Quantitative Estimation of Movement Progress during Rehabilitation after Knee/Hip Replacement Surgery

by

Roshanak Houmanfar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2014

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Mobility improvement for patients is one of the primary concerns of physiotherapy rehabilitation. In a typical physiotherapy session, the patient is instructed to perform multiple exercises, based on a specific regimen recommended by the physiotherapist for each patient. The physiotherapist then evaluates the patient's progress based on his or her performance during the exercises. Providing the physiotherapist and the patient with a quantified and objective measure of progress, based on both individual exercises and the exercise set, can be beneficial for monitoring the patient's performance. The quantified measure can also be beneficial when the physiotherapist is not available, e.g., crowded gym or rehabilitation at home.

In this thesis, two approaches are introduced for quantifying patient performance. One approach describes the movement timeseries by statistical measures and the other by a stochastic model. Both approaches formulate a distance between patient data and the healthy population as the measure of performance. Distance measures are defined to capture the performance of one repetition of an exercise or multiple repetitions of the same exercise. To capture patient progress across multiple exercises, a quality measure and overall score are formulated based on the distance measures and are used to quantify the overall performance for each session.

The proposed approaches are compared to several existing approaches, including sample distribution approaches (two sample kernel), classifier-based approaches (Naive Bayes, Support Vector Machines, and Kullback-Leibler Divergence), and dynamical movement primitives. In their original formulation, existing approaches are not capable of estimating measures of performance for multiple exercises. Therefore, the measures of performance for multiple repetitions of the same exercise are estimated using the existing approaches, while the formulation proposed in this thesis is used to estimate the overall performance for multiple exercises in one session.

The effects of different variabilities in human motion on the performance of the proposed approaches and the comparison approaches are investigated with both synthetic and patient data. The patient data consists of rehabilitation data recorded from patients recovering from knee or hip replacement surgery, the associated exercise regimen and physiotherapist evaluations of progress. The methods are evaluated quantitatively based on correlation between methods, correlation with exercise regimen difficulty, and qualitatively based on the patients' medical charts. The proposed approaches are capable of capturing the trend of progress for the synthetic dataset and are superior to the existing approaches in presence of multiple sources of variability. For patient data, the proposed approaches correlate moderately with the score obtained from the exercise regimen, and qualitatively

correspond with the patients' medical charts. The results indicate that the quantified measures of progress obtained from the proposed approaches are promising tools for supporting physiotherapy practice through monitoring patient progress.

Acknowledgements

I would like to express the deepest gratitude and appreciation for my supervisor, professor Dana Kulić, for her invaluable guidance and support through out this research. I am very grateful for her expertise, patience, time and informative discussions during meetings, and countless paper and report edits.

I am also very thankful to Doctor Michelle Karg, for her input, hours of meeting and discussion, as well as limitless paper and report edits.

I would like to thank everyone in the Automatic Rehabilitation System group for their support, and assistance throughout this research. I would also like to thank everyone in the Automatic Systems Lab for sharing their insight during group meetings and for being all-around great people who have made this experience much more enjoyable.

I would like to thank the University of Waterloo for providing funding without which this research work would not have been possible.

Lastly, I would like to thank my husband, parents, and close friends, for their unyielding support and love.

Dedication

To my husband, Behzad, for always supporting me, encouraging me, believing in me, and above all for being my strength in the journey of life. To my mother, Shahin, for her unconditional love and support, for never giving up on me, and for every decision and sacrifice she has made to get me where I am today. To my father, Iraj, for his unwavering encouragement. To my sisters, Ghazal and Mahsa, for the joy of the true friendship they have brought into my life. To my brothers, Hamed and Saeed, for being the incredible friends and individuals they are.

Table of Contents

List of Tables

List of Figures

Nomenclature

Abbreviations

HMM	Hidden Markov Model
IMU	Inertial Measurement Unit
EKF	Extended Kalman Filter
KL	Kullback-Leibler
NB	Naive Bayes
DoF	Degree of Freedom
SVM	Support Vector Machine
DMP	Dynamical Movement Primitives
3DGA	Three Dimensional Gait Analysis
BN	Bayesian Network
LASSO	Least Absolute Shrinkage and Selection Operator
KEFO	Knee Extension Flexion
KHEF	Knee Hip Extension Flexion

