1), σ_n^2 models the observation error added to the hidden variable $f(\boldsymbol{x}_i)$. In this context, this hidden variable represents the true physiological condition of the patient i. However, it is impossible to compute this variance since the true physical condition of a patient cannot be evaluated. In order to estimate the σ_n^2, this work assumes that this variance is constant throughout all ranges of physical conditions, i.e., σ_n^2 is not a function of $\boldsymbol{x}$. Then, one can estimate σ_n^2 by investigating the standard deviation of the ODI score of the normal population under an additional assumption that the true physical condition of the normal population is constant or minimally varying. The range of σ_n^2 is reported to be $0.022 \leq \sigma_n^2 \leq 0.12$ according to [FP00]. In this work, the worst scenario is considered by assigning $\sigma_n^2 = 0.12$. The empirical validation on this theoretical value is further investigated in Section 4.5.4.

4.4.5 Post-surgical MAA

Similar procedures have been taken for predicting post-surgical MAA values. In this work, rather than predicting MAAs for sinusoid and step tracks separately, a more comprehensive MAA is considered by averaging the sinusoid MAA and step MAA, i.e., $\boldsymbol{y} = \frac{1}{2}(\boldsymbol{\alpha}' + \boldsymbol{\beta}')$.

The values of σ_f^2 and $\boldsymbol{\ell}$ are computed by maximum marginal likelihood. Similarly to the ODI prediction, the value of σ_n^2 is estimated based on the control subject's performance under the same assumptions that (i) the σ_n^2 is constant

throughout all ranges of physical conditions and (ii) the true physical conditions of control subjects are constant or minimally varying. The standard deviation of $\frac{1}{2}(\boldsymbol{\alpha}'+\boldsymbol{\beta}')$ of control subjects is computed to be 0.0165, and thus $\sigma_n^2 = 0.0165$. The empirical validation on this theoretical value is also investigated in Section 4.5.4.

4.5 Experimental Results

4.5.1 Prediction of Post-surgical ODI

As discussed previously, the hyperparameters σ_f^2 and $\boldsymbol{\ell}$ are selected such that they maximize the marginal likelihood, i.e. (4.4). The observation noise σ_n^2 is set to be 0.12. With the computed hyperparameters, the expected value and the confidence range can be predicted based on (4.3). The prediction is performed by a leave-one-subject-out cross validation (LOSOCV). The best prediction result of the proposed method is achieved when the following features are used: $\{\alpha, \rho, r\}$. The prediction result is illustrated in Fig. 4.2 where the x-axis represents the actual ODI that the patient evaluated after the surgery, and the y-axis represents the predicted ODI of the proposed method. More specifically, Fig. 4.2 (a) shows the scatter of the mean values of prediction, and Fig. 4.2 (b) shows the mean values with 95% confidence range. The R^2 of this prediction was computed as 0.435, and the MAD between the prediction and the ground truth is 0.096.

The covariance matrix $K(\boldsymbol{X}, \boldsymbol{X})$ is analyzed in order to find a more meaningful interpretation for this result. $K(\boldsymbol{X}, \boldsymbol{X})$ has been constructed by finding the hyperparameters σ_f^2 and $\boldsymbol{\ell}$ based on the entire dataset (although the results in Fig. 4.2 are produced from the LOSOCV) in order to find possible correlation between all patients in the selected feature dimensions. The values of marginal likelihood hyperparameters are computed as $\sigma_f^2 = 0.77$ and $\boldsymbol{\ell} = \{1.59, 0.22, 6.55\}$ in the order of $\{\alpha, \rho, r\}$. Fig. 4.3 (a) illustrates the normalized covariance matrix; the covariance matrices are normalized to a 0 to 100 scale in order to compare

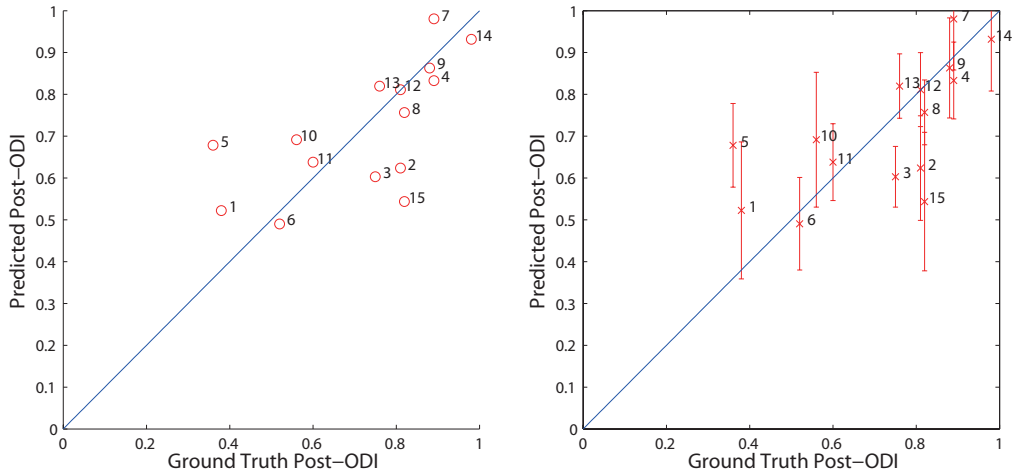

Figure 4.2: Prediction of post-surgical ODI based on pre-surgical information.

$K(\boldsymbol{X}, \boldsymbol{X})$ of ODI and MAA. This matrix is a symmetric matrix whose value is at its maximum when the two inputs are identical as identified in (4.2).

Note that there exist four predictions where the ground truth does not fall within the 95% confidence range from the expected values in Fig. 4.2: #2, #3, #5, and #15. These four points are the points that produce the largest four prediction errors. Especially, #5 and #15 have noticeably large prediction errors (i.e., the highest two errors). Interestingly, the two patients whose maximum correlation to other patients are the lowest are also #5 and #15 (computed from $K(\boldsymbol{X}, \boldsymbol{X})$ in Fig. 4.2 (a)). This implies that more patient data with similar pre-surgical characteristics within the selected feature dimension is needed to improve the prediction accuracy for these patients. On the other hand, #2 and #3 have a sufficient number of neighboring data points, but produced relatively high errors. This implies that they requires additional feature dimensions.

The prediction performances of the proposed method are also compared against the two benchmarking prediction models: multivariate linear regression (MLR) and support vector regression (SVR). As discussed earlier, MLR is the most commonly used model that assumes simple linear relationships between the input

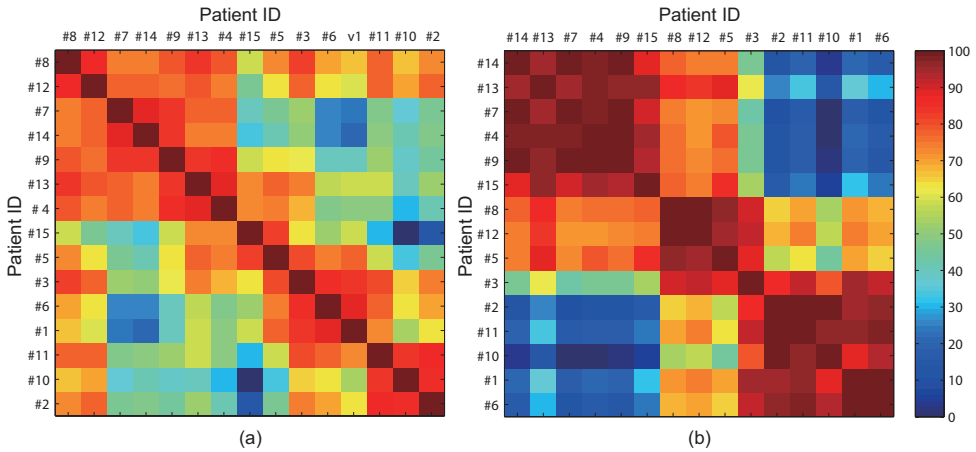

Figure 4.3: The covariance matrix in the selected feature dimensions for (a) ODI and (b) MAA.

Table 4.2: Comparison of the proposed method against the simple multivariate linear regression.

Prediction	ODI			MAA		
Method	Features	R^2	MAD	Features	R^2	MAD
GPR	$\{\alpha, \rho, r\}$	0.435	0.096	$\{\alpha, \rho, g\}$	0.761	0.011
SVR	$\{\alpha, \beta, \rho\}$	0.301	0.097	$\{\alpha, \beta, \rho\}$	0.476	0.023
MLR	$\{\rho, g, h\}$	0.286	0.110	$\{\alpha, r, h\}$	0.220	0.026

features to the dependent variable. SVR is a less parametric model similar to GPR. However, this model also reports the prediction results only with the expected value without much information regarding its confidence range. The results are summarized in Table 4.2. Similarly to the result of the proposed method, the prediction results of the best performing feature sets are reported for both MLR and SVR. Table 4.2 shows that the proposed method produces superior prediction results compare to MLR. Furthermore, the proposed method shows similar (i.e., prediction of ODI) or superior (i.e., prediction of MAA) performance compared to SVR.

4.5.2 Prediction of Post-surgical MAA

A similar procedure has been taken for predicting post-surgical MAA. As discussed earlier, σ_n^2 is set to 0.0165 and the other two hyperparameters are found based on the marginal likelihood maximization. The best prediction performance is obtained when the following features are employed: $\{\alpha, \rho, g\}$. The prediction result is illustrated in Fig. 4.4. The R^2 of this prediction is computed to be 0.761, and the MAD between the prediction and the ground truth is 0.011. There was only one data point where the ground truth value did not fall within the 95% confidence range: #15. Again, the covariance matrix $K(\boldsymbol{X}, \boldsymbol{X})$ in Fig. 4.3 (b) is consulted for explanation. Similarly to the prediction result of ODI, the patient, whose maximum value of its correlation to other patients in the selected feature dimension is minimum, is found to be #15.

The prediction results of post-surgical MAA were much more accurate compared to that of ODI as summarized in Table 4.2. Fig. 4.3 shows that the covariance values were much greater within the selected feature dimension for predicting MAA compared to ODI. In other words, there exist more patients with similar pre-surgical characteristics under the maximum marginal likelihood hyperparameters. As a consequence, the prediction results of MAA may be superior to ODI. This higher prediction accuracy of MAA may be a result of the handgrip device providing less noise compared to ODI for evaluating the true physical conditions of patients (i.e., hidden variable).