Variables

A	HMM state transition probability matrix
b_i	HMM observation probability
π	HMM initial state distribution
μ	Mean of the output distribution in an HMM
Σ	Covariance of the output distribution in an HMM
Q	HMM state sequence
O	HMM observation sequence
λ	An HMM model
$P(x\|y)$	Probability of x given y
$E(x\|y)$	Expectation of x given y
$D(\lambda_1, \lambda_2)$	Kullback-Leibler Divergence between model λ_1 and λ_2
v_k	kth feature vector
L_k	Output value of the kth feature vector
$\psi(.,.)$	Kernel function
α_k	Constant value
b	Bias value in SVM
ξ_i	Slack variable which controls the allowable misclassification of the data
w	The vector that defines the halfspace constructed by SVM
V	Feature Matrix where each column is a feature and each row is a trial
Y	Output, class label
τ	Time constant
α_z, β_z	Positive constants that determine the time response of a dynamical system
Φ	Gaussian like periodic basis functions
r	Amplitude
A	Desired amplitude
α_r	Constant
w_i	Weight of the ith basis function
c_i	Center of the ith basis function
$k(.,.)$	Kernel function
P	Sample distribution of the patient population
R	Sample distribution of the healthy population
m	Number of samples from distribution P
n	Number of samples from distribution R
MMD	Measure of difference between sample distribution
p_i	ith sample from distribution P
r_i	ith sample from distribution R

$w_1 \dots w_k$	weights of features in the LASSO technique
f_i	ith feature
$\hat{y}$	Regression estimation by LASSO
y	Output value
q	Joint angle position
$\dot{q}$	Joint angle velocity
$\ddot{q}$	Joint angle acceleration
ω	A *repetition timeseries*
γ	Joint kinematic at each time step
T	Duration of the timeseries
Ω	One exercise
Γ	Exercise set
v	Feature vector
μ_q	Mean of joint angle
min_q	Minimum of joint angle
$skew_q$	Skew of joint angle
max_q	Max of joint angle
rom_1	Range of motion of the joint angle
$V_H^{'}$	Healthy feature vector with top features
$V_P^{'}$	Patient feature vector with top features
μ_H	Mean of the top features in the healthy population
Σ_H	Standard deviation of the top features in the healthy population
δ_i	Measure of performance for a repetition timeseries
Δ_ω	Measure of performance for a repetition set
δ_{NB}	Measure of performance for a repetition timeseries using Naive Bayes
δ_{KLD}	Measure of performance for a repetition timeseries using Kullback-Leibler distance
v_p	Patient feature vector
v_H	Healthy feature vector
$\mu_{B_{ij}}$	Mean of each bin in the healthy population
$\gamma_{B_{ij}}$	Repetition timeseries in the ith bin of the jth healthy person
$\mu_{average_{B_i}}$	Average over all bins in the healthy population
δ_{k_ω}	The measure of performance for the kth subgroup in the kernel-based approach
δ_ω	Measure of performance for a repetition timeseries in the kernel-based approach
δ_{H_j}	Measure of performance for the healthy population
$\mu_{\delta_{H_j}}$	Mean of the performance in the healthy population
$\sigma_{\delta_{H_j}}$	Standard deviation of the performance in the healthy population
Q_j	Quality measure for exercise j

a	The index that penalized the quality measure
S	Overall score for an exercise set
n_ω	Number of repetitions for an exercise
$\xi_{i_{method}}$	Overall performance of approach *method* for session i
Ξ_{method}	The vector of score averages over all sessions
p_H	Probability that an exercise belongs to the hard exercise set
p_E	Probability that an exercise belongs to the easy exercise set
S_{GT}	The measure of performance obtained from the exercise regimen

Chapter 1

Introduction

The application of machine learning techniques to human motion analysis has grown rapidly over the past few years. Machine learning for human motion analysis has been used for gesture recognition during human-machine interaction [36], action recognition [1], and sport science [66]. A particularly promising application is in the field of rehabilitation and physiotherapy. Measurement and analysis of physiotherapy data have the potential to provide an objective and quantitative measure of patient progress leading to improvement in physiotherapy treatment. A short version of this work has appeared in [97].