4.5.3 Contribution of Target Tracking Results to Prediction

In order to validate the contributions of target tracking results in predicting ODI and MAA, the results of the best performing feature sets, which both include the target tracking results as shown in Table 4.2, have been compared against the best performing feature sets that do not contain the target tracking results. This

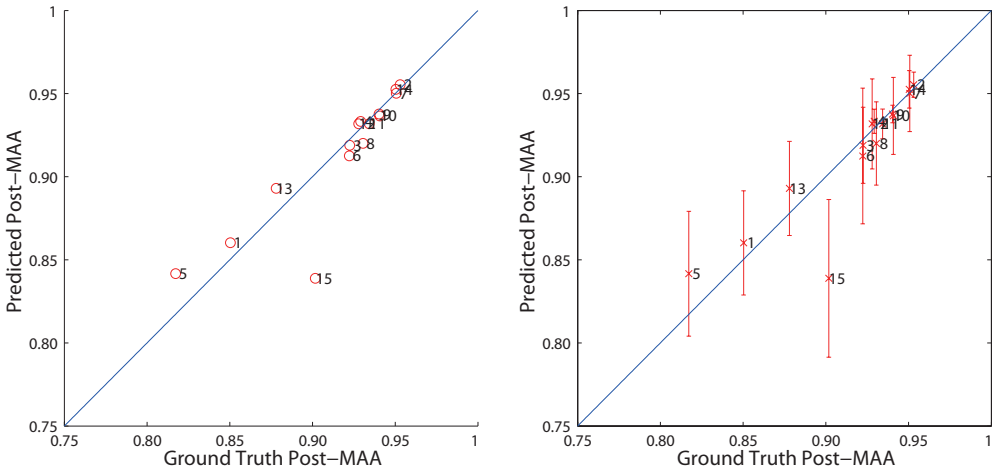

Figure 4.4: Prediction of post-surgical MAA based on pre-surgical information.

Table 4.3: Comparison between the best feature set containing the target tracking scores (FS #1) and the feature set constructed solely based on demographical and clinical variables (FS #2)

Prediction	ODI			MAA		
	Features	R^2	MAD	Features	R^2	MAD
FS #1	$\{\alpha, \rho, r\}$	0.435	0.096	$\{\alpha, \rho, g\}$	0.761	0.011
FS #2	$\{\rho, g, r\}$	0.115	0.274	$\{\rho, g, h\}$	0	0.060

allows the comparison between a feature set that contains the target tracking results (abbreviated as FS #1) and a feature set constructed solely based on demographic and clinical variables (abbreviated as FS #2). This comparison is summarized in Table 4.3. This result shows that incorporating the handgrip test for predicting post-surgical functional outcomes improves the MAD by 17.8% and 4.90% for ODI and MAA, respectively. Furthermore, the R^2 has been significantly improved from 0.115 to 0.435 for ODI and 0 to 0.761 for MAA.

4.5.4 Empirical Validation of σ_n^2

The values of σ_n^2 for estimating post-surgical ODI and MAA are set to 0.12 and 0.0165, respectively. These values are validated by comparing to the empirical σ_n^2 that maximizes the marginal likelihood over a wide range of values, such as 10^{-5} to 10^3. The empirical σ_n^2 for ODI and MAA were computed to be 0.1186 and 0.027, respectively. This is strong evidence that the theoretical estimations of σ_n^2 are valid.

4.6 Discussion and Future Work

The major objective of this study was to introduce a novel use of GPR for predicting post-operative functional outcomes based on information available prior to the operation. The data collection was a very expensive and time-consuming process as discussed in Section 4.3.1. To compensate the limitation posed by the size of the data set, the prediction was evaluated in a LOSOCV, and the achieved results were promising. This pilot study opens a new opportunity for a study with a larger patient population where more patient-specific techniques can be applied. For example, a clustering algorithm prior to the proposed prediction method may enhance the performance as different prediction models can be constructed among patients with similar characteristics. Furthermore, with a larger number of patients, more advanced feature selection algorithms such as those in [LGM13a, PRM11, PMA13] can be applied.

The proposed prediction method is validated through a dataset collected from patients with cervical SCI. In order to show that the method can be applied to a broader classes of SCI patients, a study involving lumbar (i.e., lower back) SCI patients with sensor-equipped smart shoes is being conducted.

4.7 Conclusion

This chapter introduces a prediction method for post-surgical functional outcomes by a novel use of Gaussian Process Regression. Two functional outcomes are considered in this chapter: Oswestry Disability Index and the target tracking score. The reported results show that the proposed method can predict post-surgical outcomes with 9.6% and 1.1% of error rates for ODI and target tracking, respectively. The proposed method has been compared against two widely-used benchmarking prediction models (i.e., multivariate linear regression and support vector regression), and showed superior prediction performance. This study enables new opportunities for accurate prediction of post-surgical conditions of individuals with handgrip deficits using an inexpensive and portable device. This further enables clinicians to perform more ubiquitous and convenient screening for predicting a patient's functional level before medical treatment, which may be especially beneficial to the patients, their care-givers, and physical therapists.

CHAPTER 5

Application of the Target Tracking Tests on Patients with Other Neuromotor Ailments

5.1 Objectives

Ailments such as stroke [AZD04], Parkinson's Disease (PD) [Mor00b], spinal cord injuries [Con73], and many other neuro-degenerative disorders are commonly associated with movement deficits, which affect the function of motor neurons and restrict the movements of the body such as those of the upper limbs, gait, and speech performance [Cri81].

Currently, the available assessment methods for the progress of patients with associated ailments are based on human observations of motor performances (e.g., finger-to-nose test) [JJE06,FJO75]. Clinical professionals use these measurements for preliminary scanning in order to diagnose ailments in early stage. Early detection of these ailments can dramatically reduce the risk of the severity of motor deficits [HT08].

Furthermore, medical treatments available for movement disorders are typically a combination of medication, surgical operation, and rehabilitation. These treatments are often evaluated by measuring the motor performance of the patients before and after the specific service (e.g., surgery), again, based on human observations. However, these methods suffer from the subjective nature of the measurements, which are often based on limited ordinal scales [FJO75]. This subjectivity creates a need for quantitative assessment methods, such that the

analysis of patients' motor performances can be made more accurate and objective [GDB02].

Researchers have studied various methods to objectively quantify the level of upper limb movement. Among those methods, handgrip performance has been known as a simple, accurate, and economical bedside measurement of muscle function and the progression of the movement disorders [DPa06, RPa06, LWN12, JRK09, MC97, KGG05, GLa13, LGa13a, LGa13b, LGM13b]. The grip control is of extreme importance in performing fundamental daily activities such as eating, brushing teeth, and getting dressed. However, existing works often employ equipment that are either very large in size or extremely expensive. Moreover, they often lack in-depth analysis on patients' motor performance with respect to their ailment conditions.

Given the current standard of healthcare for movement disorder patients, there is a need for innovative technologies that (i) provide wearable and portable devices that can be used on a daily basis in many settings; (ii) can be used for individuals' stratification so that such systems are applied on healthy individuals to potentially provide early alarming of any movement disorders; (iii) quantify the level of severity of the specific disorder for a patient; (iv) provide insight on disease symptoms by specifying each abnormality/symptom in terms of signal-specific features.

This chapter introduces a lightweight and inexpensive handgrip device that collects multi-dimensional sensory data associated with motor characteristics of individuals with upper limb deficits. Furthermore, a data analytic framework with associated algorithms for individuals' ailment classification, disease severity quantification, and specification of physical symptoms is discussed. The effectiveness of the proposed movement performance assessment framework is demonstrated through a dataset gathered in a clinical trial performed at St. Vincent Medical Center in Los Angeles, USA.

5.2 Related Works

Many studies have examined handgrip performance in order to reflect the motor capacity of patients with movement disorders. The mechanisms used to investigate the handgrip performances can be divided into three broad categories: (i) assessment based on precision grip, (ii) assessment based on Maximum Voluntary Contraction (MVC)[1], and (iii) assessment based on force tracking tasks.

Assessment methods based on precision grip focus on simulating tasks involving precise muscle control such as lifting, holding, or transporting a small object (e.g., chopsticks or a pencil) [Mor00a]. In [DPa06], authors utilize a force sensor embedded apparatus in order to investigate digit forces when an active and dynamic hand grasping movement is simulated. In [HHa03], a device embedded with a force sensor and an accelerometer is used to assess finger strength. In [RPa06], an instrumented glove was used to analyze finger movement of patients with subcortical stroke.

MVC-based assessment methods utilize various systems and devices to measure MVC. In [STa89], a simple dynamometer is used in order to measure the MVC as a measure of recovery and a prognostic indicator for patients with stroke. The system in [MSa00] uses a vigorimeter, which measures the air pressure using a rubber bulb. Then, a few features of MVC are analyzed to reflect the disease progress over time. In [JRK09], authors investigate the handgrip strength and endurance of healthy subjects and patients using a dynamometer, and conclude that handgrip strength and mobility for patients are strongly correlated. The system proposed in [MC97] uses instrumented objects that measure forces applied during tasks such as manipulating a book, or a fork.

Assessments based on force tracking tasks provide visual feedback of patients' hand performances, such that patients can control their grip force to minimize

[1]MVC defines the amount of force that a patient produces when she voluntarily grasps the handgrip device with maximum effort.

the difference between the target and the actual response [KZB04]. The system proposed in this chapter improves upon methods that fall within this category. The system in [KZB04] and [KGG05] utilizes various types of devices, such as a nippers pinch, spherical grip, lateral grip, and cylindrical grip, in order to capture the grip force. The system uses a sinusoidal and ramp waveform for the target. Two features are extracted in order to analyze the patients' motor performance: root mean square error and correlation between the target waveform and the user response. In [JJE06], authors propose a handgrip device that measures forces generated by individual fingers using pressure sensors. In [SN00], authors utilize a grasping apparatus to capture grip forces, and provide a target waveform of a continuous and constant force level on a computer screen.

Most of the aforementioned works (i.e., [Mor00a,DPa06,WKa05,HHa03,RPa06, STa89, MSa00, JRK09, MC97, KZB04]) focus on introducing the developed handgrip devices using a simple metric to validate the effectiveness of those devices. Thus, these works often lack in-depth analysis of the data according to their ailment conditions. Furthermore, some of the devices such as those used in [DPa06, MSa00, JRK09, KZB04, SN00] are either very large in size or extremely expensive.

5.3 System Architecture

The proposed system is composed of (i) the sensing hardware that contains a handgrip device equipped with a force sensor in order to collect the time-varying muscle controllability of patients, and (ii) the software system that visualizes the examination and stores the results in the database. An overview of the proposed system is illustrated in Fig. 5.1.

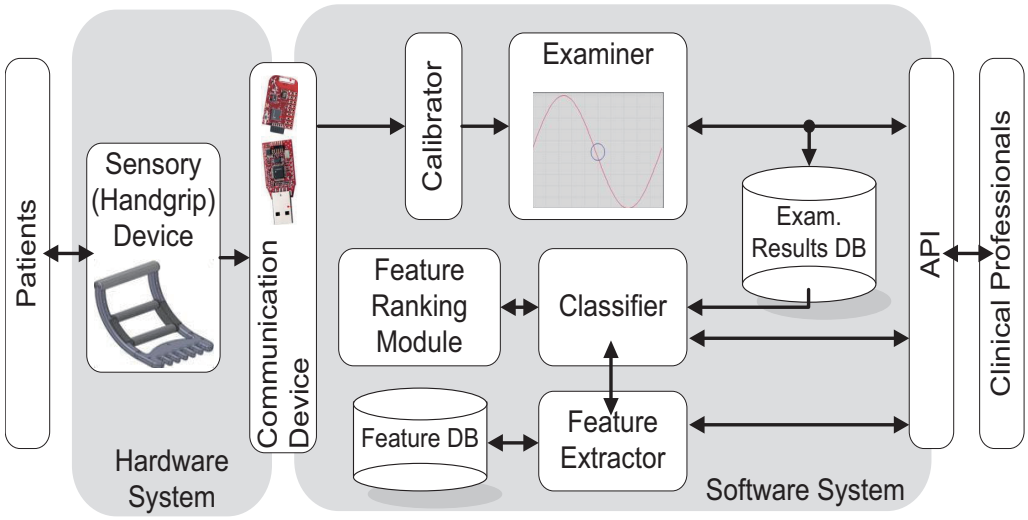

Figure 5.1: The graphical overview of the proposed system.