During a typical physiotherapy session, the physiotherapist instructs the patient to perform a number of exercises, each with several repetitions. The set of exercises chosen and the number of repetitions may be customized for each patient. The physiotherapist then evaluates the patient's progress based on their performance. In current clinical practice, the patient's performance is typically assessed using visual observation of the patient's motions, questionnaires, and goniometry. Questionnaires such as the Falls Efficacy Scale [142] are used to determine the patient's confidence in performing a task. Activity based tests such as the Community Balance and Mobility Scale [45] are used to assign a score to the patient's motion capabilities.

Goniometry is a technique of measuring joint angles which isolates a single body joint in order to evaluate a subject's range of motion [87]. While goniometry can be effective when measuring the motion of a single joint, it is a technique of measuring joint angles which isolates a single body joint in order to evaluate a subject's range of motion [87]. Goniometry is not accurate when determining the range of motion for multiple degrees of freedom during complex 3D motions.

The current measurement and assessment techniques require additional physiotherapist effort and monitoring, and are of limited utility when measuring complex whole body move-

ments. Automation of patient observation would support physiotherapy practice through automated assessment and evaluation of exercise performance and more accurate measurement of complex multi-joint movements. An automated system could also provide the therapist with numerical metrics to assess the patient's recovery process and potentially allow physiotherapists to assess the effectiveness of various treatment protocols over a population of patients.

Patient data analysis for progress monitoring is a challenging task because of the complexity of human motion. In particular, there are a number of challenges that need to be addressed: (1) high dimensionality of the data, (2) spatial and temporal variability of movement, and (3) variability of the patient population and the associated treatment regimens. Human movement consists of synchronous recruitment of multiple degrees of freedom (DoF), making single DoF comparisons (e.g., only comparing the range of motion in one joint) incomplete and possibly unreliable.

Human motion has significant temporal and spatial variability for different repetitions of the same exercise. Since humans differ in their characteristics, such as age, gender, height, and weight, variability between different subjects is also observed. When recovering from an illness or surgery there are variabilities caused by progress and improvement through rehabilitation, differing levels of pain during the course of treatment, as well as differing levels of fatigue over the course of a session. During the course of rehabilitation, patients frequently are observed to exhibit *compensation*. *Compensation* is a term that refers to recruitment of additional or different degrees of freedom [72] while performing a certain exercise. Other sources of variability include the measurement system and the algorithms used for deriving the joint angles, such as integration drift, initial sensor positioning, and sensor position change.

The purpose of progress monitoring is isolate the changes in movement associated with recovery[1] and improvement[2]. The presence of multiple other sources of variability makes this task challenging. Furthermore, the exercises are performed based on a specific regimen instructed by the physiotherapist for each patient. The regimen differs from patient to patient depending on their health status, age, and special conditions. Therefore the same exercises are not performed by an individual patient over the course of treatment, nor is the same sequence of exercises used across different patients. The proposed approach should be flexible to detect changes in the patient's performance for any set of exercises.

In this thesis, a previously developed body-worn sensor system and associated algo-

[1]Changes in patient performance which are due to the healing of the surgical wound, reduction of swelling, and reduction of the pain level.

[2]Changes in patient performance which are due to improvement in range of motion, improvement in muscle control, and improvement in speed of motion.

rithms for measuring human movement during rehabilitation is used. The system estimates the human pose at each measurement instance in terms of joint angles [76], and segments individual exercise repetitions from the continuous time series data [77]. The overall system is illustrated in Fig. 1.1. The data is collected from body worn inertial measurement unit (IMU) sensors attached to the patient and the joint angle position, velocity, and acceleration are derived. The joint angle position, velocity, and acceleration data are then segmented such that each segment begins with the start of an exercise repetition and ends when the exercise repetition is finished. In this thesis, we focus on the last two steps depicted in Fig. 1.1, i.e., extracting informative features and developing measures of performance.

1.1 Contributions

This thesis addresses the following challenges in the assessment of exercise performance during physiotherapy:

1- How to assess the patient performance for one repetition of an exercise.

2- How to assess the patient performance for multiple repetitions of the same exercise.

3- How to assess the patient performance for multiple repetitions of different exercises.

We propose two approaches for progress estimation based on the segmented motion data, based on a comparison between patient movement and healthy population movement.