5.3.1 Sensing Hardware

5.3.1.1 Handgrip Device

The handgrip device is motivated from medical devices that are currently used by clinical professionals (e.g., handgrip device in [WL81]), as shown in Fig. 5.2 (b). The proposed handgrip device is illustrated in Fig. 5.2 (a). The two cylinders bridged by the black Derlin plastic (Part-C in Fig. 5.2 (a)) are movable along the sideline (Part-B in Fig. 5.2 (a)) such that patients can grasp the device. The movable component of the handgrip device is bound to the fork-like side (Part-A in Fig. 5.2 (a)) of the device by a rubber band. Patients can use rubber bands of different tension forces in order to customize the maximum squeeze force in the handgrip device. Additionally, the fork-like side of the handgrip allows patients to further adjust the tension force by placing the rubber band at different widths. Finally, a force sensor is attached to Part-D in Fig. 5.2 (a), and it measures the force generated by the grasping action.

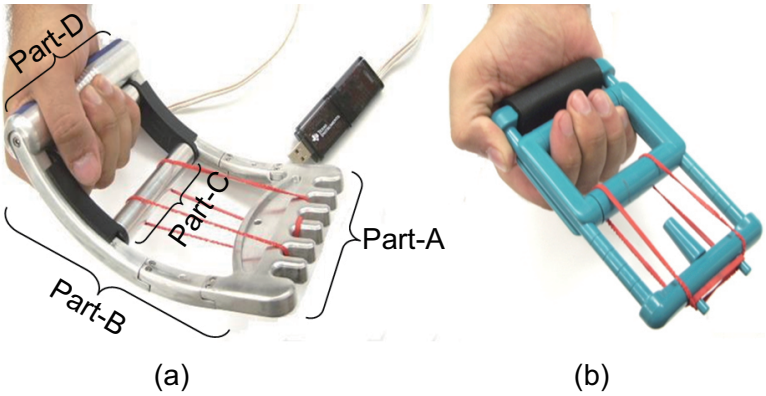

Figure 5.2: (a) The physiologically designed handgrip device of the proposed system (left) and (b) the handgrip device currently in use (right)

5.3.1.2 Sensing Platform and Communication System

A commercial FSR^{TM} force sensor (Interlink Electronics, USA) is attached to the handgrip device in order to measure the grip strength. The FSR sensor is used in the proposed system because (i) the FSR sensor responds accurately and precisely to the general range of handgrip force, and (ii) the FSR shows good performance in terms of robustness [VFa00].

In the proposed system, MSP430 (Texas Instruments, USA) is employed as the communication device, which is capable of delivering the captured sensory data in a wired or wireless manner.

5.3.2 Software Framework

The software is composed of the front-end and the back-end software systems. The front-end software system provides a graphical interface for patients to perform the tracking examination, which will be discussed in detail in Section 5.4. It also provides a number of test parameters, which may change the attributes of the examination such as test duration or difficulty of the test. The back-end software system is a database system, which stores the information such that an in-depth

data analysis becomes possible.

5.4 Examination Procedure

The examination is designed to assess a patient's grip strength as well as a patient's ability to precisely control the strength [KZB04, KGG05]. Fig. 5.3 illustrates an example of a tracking examination using a sinusoidal waveform. When the examination begins, the target waveform is horizontally shifted to the left at a constant speed and, as a result, the user observes a flow of the waveform within the screen. The blue circle in the middle of the screen corresponds to the level of pressure acquired from the force sensor. The blue circle is always located in the middle of the x-axis and its position in the y-axis changes according to the provided pressure. The objective of the examination is to control the grip strength to minimize the difference between the target waveform and the patient response. The software stores the target waveform, the patient's response, and various test attributes (e.g., duration or difficulty of the test) in the back-end database.

Patients may have different handgrip strengths due to various physical conditions. Therefore, the system first measures the MVC prior to the actual examination and normalizes the examination based on the MVC value for each patient. As a result, the labels on the y-axis in Fig. 5.3 represent the percentage of the acquired grip strength compared to the MVC measured in the calibration process (i.e. 100% in the y-axis refers to the patient's MVC).

In summary, the examination considers both the maximum strength and the preciseness of a patient's grip control as explained in [Jon00].

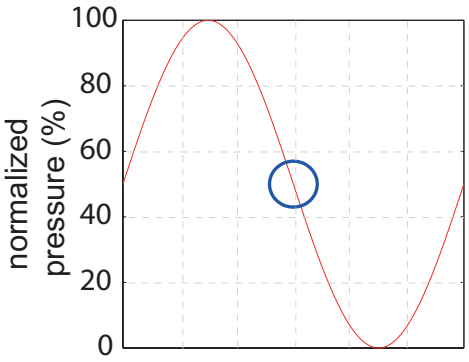

Figure 5.3: Illustration of an examination using a sinusoidal waveform

5.5 Data Analysis

This section provides detailed discussion about (i) metrics that quantify general motor performance of patients and (ii) a comparative analysis methodology that can be used to summarize the characteristics of physical symptoms of patients.

5.5.1 Comprehensive Metric to Quantify Motor Performance

In order to quantify comprehensive motor performance based on a tracking examination, various metrics have been used such as mean absolute difference (MAD) [Pou74], root mean square error (RMSE) [ON95], mean square error (MSE) [NNO93], mean absolute variance (MAV) [NM80], and the standard deviation of error (SDE) [Jon00] between the target waveform and the patient's response. In this work, MAD and MAV are employed in order to quantify comprehensive motor function.

5.5.2 Ailment-Specific Analysis

This section describes the in-depth data analysis method that extracts meaningful information about the characteristics of a group of patients by comparing the examination results with other subjects. For example, consider a scenario where

an elderly candidate patient experiences some upper limb movement disorders and other symptoms related to a specific ailment, e.g., Cerebral Vascular Accident which is also known as stroke. Then, clinical professionals can perform an ailment-specific analysis on previously collected signals of patients with CVA, and reflect the results to the examination outcomes of this elderly patient. The results will allow the clinical professionals to observe (i) if the proposed system can make a clear distinction in terms of motor function between the two groups, and more importantly (ii) which motor characteristics are uniquely observed among patients with CVA that help such distinction. The reason that we perform the ailment-specific analysis is because different ailments are associated with different motor characteristics, and therefore the motor performance of patients sharing the same ailment must be analyzed based on the specific motor characteristics of that ailment. For instance, patients with CVA often carry cognitive impairments, which result in delays between the target waveform and the patient generated waveform due to delayed motor response. On the other hand, patients with Chronic Inflammatory Demyelinating Polyneuropathy (CIDP) only carry physical motor impairments and do not show significant delay in the results. This ailment-specific analysis allows the clinical professional to examine the candidate patient based on features that represent motor symptoms of the associated ailment for the purpose of (i) quantifying the severity of such symptoms, and (ii) possibly tracking the improvement over time.

In order to address this objective, the ailment-specific analysis employs the *significant-feature identification* algorithm [SM03, SS98], which utilizes feature ranking, feature selection, and classification algorithm. The significant-feature identification algorithm performs the classification iteratively in order to observe the most frequently selected feature subsets and their associated classification accuracy.

The ailment-specific analysis begins with forming the group of signals of in-

terest and the group of signals to be compared against, which are used as ground truth labels in order to evaluate the classification performance. Throughout this chapter, the term *group of interest* (GOI) is used to generically represent the signals in which one is particularly interested to analyze. Note that this work also uses the term *positive signals* and *negative signals* to define the signals in the GOI and the signals in the other group, respectively.

The ailment-specific analysis method employs a hold-out strategy which formulates the *learning* and the *validation* dataset in order to perform the classification iteratively [BZ08]. The entire dataset is first divided into the *learning set* and the *validation set* using a leave-one-out cross validation (LOOCV). The *learning set* is used to extract significant features and to construct a classification model, and the *validation set* is used to evaluate the performance of the model. Then, the *learning set* is tested over the feature search space constructed by the feature ranking and feature selection algorithms. In order to do so, each element in the feature search space is evaluated based on another LOOCV within the *learning set*; the *learning set* is further divided into the *training set* and the *testing set* [BZ08]. The feature extraction, feature ranking, and feature selection algorithms performed on the *learning set* are graphically illustrated in Fig. 5.4. The classification model and the signification feature set constructed by the *learning set* are finally evaluated using the *validation set*. Finally, the models and the feature sets constructed by the outer layer cross validation are further processed by *significant-feature identification* method to extract the true significant feature-set.

5.5.2.1 Feature Extraction

Suppose that a set of extracted features from a single examination result is represented as a horizontal array $\overline{s} = [s_1\ s_2 \cdots s_T]$, where T is the number of features. Each feature s_i is computed using a feature extraction function $f_i\left(w_g[n], w_r[n]\right)$, which is either in the time- or frequency-domain.

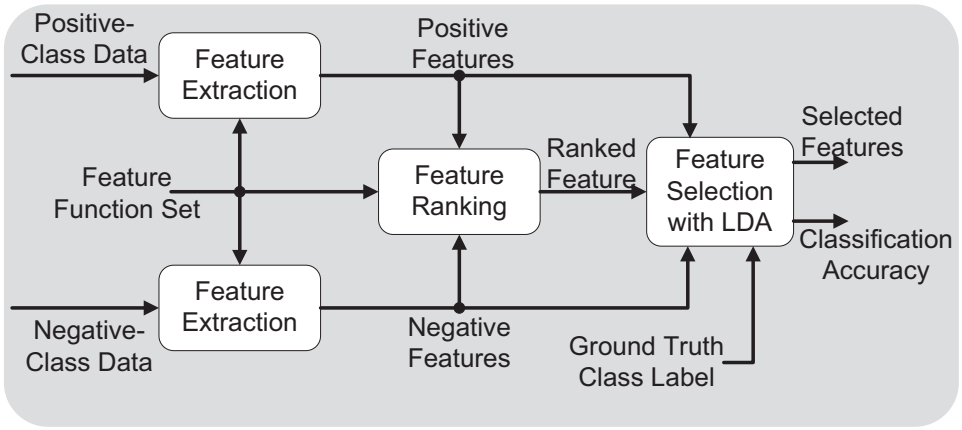

Figure 5.4: The graphical overview and data flow of the analysis methodology on the *learning set*.

Representing the total number of test instances (i.e. both *learning* and *validation* sets) as M, a $(M \times T)$ feature matrix can be computed:

$$S = \begin{bmatrix} \overline{s}^1 \\ \overline{s}^2 \\ \vdots \\ \overline{s}^M \end{bmatrix} = \begin{bmatrix} s_1^1 & s_2^1 & \cdots & s_T^1 \\ s_1^2 & s_2^2 & \cdots & s_T^2 \\ \vdots & & \ddots & \vdots \\ s_1^M & s_2^M & \cdots & s_T^M \end{bmatrix} \tag{5.1}$$

$$= \begin{bmatrix} \overline{s}_1 & \overline{s}_2 & \cdots & \overline{s}_T \end{bmatrix}, \tag{5.2}$$

where $\overline{s}^j$ with $j \in [1, M]$ is a horizontal array of features, and $\overline{s}_i$ with $i \in [1, T]$ is a vertical array composed of values of a feature $f_i(\cdot)$ computed from the entire signals.

5.5.2.2 Feature Ranking and Feature Selection

As explained earlier, the feature ranking and feature selection algorithms are performed on the *learning set*. The reason that feature ranking and feature selection techniques are employed as follows. First, given a pre-defined set of features $\overline{s}$, not all of these features play an important role in classifying a certain GOI. This may lead to a problem during the classification process that a collection of many

useless features can be accumulated to overwhelm some useful features. Second, it is more computationally efficient since it filters out features that are less useful. Lastly, information about the ranks of features according to the level of contributions in the classification process is valuable to us because features with higher rank can be defined as physical symptoms found only in that GOI.

The proposed system employs the estimated Pearson correlation coefficients to rank the features according to the level of correlation to the class labels (i.e., positive or negative signals). The Pearson correlation coefficient for a feature f_i, can be estimated using

$$R(i) = \frac{\sum_{j=1}^{M'} \left(s_i^j - E(\overline{s}_i)\right)\left(y^j - E(\overline{y})\right)}{\sqrt{\sum_{j=1}^{M'} \left(s_i^j - E(\overline{s}_i)\right)^2 \sum_{j=1}^{M'} \left(y^j - E(\overline{y})\right)^2}}, \tag{5.3}$$

where y^j with $j \in [1, M']$ represents the class of the signal j (i.e., $+1$ for positive and -1 for negative class signal). M' represents the size of the *learning set* (i.e., $M' = M - 1$ since LOOCV is applied). $E(\cdot)$ represents a function computing the mean value of the input vector. Then, we use $R(i)^2$ as a feature ranking criterion that estimates goodness of linear fit of an individual feature to the class vector $\overline{y}$ [GE03].

Given the rank of all features, the well-known *forward selection* strategy is used to construct the search space, which starts with the highest ranked feature and gradually adds a feature that is the next highest. Then, the size of the search space is reduced to $T - 1$. Each feature subset is evaluated using Linear Discriminant Analysis (LDA), and the feature subset with the highest averaged classification accuracy is selected.

5.5.2.3 Identifying Significant-Features

When the best performing feature subset is selected, the *learning set* is projected onto the selected features and the classification model is trained. Then, the *val-*

idation set is classified and evaluated using LDA. The classification accuracy is computed by averaging the number of correctly classified instances over M validation sets (created by the outer LOOCV) using LDA. Moreover, the algorithm produces M feature subsets that are selected by the feature selection technique. In order to investigate the significant feature subsets, we compute the followings over the cross validations: (i) the most frequently appearing feature subset, (ii) the top K features for the accumulated ranking score, and (iii) the associated classification accuracy. Intuitively, the more frequently selected feature subset carries more relevant information about the patients that we are interested [SS98].

5.6 Validation

5.6.1 Clinical Trial

We have conducted a clinical trial at St. Vincent Medical Center (Los Angeles, CA). A total of 16 subjects participated in this clinical trial and provided 78 examination instances. Among the participating subjects, 12 of the study's subjects were patients (mean age of 70.5 years with a standard deviation of 10.5), and they produced a total of 67 examination results. The remaining 4 subjects were healthy individuals (mean age of 64.3 years with a standard deviation of 12.1) and the data collected from these subjects resulted in 11 examination instances. All patients had motor deficits in their upper limbs and they were examined prior to any operational treatments. All subjects performed the examination while he/she sat upright, had the elbow flexed 90 degrees with the arm close to the body, and had the wrist and forearm resting on a table. Prototype software that was designed from our laboratory was used to perform the examination. The user interface provides real-time visual feedback to the subject to perform motor movements consistent with the desirable waveform. The interface is developed in C# .NET and it communicates with the back-end server, which is a SQL database.

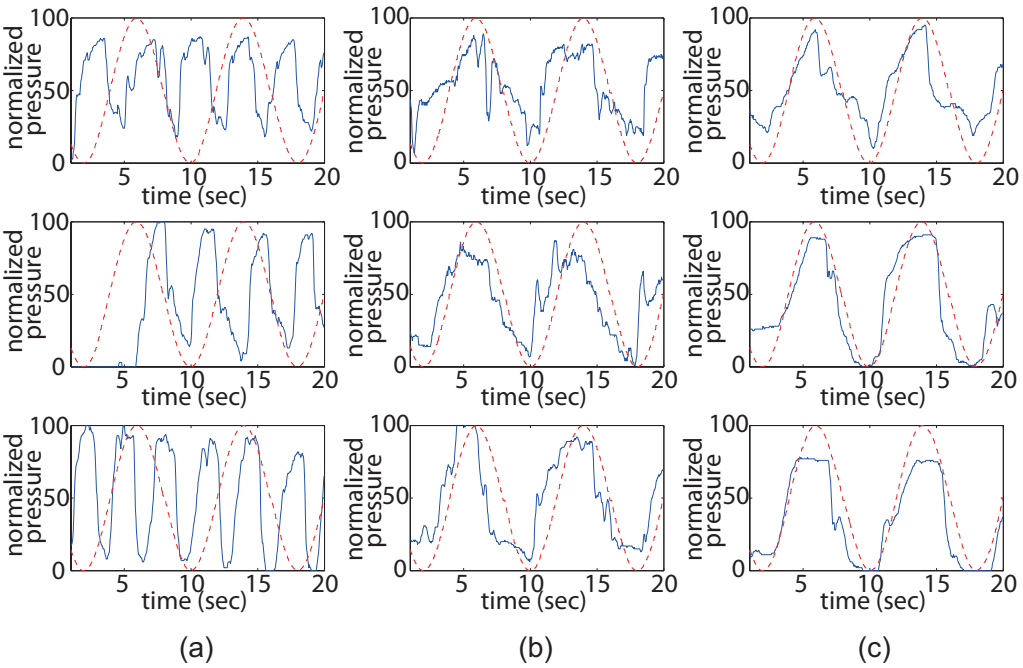

Figure 5.5: Randomly selected sample signals of subjects with various conditions. The three signals in column (a) belong to patients with CVA. The signals in column (b) belong to patients with CIDP. The signals in column (c) are sample signals of healthy subjects.

The 16 subjects that participated in the clinical study were grouped based on their primary medical problems. For instance, if a patient was actively diagnosed for CVA, that patient was assigned to the CVA group. This resulted in the formation of two primary ailment groups: (i) a group of patients with CVA and (ii) a group of patients with CIDP. These GOIs will be the focus of our experimental validation for the rest of this article.

These two groups are chosen such that each ailment group can be analyzed with sufficient data that can be used for both learning and validation. For example, the number of examination results for the patients with CVA was 17 and the number of results for patients with CIDP was 24. The rest of the patients were diagnosed with various ailments with upper limb deficits such as Parkinson's disease and Intra-Cerebral Hemorrhage but were eliminated from the analysis due to the relatively small number of signals (e.g., we had only one patient diagnosed with each of these ailments).

Fig. 5.5 illustrates sample examination results of the two ailment groups and the healthy subjects that we consider in this analysis. As illustrated, the examination results provide clear visual distinction between the health subjects and patients. CVA patients seem to exhibit delayed movements often having difficulties of coordinating the speed of the moving sinusoidal waveform. It may be a result of both physical and cognitive problems, which are common symptoms of CVA. The signals of CIDP have high level of noise compared to those of healthy subjects, which may be a result of a tremor effect. On the other hand, the signals of healthy subjects are smooth and well correlated to the target waveform.

5.6.2 Features used in the Analysis

This section presents features that are used in the data analysis. A total of 45 candidate feature functions are defined, where the first 36 feature functions are

in the time-domain and the following 9 feature functions are in the frequency-domain.

The first time-domain feature extraction function, denoted as f_1, represent MAD between the target waveform and the waveform generated by the subject. This function provides the overall level of performance in term of preciseness. f_2 computes the maximum instantaneous change in magnitude of the subject-generated waveform in order to investigate how well a subject manipulates the grip strength. f_3 computes the minimum time required for the subject-generated waveform to cross the target waveform from the time that the examination begins in order to investigate the subject's recovery time from deviation. The fourth time-domain function, f_4, computes the total number of intersections of two waveforms in order to investigate a subject's ability to control the grip strength to stay near the target waveform. f_5 investigates the number of changes in the sign of the slope of the patient-generated waveform in order to correlate the examination results to possible tremor effects. f_6 and f_7 compute the number of times that the subject-generated waveform crosses horizontal lines at magnitude y = 50% and y = 25%, respectively. The time-domain functions from f_8 to f_{22} are constructed as the following. 20 second-long waveforms generated by subjects are quantized into 15 segments of uniform length (i.e., each segment contains the data of 20/15 seconds). Then, f_8 to f_{22} contain the mean values of the magnitude of these segments. These quantized segments are used to evaluate the changes in grip strength over time. Furthermore, the time-domain functions from f_{23} to f_{36} compute the difference in mean magnitude of the two neighboring segments. These features are used to evaluate how fast the grip strength of a patient changes over time. The frequency-domain functions used in this analysis are computed as the following. The first frequency-domain function, f_{37}, computes the average difference in magnitude between the DFT of the target waveform and the DFT of the subject-generated waveform over all the possible frequency range. The frequency-domain functions

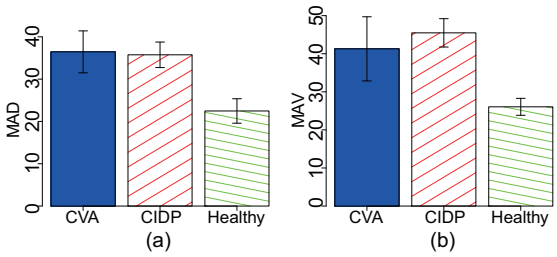

Figure 5.6: (a) MAD and (b) MAV for three groups: patients with CVA, patients with CIDP, and healthy subjects

from f_{38} to f_{45} divide the frequency range from 0 Hz to 16 Hz into 8 segments of uniform length (i.e., 2 Hz), and compute the spectrum energy for each segment in order to investigate the tremor effect at various frequency ranges.

5.6.3 Comprehensive Quantification of Motor Performance

This section presents the results of quantification of motor performance based on the metrics discussed in Section 5.5.1. The results of the two groups of major ailments are compared against the results of healthy subjects and they are illustrated in Fig. 5.6. The comparative results of the two ailment groups to the healthy group show that there exist significant degradations in motor performance in the ailment groups. These results show that the proposed system can be used for a preliminary screening device to quantify grip motor performance and compare the results against healthy subjects. This further implies that (i) the system may distinguish patients with upper limb deficits and (ii) the system provides a reference motor performance of subjects in a healthy physical status such that the improvement of a patient's motor capacity can be tracked.