Feature-based method for individual exercise and exercise set assessment: In the first approach, descriptive features are extracted from joint angle positions, velocities, and accelerations. The feature extraction method proposed in this approach is motivated by the biomechanics literature [117] [146]. For each exercise, informative features are selected from the feature set. The measure of performance for a *repetition timeseries* is computed using the distance between the feature vector of the patient and the mean of the healthy population in the feature space. The measure of performance for a *repetition set* is calculated by computing the average of the measures of performance of the *repetition timeseries*. The measure of performance for an *exercise set* is calculated by calculating the weighted average of the measures of performance of the different *repetition sets*.

Stochastic model based method for individual exercise set assessment: In the second approach, the descriptive features are extracted from the HMM of the joint angle positions, velocities, and accelerations. The feature extraction method proposed in the second approach provides a model of the timeseries, which is more common in the machine learning literature [67] [71] [3]. For each exercise, the informative features are selected from the feature set. The measure of performance for a *repetition set* is determined using the distance between the feature vector of the patient and the mean of the healthy population in the feature space. The measure of performance for an *exercise set* is calculated by computing the weighted average of the measures of performance of the *repetition sets*.

Progress estimation from multiple exercises: The distance measures obtained using each proposed approach are utilized to quantify the performance quality of a single repetition of an exercise, a set of repetitions of the same exercise by comparing the performance of patients with the performance of the healthy population. The overall measure of performance for a set of different exercises is obtained by computing the weighted average of the measures of performance of the repetition sets in the exercise set, while considering the number of repetitions for each exercise and exercise difficulty. These measures are used to provide the patients and the physiotherapists with feedback on the patient's performance during rehabilitation.

Systematic comparison and evaluation of proposed and existing methods on synthetic data: The proposed approaches are compared to several existing approaches, including sample distribution approach, classifier based approaches, and the DMP-based approach. The sample distribution approach is based on the similarity of the sample distributions for two timeseries, and is also a common approach in the machine learning literature [39]. This method enables the detection of the small perturbations in human motion and is therefore capable of capturing differences in the smoothness of the motion. The classifier-based approach uses the same statistical features as the first proposed approach. In this approach, a classifier is trained to separate the healthy population from the patients and the classification criterion is used as the measure of performance. The DMP-based approach uses the features extracted from the DMP models of the joint angle positions, velocities, and accelerations, and uses the distance between the features of the healthy population with those of the patient population as a measure of progress.

The proposed approaches and the existing approaches are evaluated using a synthetic dataset. The synthetic dataset is used to systematically investigate the effects of the variabilities typically seen in human motion on the performance of the proposed approaches and the existing approaches.