5.6.4 Ailment-Specific Analysis Results

This section presents the data analysis results when the signals of patients with CVA and CIDP are compared against various groups of subjects: (i) all patients

without the ailment, (ii) healthy subjects, and (iii) the union of (i) and (ii) (i.e., all subjects without the ailment). For example, when the positive-class signals are defined as the signals of patients with CVA, the negative-class signals are defined as (i) signals of patients without CVA (i.e., excluding healthy subjects), (ii) signals of healthy subjects, and (iii) combined signals of patients without CVA and healthy subjects. In this work, we generically use the term *classification instances* for these three different positive and negative class combinations and label each instance as (i) a non-CVA instance, (ii) a healthy subject instance, and (iii) a combined instance.

5.6.4.1 CVA Detection

The data analytic results are summarized in Table 5.1. The sixth and the seventh columns represent the number of the selected features and their labels, respectively. The last column represents the top single feature with the highest accumulated ranking score (i.e., $K = 1$ in Section 5.5.2.3). The corresponding classification accuracies for the *non-CVA patients instance*, *healthy subjects instance*, and the *combined instance* are 94.64%, 95.00%, and 92.54%, respectively. These high classification accuracies show that the physiological signals produced by the proposed system contain ailment-specific information, which can be further interpreted as the motor characteristics of that ailment. For CVA patients, interestingly, the most frequent features for all three classification instances are identical (i.e., f_{36} and f_{39}). Fig. 5.7 illustrates the empirical distributions of the three features, which show clear separation between the positive (blue) and the negative (shaded red) classes.

Feature f_{36} represents the MAD of the last two temporal segments as was explained in the previous subsection, and the results in Fig. 5.7 show that CVA patients have relatively higher values compared to the rest of the subjects. Thus, it may help the physicians to see that the selected patients may dramatically lose the

Table 5.1: Summary of the experimental results when the positive class is defined as (i) CVA and (ii) CIDP.

Positive Class	Negative Class	Classification Accuracy	True Positive	True Negative	Size of most Frequent Subset	Most Frequent Subset Features	Top Ranked Feature
CVA	non-CVA Patients	0.9464	0.8889	0.9574	2	f_{36}, f_{39}	f_{36}
	Healthy subjects	0.9500	0.8889	1.0000	2	f_{39}, f_{36}	f_{36}
	Combined	0.9254	0.7778	0.9483	2	f_{36}, f_{39}	f_{36}
CIDP	non-CIDP Patients	0.9107	0.9167	0.9063	7	f_{32}, f_{14}, f_{13}, f_{24} f_{36}, f_{26}, f_{45}	f_{32}
	Healthy subjects	0.9142	1.0000	0.7200	6	f_{32}, f_{25}, f_{27} f_{11}, f_{2}, f_{19}	f_{32}
	Combined	0.8507	0.8333	0.8605	10	f_{32}, f_{13}, f_{27}, f_{36}, f_{24} f_{11}, f_{8}, f_{14}, f_{26}, f_{45}	f_{32}

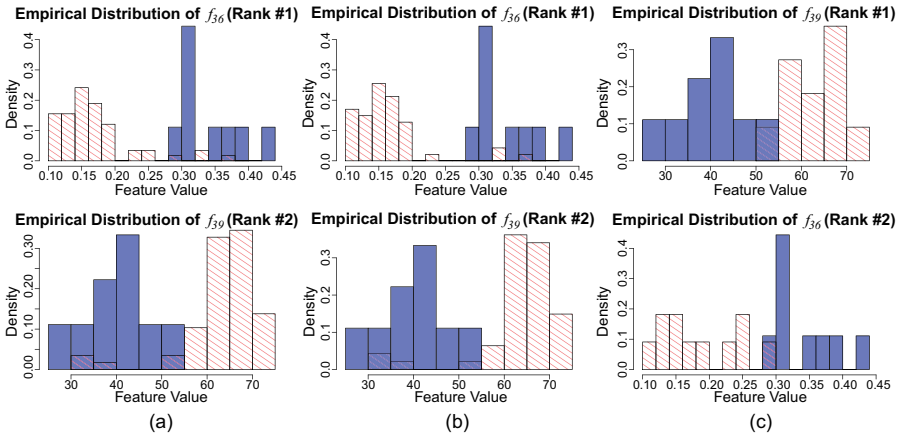

Figure 5.7: Empirical distribution of the selected features when the signals of CVA patients are compared against (a) combined group (f_{36} and f_{39}), (b) patients with CVA (f_{36} and f_{39}), and (c) healthy subjects (f_{39} and f_{36}).

preciseness of their grip control (or simply the grip strength) as the test proceeds towards the end. Moreover, feature f_{39} represents the spectrum energy of the patient response at the frequency range between 2 to 4 Hz, and Fig. 5.7 shows that the selected patients have relatively low spectral energy at this frequency range.

5.6.4.2 CIDP Detection

When the positive signals are defined as the signals of CIDP patients and compared against *non-CIDP patients instance*, *healthy subjects instance*, and the *combined instance*, the classification accuracies are 91.07%, 91.42%, and 85.07%, respectively. According to Table 5.1, a number of interesting observations are made on the results for CIDP patients.

First, feature f_{32} is found in all instances and always ranked the highest. It implies that this feature best represents the characteristic of the signals of CIDP, which is selected in the most frequent feature subsets regardless of the definition of the negative class.

Second, the most frequent feature set for the *combined instance* seems to be the union of the feature sets of the rest of the two instances. For example, f_{13} and f_{26} are found in the *patients without CIDP* instance and f_{27} and f_{11} are found in the *healthy subject* instance. This may result from the fact that the signals of the *combined instance* is the union of signals of the rest of the two instances.

f_{36} represents the difference in the average magnitude errors of the last two temporal segments, and according to our investigation, CIDP patients have relatively low value of f_{36}. This result shows that the selected patients can maintain the grip preciseness (or grip strength) until the very end of the examination as compared to the rest of the subjects. Moreover, f_{13} indicates the MAD between the target waveform and the patient's response at the 6^{th} temporal segment, where it contains a minimal point in sinusoidal waveform. The empirical distribution indicates that CIDP patients have relatively low error rates in that segment. More interestingly, f_{27}, which is selected as one of the top ranked features, shows that the difference in MAD between the 5^{th} and 6^{th} temporal segments is relatively high. These results indicate that CIDP patients lose their grip muscle control when the amplitude of the waveform starts to increase. Thus, patients may have trouble with the required grip strength.

5.6.5 Patient-Independent Ailment-Specific Analysis

Patient-independent classification involves validating the signals of patients that are excluded from those being used to train the classification model. (e.g., *learning* on CVA patient #1, #2, #3 and *validating* on CVA patient #4). This preliminary study is particularly important to investigate if the system can provide ailment specific information without previous history of a new patient. Furthermore, it demonstrates the robustness and independence of the proposed system to the data on which the system is initially constructed. In this experiment, four different classification instances are considered as shown in Table 5.2. For each classifica-

Table 5.2: Summary of the experimental results on patient independent ailment classification.

Positive Class	Negative Class	Classification Accuracy	True Positive	True Negative
CVA	Healthy	0.9167	0.9091	0.9231
Healthy	CVA	1.0000	1.0000	1.0000
CIDP	Healthy	0.7610	0.8276	0.5385
Healthy	CIDP	0.8542	0.8462	0.8276

tion instance, the classification accuracy is averaged among all *learning-validation* dataset combinations (i.e., leave-one-patient-out cross validation).

The selected features were different among different *learning-validating* set combinations within a classification instance. For the *CVA vs. Healthy Subject* instance, the number of features in the most frequent feature subsets for the four combinations were 2, 2, 3, and 3. In order to investigate the most frequent feature subsets of this patient-independent classification results, the following analysis has been performed.

First, the top ten ranked features from each *learning-validating* combination have been extracted. The reason that we observe the top ten features (out of 45 features) is to verify that relatively small number of motor characteristics can generalize the signals of the same ailment through patient independent analysis. Then the features that have appeared in all of those four top ten feature sets are observed. These features for CVA patients are $F_{ind} = [f_{39}\ f_{36}\ f_{43}\ f_{42}\ f_{38}\ f_{37}]$. In other words, the same six features are shared among the top ten ranked features when the classification is performed in a patient-independent way. This is a strong evidence that similar motor characteristics are observed among patients with CVA.

Next, the top ten ranked features are extracted from the results of the *CVA*

vs. Healthy Subject instance from Section 5.6.4 (i.e., the result shown in the third row of Table 5.1). The top ten ranked features were $F_{dep} = [f_{39}\ f_{36}\ f_{43}\ f_{42}\ f_{37}\ f_{46}\ f_{38}\ f_{45}\ f_{40}\ f_{16}]$. It can be easily observed that $F_{ind} \subset F_{dep}$, that is, similar features are observed across all the patients with CVA. Note also that features f_{39} and f_{36}, which were the selected features in Section 5.6.4, also appeared as the top two features for all *learning-validating* combinations in the patient independent classification.

Similar observations were made with CIDP patients.

5.7 Discussion and Future Work

Our main goal in this study was to demonstrate the efficiency of the presented handgrip device and its back-end data analysis for diagnosis of hand movement deficits. Given that clinical data collection is an expensive and time consuming process, we decided to validate our system based on a pilot clinical trial with limited patient population. For this, we performed an in-depth data analysis on the grip tracking signals collected from 16 subjects at St. Vincent Medical Center (Los Angeles, CA) and have achieved promising results. We believe that this proof-of-concept opens new routes for us to apply the proposed methods to a larger population.

With larger datasets, the system also can be configured to consider varying levels of granularity in the levels of motor disorders. For example, the NIH Storke Scale (NIHSS) is a widely used tool that classifies the level of impairment caused by a stroke [Hag11]. The proposed measurement model can be independently constructed on each of these classes, which allows for an accurate analysis of motor symptoms at different levels of severity.

The validation of the system to reflect the changes in motor performance before and after a medical treatment (e.g., operational surgery) is also an important issue

to be addressed. In an ongoing clinical trial, we are conducting a longitudinal study on a larger patient population through a collaboration with the UCLA Department of Neurosurgery, which includes patients with spinal cord injury in cervical region.

5.8 Conclusion

In this chapter, we introduce a portable handgrip device and its associated data analysis method, which together quantify hand-movement performance for patients with movement disorders. Two data analysis methods are discussed: a comprehensive metric for quantifying motor performance and an ailment-based comparative analysis. We showed that the comprehensive metric that quantifies motor performance can successfully distinguish patients with hand-motor deficits from healthy subjects. More importantly, the subset features that contribute the most to these classification instances are discussed in detail in order to provide intuitive analysis on ailment conditions. This study enables new opportunities for accurate quantification of an individual's ailment condition, disease severity, and specific physiological symptoms.