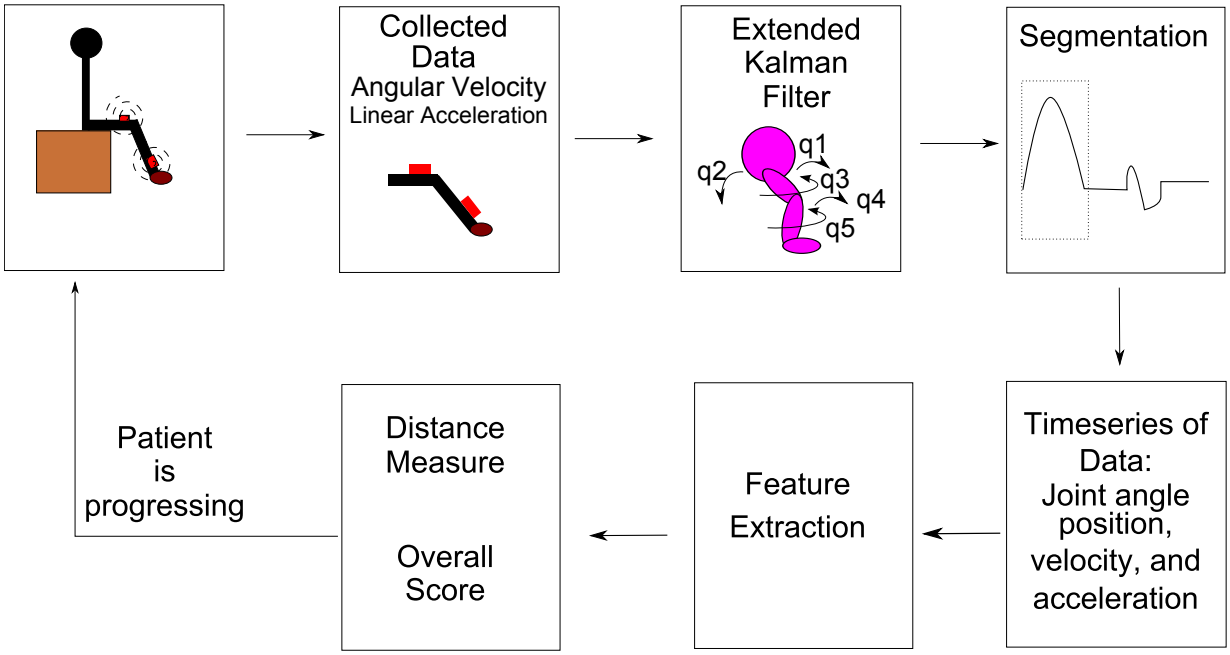


Figure 1.1: An overview of the complete measurement and analysis system. The IMU sensors are mounted on the patient's knee and ankle. Angular velocity and linear acceleration are collected from the sensors and the five joint angles are estimated using the extended Kalman filter [76] (q_1: extension/flexion of the hip, q_2: internal/external rotation of the hip, q_3: abduction/adduction of the hip, q_4: extension/flexion of the knee, q_5: internal/external rotation of the lower limb). The data is then segmented [77] so that each segment starts with the beginning of one *repetition timeseries* and ends when the repetition is completed. Features are extracted from the joint angles' segmented timeseries and are used to obtain the measures of progress.

Validation of proposed method on patient dataset: The clinical dataset is used to assess the performance of the approaches for real patient data. The performance of both proposed approaches and the existing approaches are evaluated on the real patient data. The physiotherapist assessment can be implicitly obtained from patients exercise regimen. The result of each approach is compared to the score obtained from the exercise regimen. Furthermore, the results of the proposed approaches are compared to patient's medical chart.

1.2 Outline

This thesis is organized as follows: Chapter 2 overviews the related work and motivates the application of continuous measures in physiotherapy rehabilitation. The current state of the art mostly consists of works that either use a classifier for distinguishing two populations based on their motion, or use classification criteria as a measure of similarity between two populations. Very few existing methods consider the gradual changes occurring in patient performance and none of the works are extended to include exercise sets with more than one exercise type. Furthermore, many of the related works are not validated on a patient dataset.

Chapter 3 overviews the mathematical background on machine learning techniques that are used in Chapters 4-6. The Naive Bayes, Support Vector Machines, and Kullback-Leibler divergence classification approaches are described in this chapter and are later used for comparison with the proposed approaches.

Chapter 4 presents the proposed approaches for estimating patient progress during physiotherapy rehabilitation. The proposed approaches consist of the feature-based approach, and the HMM-based approach for assessing individual exercises, as well as an approach for assessing progress based on a set of multiple exercises. Furthermore, this chapter provides the formulation for the comparison approaches, including the kernel-based approach, the DMP-based approach , and the classifier-based approach.

Chapter 5 presents the results of applying each of the proposed approaches on a synthetic dataset. The dataset is designed to simulate progress in patients' range of motion and speed differences. The data set is then subjected to the different types of possible variabilities similar to those observed in human motion, and the results of each approach under each enforced condition are evaluated.

Chapter 6 presents the results of applying each of the proposed approaches on a real patient dataset. The dataset contains the motion data for 18 patients undergoing rehabilitation after a knee or a hip replacement surgery. The results of each approach are compared to the qualitative assessment of the patients' motion, which are obtained from medical charts, and a quantitative estimate of performance based on the exercise regimen difficulty instructed by the physiotherapist during rehabilitation.

Lastly, Chapter 7 discusses the properties of the proposed approaches, summarizes the key contributions of the thesis, and outlines directions for the future work.

Chapter 2

Related Work

This chapter provides an overview of the related work on human motion analysis and its application to rehabilitation. In Sec. 2.1, different applications of human movement analysis are introduced, and examples of techniques utilized by each field of application are provided. Sec. 2.2 describes the current state of the art for human motion analysis in rehabilitation. The methodology used by each publication in this category is introduced and a comparison between different approaches is provided.

2.1 Human Motion Analysis

Human movement analysis is an active area of research with a wide range of applications including action recognition [1], gait recognition [6], gait identification [93] [32] [54], gesture recognition [36], motion imitation in robotics [84] [67], affective human computer interaction [48], sport science [66], medical diagnosis [125], and rehabilitation [2].