CHAPTER 6

Conclusion

This dissertation provided a comprehensive overview of target tracking tests based on a novel handgrip device. The contributions of this thesis can be summarized as the following.

1. Validation of reliability and validity of the target tracking test using machine learning algorithms.
2. Quantifying hyperexcited hand movements of CSM patients.
3. Introducing a method that predicts the postoperative functional outcomes based on the patients' preoperative information using Gaussian Process Regression.
4. Demonstrating the results of applying the target tracking tests on other neuromotor ailments such as CVA and CIDP.

The results reported in this dissertation show a great potential of the device to be used as an effective physical monitoring tool for CSM patients. The mean classification accuracy of the detection of motor dysfunction of CSM patients compared to control subjects was 78.9%. The system's correlation to the perceived level of motor dysfunction produced a mean error rate of 11.0% and its responsiveness achieved a maximum classification result of 88.2%.

The detection of hyperexcitation of hand muscle showed 99.5%, and its clinical significance was demonstrated by comparing the improvement (relaxation) of

hyperexcitation to the postoperative perceived motor deficits.

The prediction of the postoperative functional outcomes using Gaussian Process Regression showed 9.6% and 1.1% error rates for ODI and target tracking results, respectively.

The system also demonstrated its ability to detect (classify) other types of ailments that carry neuromotor deficits: 95.00% of detection accuracy for CVA and 91.42% for CIDP.

6.1 Future Work in Target Tracking Tests

All the data analyses performed in this dissertation are based on cross validations (or hold-out strategy as in Chapter 2 and Chapter 4) in order to minimize the chances of overfitting and to provide more fair results rather than optimistic results. However, the sample size considered in this work (maximum of 30 CSM patients and 30 control subjects) is still not sufficient to generalize the target patient population. This cohort study is an on-going study that will continue collecting data. The future work should include a larger sample size (e.g., 100 patients) to provide more generalized results. Furthermore, with a larger sample size, future work should consider a stratified (or clustered) data analysis based on demographic and clinical information such as the location of the disc herniation, age, ODI, and the handgrip score. This will provide a classification or regression model based on features that are specific to the motor characteristics of the subgroups, which would enhance the quantification of motor functions.

As discussed in Chapter 4, the handgrip device and the target tracking system have a great potential to be used for remote patient monitoring since the system is inexpensive, highly mobile, simple to use, and has relatively short testing duration. Human-computer interactive factors that will enhance patients' adherence to the system for longitudinal monitoring is an interesting future work. More

specifically, since CSM is a chronic ailment, human-computer interactive factors for elderly patients is necessary; gamification of medical devices (e.g., [FCZ11] is an interesting example.

6.2 Future Work in Degenerative Lumbar Spinal Disorder

The works presented in this paper only consider patients with cervical disorders with hand movement deficits. Degenerative spinal disorder is also common in lower back, which is also known as Lumbar Spinal Stenosis (LSS). One of the major complains of LSS patients is the impaired walking due to radiation of pain or neuromotor dysfunction. To address similar objectives (e.g., quantify the level of motor performance) for LSS patients, we have implemented a smart shoes equipped with (i) seven pressure sensors mounted on the insole, (ii) 3-axis accelerometer, and (iii) 3-axis gyroscope on the top of each foot. The sensor-equipped shoe is illustrated in Fig. 6.1. The smart shoes are also equipped with an embedded system with wireless transmitter that captures the sensory signals in real-time and sends them to the gateware (i.e., a laptop in our clinical setting). Various dimensions of gait performance can be analyzed, for example, (i) walking velocity, (ii) stride length, (iii) stride period, (iv) symmetry index, (v) distribution of weights on the foot, and many more. A clinical study has been conducted on approximately 16 patients at the time that this thesis is written, and is currently on-going. The quantification methodologies introduced in this paper can be similarly applied to the gait data. However, data pre-processing and parameter optimization need to be re-designed to the new feature dimensions.

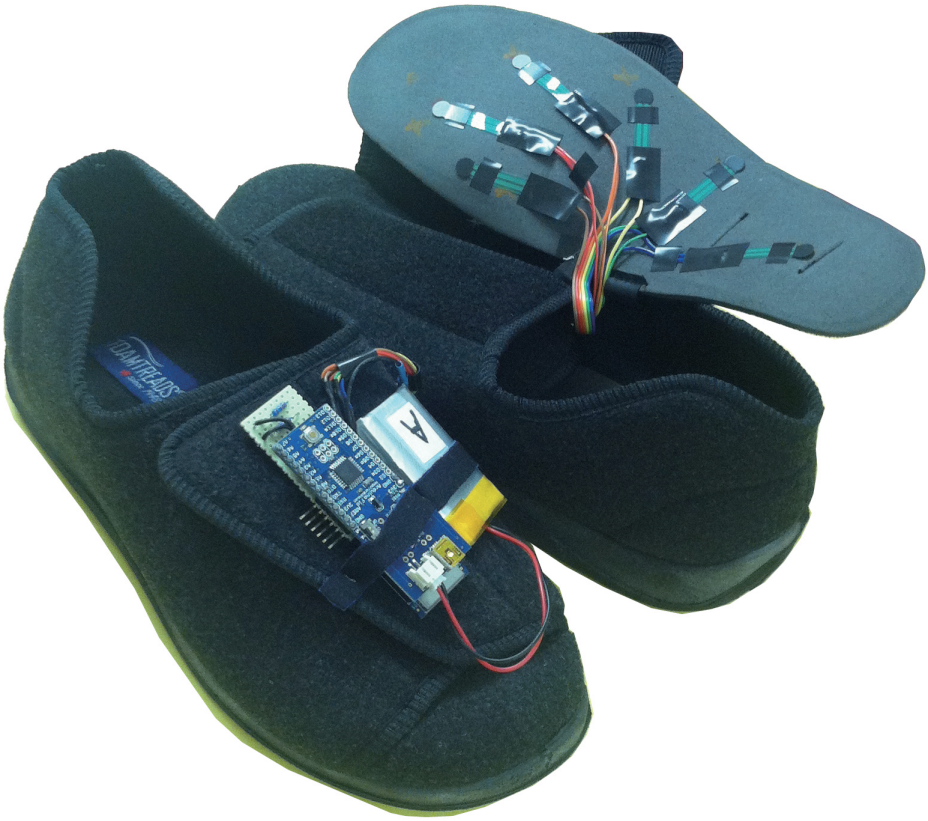

Figure 6.1: Picture of the smart shoe designed for lumbar spinal patients.

References

[AFM09] H. Al-Nashash, N.A. Fatoo, N.N. Mirza, R.I. Ahmed, G. Agrawal, N.V. Thakor, and A.H. All. "Spinal Cord Injury Detection and Monitoring Using Spectral Coherence." *Biomedical Engineering, IEEE Transactions on*, **56**(8):1971–1979, 2009.

[AKE98] E. M. Adang, G Kootstra, G. L. Engel, J. P. van Hooff, and H. L. Merckelbach. "Do retrospective and prospective quality of life assessments differ for pancreas-kidney transplant recipients?" *Transpl Int.*, **11**:11–15, 1998.

[AZD04] F. Alarcon, J.C. Zijlmans, G. Duenas, and N. Cevallos. "Post-stroke movement disorders: report of 56 patients." *J Neurol Neurosurg Psychiatry*, **75**(11):1568–1574, 2004.

[Azz12] Ahmed M. Azzam. "Effect of Hand Function Training on Improvement of Hand Grip Strength in Hemiplegic Cerebral Palsy in Children." *J Nov Physiother*, **2**, 2012.

[BLK91] E C Benzel, J Lancon, L Kesterson, and T Hadden. "Cervical laminectomy and dentate ligament section for cervical spondylotic myelopathy." *J Spinal Disord*, **4**(3):186–95, 1991.

[BNH12] H. L. Barden, M.T. Nott, R. Heard, C. Chapparo, and I. J. Baguley. "Clinical assessment of hand motor performance after acquired brain injury with dynamic computerized hand dynamometry: construct, concurrent, and predictive validity." *Arch Phys Med Rehabil.*, **93**:2257–2263, 2012.

[Bom00] C. Bombardier. "Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations." *Spine*, **25**(24):3100–3103, 2000.

[BS87] R. W. Bohannon and M. B. Smith. "Interrater reliability of a Modified Ashworth Scale of muscle spasticity." *Phys Ther*, **67**:206–207, 1987.

[BWB05] Ralph H.B. Benedict, Elizabeth Wahlig, Rohit Bakshi, Inna Fishman, Frederick Munschauer, Robert Zivadinov, and Bianca Weinstock-Guttman. "Predicting quality of life in multiple sclerosis: accounting for physical disability, fatigue, cognition, mood disorder, personality, and behavior change." *Journal of the Neurological Sciences*, **231**(12):29 – 34, 2005.

[BZ08] R. Bellazzi and B. Zupan. "Predictive data mining in clinical medicine: current issues and guidelines." *Int. J. Med. Inform.*, **77**(2):81–97, 2008.

[Con73] J.P. Conomy. "Disorders of body image after spinal cord injury." *Neurology*, **23**(8):842, 1973.

[Cri81] E.M. Critchley. "Speech disorders of Parkinsonism: a review." *J. Neurol. Neurosurg. & Psychiatry*, **44**(9):751–758, 1981.

[CWC11] Chad E. Cook, Mark Wilhelm, Amy E. Cook, Christopher Petrosino, and Robert Isaacs. "Clinical Tests for Screening and Diagnosis of Cervical Spine Myelopathy: A Systematic Review." *Journal of Manipulative and Physiological Therapeutics*, **34**(8):539 – 546, 2011.

[DC86] Richard A. Deyo and Robert M. Centor. "Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance." *Journal of Chronic Diseases*, **39**(11):897 – 906, 1986.

[Dij99] M. P. Dijkers. "Correlates of life satisfaction among persons with spinal cord injury." *Arch Phys Med Rehabil*, **80**(8):867–876, 1999.

[DM91] J J Denno and G R Meadows. "Early diagnosis of cervical spondylotic myelopathy. A useful clinical sign." *Spine*, **16**(12):1353–1355, 1991.

[DPa06] C.E. Dumont, M.R. Popovic, and et al. "Dynamic force-sharing in multi-digit task." *Clin Biomech*, **21**(2):138–146, Feburary 2006.

[dPP03] J de J Salazar-Torres, A. D. Pandyan, C. I. Price, R. I. Davidson, M. P. Barnes, and G. R. Johnson. "Biomechanical Characterization of the Stretch Reflex Activity as an Approach to Spasticity Measurement and Modeling-A pilot study." In *Proceedings of the 25' Annual lnternational Conference of the IEEE EMBS*, Cancun, Mexico, September 2003.

[DSH06] M Doita, H Sakai, T Harada, K Nishida, H Miyamoto, T Kaneko, and M Kurosaka. "Evaluation of impairment of hand function in patients with cervical myelopathy." *J Spinal Disord Tech*, **19**(4):276–280, 2006.

[ETK96] Wagih S. El Masry, Masahiko Tsubo, Shinsuke Katoh, Yasser H. S. El Miligui, and Ameen Khan. "Validation of the American Spinal Injury Association (ASIA) Motor Score and the National Acute Spinal Cord Injury Study (NASCIS) Motor Score." *Spine*, **21**(5):614–619, 1996.