Most works in the action recognition category focus on correctly labelling a motion using a classifier [96] [7] [38] [49] [81]. A common approach in this application field is to train classifiers based on a known motion set and then use the classifiers to label motions performed by the demonstrator [124] [7] [50]. For example, *Udin et al.* [124] assume that the demonstrator performs a motion from a specific set of motions. They train HMM models for each motion and for every new motion performed by the demonstrator. They compute the probability of the sequence belonging to each model, and choose the most probable model as the label.

The gait recognition category focuses on labelling different types of gait, such as healthy and elderly gait patterns [6]. In most cases, features are extracted from the collected

timeseries, they are then used in training classifiers for a specific set of gait motions. The classifiers are then used to label different types of gait [6] [31] [136].

In gait identification applications, the objective is either to distinguish between different individuals based on gait data collected from each subject (between subject differences) or to detect differences between multiple trials of one subject's gait patterns (within subject changes). For example, *Kale et al.* [54] use the width of the outer contour of the silhouette of a walking person as the image features. A set of key stances during the walking cycle are chosen, and the Euclidean distance of a given image from this stance set is considered as the set of the lower dimensional features. A continuous Hidden Markov Model is trained using several such lower dimensional vectors, and is used for gait identification.

Making human-computer interaction more intuitive is the main objective of affective human computer interaction. One of the means to facilitate this interaction is for the computer to detect different poses and gestures performed by the user [48].

In the gesture recognition field, the objective is to assign labels to human postures and gestures. *Camurri et al.* [17] detect expressive gestures by analysing motion data collected from participants. They introduce new features that correspond to tempo changes, underlying rhythm of the motion, and contraction or expansion of limbs with respect to the body center. They use these features to train a decision tree, which is then used to classify different expressive gestures.

In motion imitation for robotics, the focus is on transferring human knowledge of movement to a robot. *Nakazawa et al.* [84] use a motion capture system to collect human motion data. They detect the motion primitives by using the correlation index of the end effector's position between different trials. They generate new movements by concatenating multiple motion primitives. The resulting motions are implemented on a humanoid robot using inverse kinematics and a dynamic balancing technique. *Kulic et al.* [67] use an incremental clustering algorithm to incrementally learn different motions performed by the demonstrator without prior knowledge about the motion set.

In the sport science field, some works focus on detecting gradual motion changes that occur due to sports training [66] [35]. *Kulic et al.* [66] use an HMM-based model for modelling human motion and then use Kullback-Leibler divergence to capture the changes in human motion as a result of sports training.

Human motion analysis is also used for medical diagnosis purposes. Most of the publications in this category focus on distinguishing the healthy population from the patient population using classification algorithms [120] [125] [28]. *Toro et al.* [120] use a hierarchical analysis on sagittal kinematic gait data derived from healthy children and children with cerebral palsy to distinguish between the two populations.

The objective of the above-mentioned applications is either to recognize what movement is performed, e.g., [49] or how a movement is performed, e.g., [36]. Often, human motion analysis is divided into four stages as follows:

- Collecting and Preprocessing the Data
- Feature Extraction and Feature Selection
- Classification or Regression
- Validation

Depending on the application, the existing works differ in how they approach the different stages of human motion analysis.

2.1.1 Collecting and Preprocessing the Data

The data collection procedure depends significantly on the application and the purpose of the study. The collected data often contains undesirable variabilities caused by the limitations of the data collection procedure. Therefore, often preprocessing steps are required to lessen the effect of such variabilities on the data.

Camera and other visual tools are typically used to collect data where ease of data collection is essential and sensors or markers cannot be placed on the subjects either due to mobility concerns, e.g., children with cerebral palsy [31], or due to application, e.g., video surveillance [133]. The data collected using visual cameras require image processing techniques to extract human motion from image frames. This can be a hard task because often extracting motion information is affected by the background scene, occlusion and self-occlusion, color and lighting variability, noise, and other well known computer vision challenges [129]. Preprocessing the data to extract human motion from the image frames often involves motion segmentation and object classification, where the former aims at detecting regions corresponding to moving objects such as people and vehicles, and the latter focuses on distinguishing the relevant targets from other moving figures. Common motion segmentation techniques include background subtraction, temporal differencing, and optical flow, and common object classification techniques include shape-based classification, and motion-based classification [129].