[FCK07] Mitsuru Fukui, Kazuhiro Chiba, Mamoru Kawakami, Shinichi Kikuchi, Shinichi Konno, Masabumi Miyamoto, Atsushi Seichi, Tadashi Shimamura, Osamu Shirado, Toshihiko Taguchi, Kazuhisa Takahashi, Katsushi Takeshita, Toshikazu Tani, Yoshiaki Toyama, Kazuo Yonenobu, Eiji Wada, Takashi Tanaka, and Yoshio Hirota. "Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. Part 2. Verification of its reliability." *Journal of Orthopaedic Science*, **12**(6):526–532, 2007.

[FCZ11] N. Friedman, V. Chan, D. Zondervan, M. Bachman, and D.J. Reinkensmeyer. "MusicGlove: Motivating and quantifying hand movement rehabilitation by using functional grips to play music." In *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pp. 2359–2363, Aug 2011.

[Fin10] Joel A. Finkelstein. "Response Shift Following Surgery of the Lumbar Spine." *Thesis, University of Toronto*, 2010.

[FJH12] Michael G. Fehlings, Neilank K. Jha, Stephanie M. Hewson, Eric M. Massicotte, Branko Kopjar, and Sukhvinder Kalsi-Ryan. "Is surgery for cervical spondylotic myelopathy cost-effective? A cost-utility analysis based on data from the AOSpine North America prospective CSM study." *JNS: Spine Special Supplements*, **17**(1):89–93, 2012.

[FJO75] A.R. Fugl-Meyer, L. Jaasko, S. Olsson, and et al. "The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance." *Scand J Rehabil Med.*, **7**(1):13–31, 1975.

[FP00] J.C.T. Fairbank and Paul B. Pynsent. "The Oswestry Disability Index." *Spine*, **25**(22):2940–2953, 2000.

[FYK80] R.G. Feldman, R.R. Young, and W.P. Koella. *Spasticity: Disordered Motor Control*. Yearbook Medical Publishers, Chicago, USA, 1980.

[GDB02] 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." *Neurorehabil Neural Repair*, **16**(3):232–240, 2002.

[GE03] I. Guyon and A. Elisseeff. "An introduction to variable and feature selection." *Journal of Machine Learning Research*, **3**:1157–1182, March 2003.

[GHW93] M Gronblad, M Hupli, P Wennerstrand, E Jarvinen, A Lukinmaa, J P Kouri, and E O Karaharju. "Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients." *Clin J Pain*, **9**(3):189–195, 1993.

[GJ10] H. Ghasemzadeh and R. Jafari. "Body Sensor Networks for Baseball Swing Training: Coordination Analysis of Human Movements Using Motion Transcripts." In *Proc. IEEE Intl. Conf. PerCom 2010*, Mannheim, Germany, March 2010.

[GKW97] M Grevitt, R Khazim, J Webb, R Mulholland, and J. Shepperd. "The short form-36 health survey questionnaire in spine surgery." *J Bone Joint Surg Br*, **79**(1):48–52, 1997.

[GLa13] Ruth Getachew, Sunghoon Ivan Lee, and et. al. "Utilization of a Novel Digital Measurement Tool for Quantitative Assessment of Upper Extremity Motor Dexterity in Cervical Spondylotic Myelopathy." In *29th Annual Meeting of the AANS/CNS*, Phoenix, USA, March 2013.

[H 10] H. Ghasemzadeh and R. Jafari. "Coordination Analysis of Human Movements with Body Sensor Networks: A Signal Processing Model to Evaluate Baseball Swings." *IEEE Sensors Journal Special Issue on Cognitive Sensor Networks (SJ)*, **11**:603–610, 2010.

[Hag11] Ver Hage. "The NIH stroke scale: a window into neurological status." *Nurse.Com Nursing Spectrum (serial online)*, **24**:44–49, 2011.

[HBe05] R. Hillestad, J. Bigelow, and et al. "Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, and Costs." *Health Affairs*, **24**(5):1103–1117, 2005.

[HHa03] J. Hermsdorfer, E. Hagl, and et. al. "Grip force control during object manipulation in cerebral stroke." *Clin. Neurophys.*, **114**:915–929, 2003.

[HN92] M Haas and J Nyiendo. "Diagnostic utility of the McGill Pain Questionnaire and the Oswestry Disability Questionnaire for classification of low back syndromes." *J Manipulative Physiol Ther*, **15**:90–98, 1992.

[HRS12] Yiran Huang, Mahsan Rofouei, and Majid Sarrafzadeh. "Automated Wolf Motor Function Test (WMFT) for Upper Extremities Rehabilitation." In *The 9th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 91–96, London, UK, May 2012.

[HSL13] Thomas M.H. Hope, Mohamed L. Seghier, Alex P. Leff, and Cathy J. Price. "Predicting outcome and recovery after stroke with lesions extracted from MRI images." *NeuroImage: Clinical*, **2**(0):424 – 433, 2013.

[HT04] M Hochman and S Tuli. "Cervical Spondylotic Myelopathy: A Review." *The Internet Journal of Neurology*, **4**(1), 2004.

[HT08] E. Hudyma and G. Terlikowski. "Computer-aided detecting of early strokes and its evaluation on the base of CT images." In *IMCSIT*, pp. 251–254, Busan, South Korea, November 2008.

[HWL09] Yu-wei Hsieh, Ching-yi Wu, Keh-chung Lin, Ya-fen Chang, Chialing Chen, and Jung-sen Liu. "Responsiveness and Validity of Three Outcome Measures of Motor Function After Stroke Rehabilitation." *Stroke*, **40**(4):1386–1391, 2009.

[JJE06] R. Jafari, D.L. Jindrich, V.R. Edgerton, and M. Sarrafzadeh. "CMAS: Clinical Movement Assessment System for Neuromotor Disorders." In *IEEE BioCase 2006*, London, UK, November 2006.

[Jon00] R.D. Jones. *Measurement of sensory-motor control performance capacities: tracking tasks*. CRC Press, 2000.

[JRK09] Lene H Jakobsen, Ingeborg K Rask, and Jens Kondrup. "Validation of handgrip strength and endurance as a measure of physical function and quality of life in healthy subjects and patients." *Nutrition*, 2009.

[JXC13] M. Jadaliha, Yunfei Xu, Jongeun Choi, N.S. Johnson, and Weiming Li. "Gaussian Process Regression for Sensor Networks Under Localization Uncertainty." *Signal Processing, IEEE Transactions on*, **61**(2):223–237, 2013.

[KGG05] G. Kurillo, M. Gregoric, N. Goljar, and T. Bajd. "Grip force tracking system for assessment and rehabilitation of hand function." *J. Technology and Health Care*, **13**(3):137–149, 2005.

[KHM95] Gnter Kriz, Joachim Hermsdrfer, Christian Marquardt, and Norbert Mai. "Feedback-based training of grip force control in patients with brain damage." *Archives of Physical Medicine and Rehabilitation*, **76**(7):653 – 659, 1995.

[KKF97] Bruno Kopp, Annett Kunkel, Herta Flor, Thomas Platz, Ulrike Rose, Karl-Heinz Mauritz, Klaus Gresser, Karen L. McCulloch, and Edward Taub. "The arm motor ability test: Reliability, validity, and sensitivity to change of an instrument for assessing disabilities in activities of daily living." *Archives of Physical Medicine and Rehabilitation*, **78**(6):615 – 620, 1997.

[Kli10] Eric Klineberg. "Cervical Spondylotic Myelopathy: A Review of the Evidence." *Orthopedic Clinics of North America*, **41**(2):193 – 202, 2010.

[Kon95] G.V. Kondraske. *A working model for human system-task interfaces.* CRC Press, Boca Raton, FL, 1995.

[Kop00] J. A. Kopec. "Measuring functional outcomes in persons with back pain: a review of back-specific questionnaires." *Spine*, **15**:3110–3114, 2000.

[KWK96] Gert Kwakkel, Robert C. Wagenaar, Boudewijn J. Kollen, and Gustaaf J. Lankhorst. "Predicting Disability in StrokeA Critical Review of the Literature." *Age and Ageing*, **25**(6):479–489, 1996.

[KZB04] G. Kurillo, A. Zupan, and T. Bajd. "Force tracking system for the assessment of grip force control in patients with neuromuscular diseases." *Clinical Biomechanics*, **19**:1014–1021, 2004.

[LBF97] Richard Leclaire, Francois Blier, Luc Fortin, and Roland Proulx. "A cross-sectional study comparing the Oswestry and Roland-Morris functional disability scales in two populations of patients with low back pain of different levels of severity." *Spine*, **22**:68–71, 1997.

[LBK99] Peter S. Lum, Charles G. Burgar, Deborah E. Kenney, and H. F. Machiel Van der Loos. "Quantification of Force Abnormalities During Passive and Active-Assisted Upper-Limb Reaching Movements in Post-Stroke Hemiparesis." *IEEE Tran. on Biomed. Engineering*, **46**:652–662, 1999.

[LGa13a] Sunghoon Ivan Lee, Ruth Getachew, and et. al. "Utilization of a Novel Digital Device and an Analytic Method for Accurate Measurement of Upper Extremity Motor Function." In *81st AANS Annual Scientific Meeting*, New Orleans, USA, April 2013.

[LGa13b] Sunghoon Ivan Lee, H. Ghasemzadeh, and et. al. "Objective Assessment of Spastic Hand Hypertonia using a Novel Digital Device." In *The 20th IAGG World Congress of Gerontology and Geriatrics*, Seoul, Korea, June 2013.

[LGD98] Maurice H. Lipper, Jonas H. Goldstein, and Huy M. Do. "Brown-Sequard Syndrome of the Cervical Spinal Cord after Chiropractic Manipulation." *Am. J. Neuroradiol.*, **17**:1349–1352, 1998.

[LGM13a] S. I. Lee, H. Ghasemzadeh, B. J. Mortazavi, and M. Sarrafzadeh. "Pervasive Assessment of Motor Function: A Lightweight Grip Strength Tracking System." *Biomedical and Health Informatics, IEEE Journal of*, **17**(6):1023–1030, 2013.

[LGM13b] Sunghoon Ivan Lee, Hassan Ghasemzadeh, Bobak Jack Mortazavi, Andrew Yew, Ruth Getachew, Mehrdad Razaghy, Nima Ghalehsari, Brian H. Paak, Jordan H. Garst, Marie Espinal, Jon Kimball, Daniel C. Lu, and Majid Sarrafzadeh. "Objective assessment of overexcited hand movements using a lightweight sensory device." In *Body Sensor Networks (BSN), 2013 IEEE International Conference on*, May 2013.

[LHC02a] H. M. Lee, Y Z Huang, J-J. J. Chen, and I. S. Hwang. "Quantitative analysis of the velocity related pathophysiology of spasticity and rigidity in the elbow flexors." *J. Neurol. Neurosurg. Psychiatry*, **72**:621–629, 2002.

[LHC02b] H.M. Lee, Y.Z. Huang, J.J. Chen, and I.S. Hwang. "Quantitative analysis of the velocity related pathophysiology of spasticity and rigidity in the elbow flexors." *J. Neurol. Neurosurg. Psych.*, **72**:612–629, 2002.

[LLM05] H. K. Lim, D. C. Lee, W. B. McKay, M. M. Priebe, S. A. Homes, and A. M. Sherwood. "Neurophysiological assessment of lower-limb voluntary control in incomplete spinal cord injury." *Spinal Cord.*, **43**:283–290, 2005.

[LPH11] Christel M. van Leeuwen, Marcel W. Post, Trynke Hoekstra, Lucas H. van der Woude, Sonja de Groot, Govert J. Snoek, Dineke G. Mulder, and Eline Lindeman. "Trajectories in the Course of Life Satisfaction After Spinal Cord Injury: Identification and Predictors." *Arch Phys Med Rehabil*, **92**(2):207–213, 2011.

[LWN12] Sunghoon Ivan Lee, J. Woodbridge, A. Nahapetian, and M. Sarrafzadeh. "MARHS: Mobility Assessment System with Remote Healthcare Functionality for Movement Disorders." In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, Miami, USA, January 2012.

[Ma03] R. J. Marino and et al. "International standards for neurological classification of spinal cord injury." *J Spinal Cord Med*, **26**:s50–s56, 2003.

[MC97] W.D. Memberg and P.E. Crago. "Instrumented objects for quantitative evaluation of hand grasp." *J of Rehabilitation Research and Development*, **34**(1):82–90, January 1997.

[MI94] C.L. MacKenzie and T Iberall. *Advances in Psychology: The Grasping Hand*. Elsevier Science BV, Amsterdam, 1994.

[Mor00a] M.E. Morris. "Movement disorders in people with Parkinson's disease:a model for physical therapy." *Phys. Ther.*, **80**:578–597, 2000.

[Mor00b] M.E. Morris. "Movement Disorders in People With Parkinson Disease: A Model for Physical Therapy." *Physical Therapy*, **80**(6):578–597, June 2000.

[MSa00] I.S. Merkies, P.I. Samijn, and et. al. "Assessing grip strength in healthy individuals and patients with immune-mediated polyneuropathies." *Muscle Nerve*, **52**:1393–1401, 2000.

[MUC01] David M. Morris, Gitendra Uswatte, Jean E. Crago, Edwin W. Cook III, and Edward Taub. "The reliability of the Wolf Motor Function Test for assessing upper extremity function after stroke." *Archives of Physical Medicine and Rehabilitation*, **82**(6):750 – 755, 2001.

[NH05] D.A. Nowak and J. Hermsdorfer. "Grip force behavior during object manipulation in neurological disorders: toward an objective evaluation of manual performance deficits." *Mov Discord*, **20**(1):11–25, 2005.

[NM80] P.D. Neilson and M.D. "Influence of control-display compatibility on tracking behavior." *Quart. J. Exp. Psychol*, **32**:125–135, 1980.

[NNO93] P.D. Neilson, M.D. Neilson, and N.J O'Dwyer. "What limits high speed tracking performance?" *Hum. Mov. Sci.*, **12**:85–109, 1993.

[NPL11] S K Naik, C Patten, N Lodha, S A Coombes, and J H Cauraugh. "Force control deficits in chronic stroke: grip formation and release phases." *Exp Brain Res*, **211**(1):1–15, 2011.

[ON95] N.J O'Dwyer and P.D. Neilson. *Adaptation to a changed sensory-motor relation: immediate and delayed parametric modification.* Human Kinetics, Champaign, IL, 1995.

[PA00] D Postulart and E. M. Adang. "Response shift and adaptation in chronically ill patients." *Med Decis Making*, **20**:186–193, 2000.

[PBC10] S D Pradhan, B R Brewer, G E Carvell, P J Sparto, A Delitto, and Y. Matsuoka. "Assessment of fine motor control in individuals with Parkinson's disease using force tracking with a secondary cognitive task." *J Neurol Phys Ther*, **34**(1):32–40, 2010.

[PJF12] L.S. Prichep, A. Jacquin, J. Filipenko, S.G. Dastidar, S. Zabele, A. Vodencarevic, and N.S. Rothman. "Classification of Traumatic Brain Injury Severity Using Informed Data Reduction in a Series of Binary Classifier Algorithms." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, **20**(6):806–822, 2012.

[PMA13] L Palmerini, S Mellone, G Avanzolini, , F Valzania, and L Chiari. "Quantification of Motor Impairment in Parkinson's Disease Using an Instrumented Timed Up and Go Test." *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, **21**(4):664–673, 2013.

[PMD00] F. Pisano, G. Miscio, C. Del Conte, D. Pianca, E. Candeloro, and R. Colombo. "Quantitative measures of spasticity in post-stroke patients." *Clin Neurophysiol*, **111**:1015–1022, 2000.

[Pou74] E.C. Poulton. *Tracking skill and manual control.* Academic Press, New York, 1974.

[PRM11] L Palmerini, L Rocchi, S Mellone, F Valzania, and L Chiari. "Feature Selection for Accelerometer-Based Posture Analysis in Parkinson's Disease." *Information Technology in Biomedicine, IEEE Transactions on*, **15**(3):481–490, 2011.

[PWA98] Marcel W.M. Post, Luc P. de Witte, Floris W.A. van Asbeck, Alphons J. van Dijk, and August J.P. Schrijvers. "Predictors of health status and life satisfaction in spinal cord injury." *Archives of Physical Medicine and Rehabilitation*, **79**(4):395 – 401, 1998.

[Ros06] B.A. Rosner. *Fundamentals of Biostatistics*. Thomson-Brooks/Cole, 2006.

[RPa06] Preeti Raghavan, Electra Petra1, and et al. "Patterns of impairment in digit independence after subcortical stroke." *J Neurophysiol*, **95**(1):269–278, 2006.

[RW06] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[SC01] Anoushka Singh and A. Crockard. "Comparison of seven different scales used to quantify severity of cervical spondylotic myelopathy and post-operative improvement." *J Outcome Meas.*, **5**:798–818, 2001.

[SD07] S. A. Sisto and T. Dyson-Hudson. "Dynamometry testing in spinal cord injury." *J Rehabil Res Dev*, **44**(1):123–136, 2007.

[SF01] Lali Sekhon and Michael G. Fehlings. "Epidemiology, Demographics, and Pathophysiology of Acute Spinal Cord Injury." *Spine*, **26**(24S):S2–S12, 2001.

[SFM08] O. Stegle, S.V. Fallert, D. J C MacKay, and S. Brage. "Gaussian Process Robust Regression for Noisy Heart Rate Data." *Biomedical Engineering, IEEE Transactions on*, **55**(9):2143–2151, 2008.

[SGC05] A. Singh, K. Gnanalingham, A. Casey, and A. Crockard. "Use of quantitative assessment scales in cervical spondylotic myelopathy: survey of clinician's attitudes." *Acta Neurochir (Wien)*, **147**:1235–1238, 2005.

[SM03] A.H. Sung and S. Mukkamala. "Identifying important features for intrusion detection using support vector machines and neural networks." In *Symposium on Applications and the Internet*, pp. 209–216, 2003.

[SMM11] Sze-Ee Soh, Meg E. Morris, and Jennifer L. McGinley. "Determinants of health-related quality of life in Parkinsons disease: A systematic review." *Parkinsonism & Related Disorders*, **17**(1):1 – 9, 2011.

[SN00] W.E. Sharp and K.M Newel. "Coordination of grip configurations as a function of force output." *J. Motor Behavior*, **32**:73–82, 2000.

[SS98] C. Spence and P. Sajda. "The role of feature selection in building pattern recognizers for computer-aided diagnosis." In *Proceedings of SPIE*, volume 3338, pp. 1434–1441, June 1998.

[STa89] A. Sunderland, D. Tinson, and et. al. "Arm function after stroke. An evaluation of grip strength as a measure of recovery and a prognostic indicator." *J. Neurol. Neurosurg. & Psychiatry*, **52**:1267–1272, 1989.

[STC13] Anoushka Singh, Lindsay Tetreault, Adrian Casey, Rodney Laing, Patrick Statham, and Michael G. Fehlings. "A summary of assessment tools for patients suffering from cervical spondylotic myelopathy: a systematic review on validity, reliability and responsiveness." *European Spine Journal*, pp. 1–20, 2013.

[STH95] P. Sanderson, B. Todd, G. Hold, and et. al. "Compensation, work status, disability in low back pain patients." *Spine*, **20**:554–556, 1995.

[TB10] J. A. Tracy and J. D. Bartleson. "Cervical Spondylotic Myelopathy." *The Neurologist*, **16**, 2010.

[THC08] M Tonack, S L Hitzig, B C Craven, K A Campbell, K A Boschen, and C F McGillivray. "Predicting life satisfaction after spinal cord injury in a Canadian sample." *Spinal Cord*, **46**:380–385, 2008.

[TM94] D.C. Turk and D.A. Marcus. "Assessment of chronic pain patients." *Semin. Neurol.*, **14**:206–212, 1994.

[TS96] M. A. Testa and D. C. Simonson. "Assessment of quality-of-life outcomes." *N Engl J Med*, **334**(13):835–840, 1996.

[VFa00] F. Vecchi, C. Freschi, and et. al. "Experimental evaluation of two commercial force sensors for applications in biomechanics and motor control." In *Proc. IFESS Conf.*, Aalbor, DK, June 2000.

[VKW11] Janne M. Veerbeek, Gert Kwakkel, Erwin E.H. van Wegen, Johannes C.F. Ket, and Martijn W. Heymans. "Early Prediction of Outcome of Activities of Daily Living After Stroke: A Systematic Review." *Stroke*, **42**(5):1482–1488, 2011.

[WKa05] R. Wenzelburger, F. Kopper, and et. al. "Hand coordination following capsular stroke." *Brain*, **128**:64–74, 2005.

[WL81] C. Waggoner and R. LeLieuvre. "A method to increase compliance to exercise regimens in rheumatoid arthritis patients." *Journal of Behavioral Medicine*, **4**(2):191–201, June 1981.

[WPM10] E. Wade, A. R. Parnandi, and M. J. Mataric. "Automated administration of the Wolf Motor Function Test for post-stroke assessment." In *2010 4th International Conference Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, Munchen, Germany, March 2010.

[WW04] Sandy McCombe Waller and Jill Whitall. "Fine motor control in adults with and without chronic hemiparesis: baseline comparison to nondisabled adults and effects of bilateral arm training." *Archives of Physical Medicine and Rehabilitation*, **85**(7):1076 – 1083, 2004.

[XCO11] Yunfei Xu, Jongeun Choi, and Songhwai Oh. "Mobile Sensor Network Navigation Using Gaussian Processes With Truncated Observations." *Robotics, IEEE Transactions on*, **27**(6):1118–1131, 2011.

[You00] William F. Young. "Cervical Spondylotic Myelopathy: A Common Cause of Spinal Cord Dysfunction in Older Persons." *Am Fam Physician*, **62**:1064–1070, 2000.