In cases where mobility, motion accuracy, and the privacy of participants are important, wearable sensors that collect information about speed, angle, position, and acceleration are utilized. This data collection approach allows collecting the acceleration and velocity of

each limb directly but is susceptible to sensor noise and is affected by communication issues between the sensors and the database recorder. The source of undesired variations in the data collected from wearable sensors depends on the sensor type. The accelerometers are often affected by high frequency noise. Some common approaches for removing the high frequency noise include wavelet decomposition [33], and low pass filtering [76]. Furthermore, the sensors are not always mounted in a consistent way between different trials, and could rotate during a single trial. This results in alignment differences between different trials, and the samples of the same trial. This source of variability can be reduced by initializing the values of the sensors while maintaining a specific posture [73] [33]. The gyroscopes have a floating initial value, which is affected by temperature and acceleration. This type of variability can be lessened using a number of calibration techniques [33] [90].

Some of the works use motion capture to collect data [147] [109] [63]. In this method, reflective markers are placed on selected relevant landmarks which are often close to each joint, and the position of the markers are either utilized directly for motion analysis, or are utilized to derive the joint angles using inverse kinematics. The system is very accurate in estimating the position of the markers and is therefore a common approach in the biomechanics and robotics literature [80] [11]. The cameras in the motion capture system need to be fixed which makes the system immobile. Furthermore, the approach requires specific hardware and special software programs, and the cost of the software and the equipment can be prohibitive. The collected data can contain missing markers due to self occlusion. Moreover, if the markers are placed too close to each other, the system can mistakenly interchange the position of each marker. These errors need to be fixed during post processing by correcting the marker labels and interpolating the position of the missing markers.

The mentioned approaches are not the only methods proposed for collecting the data and data collection approaches could be customized based on the application of the study. For instance, Sun et al. [113] use ultrasonic modules and calculate joint angles from the distances measured by the modules.

2.1.2 Feature Extraction and Feature Selection

Depending on the application and the data collection approach, different feature extraction and feature selection techniques are utilized. When the collected data is available in the form of image sequences, some studies use active contours to model the motion of the moving object, where the idea is to have a representation of the bounding contour of the object and update it dynamically over time. The other common approach is to track the motion using features such as points, lines, or blobs [129].

Where the data is collected using wearable inertial sensors, some studies use the time series of the collected data to estimate the joint angles of the human body for better comprehension [76], whereas others use the raw data to avoid inaccuracies caused by inaccurate modelling, and integrating or differentiating variabilities caused by sensor limitations [117].

To avoid losing information, some studies use the entire timeseries for their analysis [117]. However, this approach increases the calculation complexity, and increases the feature space. Some studies use statistical features extracted from the timeseries [125], whereas others use a model the timeseries in their analysis [67]. The former is a common approach in the biomechanical literature [103] and is fast to compute, and the latter is a common approach in the machine learning literature, and reduces the effect of outliers while considering the temporal and spatial variabilities and attributes [146] [66].

For example, Sun et al. [113] use the time series directly as the feature, whereas Zhang et al. [144] and Uddin et al. [124] extract the temporal and spatial characteristics of the time series and consider those as their features. Statistical characteristics are one of the very common features chosen by most works. Świtoński et al. [115] also consider frequency based features such as the first five values of the Fourier Transform on the time series in their feature sets, and Glowinski et al. [36] include energy and jerkiness in their feature set to detect frustration. Some works use the projection of the data as their features; for example, Jiang et al. [50] use the left eigenvector of the singular value decomposition of the observation matrix.

2.1.3 Classification and Regression Techniques

Based on the purpose of the study, classifiers or regression techniques are used. Classification techniques are used when the purpose of the study is to distinguish two or more classes from each other. Classifiers are trained using either the descriptive features, e.g., [146] [50], or the model of the timeseries, e.g., [67]. Regression techniques are often used when the purpose of the study is to assign a continuum value as the degree of membership of a subject to a certain class. The regression techniques often utilized the descriptive features of a timeseries to construct a regression function, e.g., [115].

For example, Sun et al. [113] creates a model for the time series using robust fitting and check if the subject's time series falls in the prediction interval. In this case, the subject is considered healthy. Taylor et al. [117] use an Adaboost classifier to distinguish correct motions from incorrect motions.

In a few cases, a combination of classification and regression is utilized. Dao and Tho [28] use a predictive decision tree for regression purposes, where each node contains

the degree of belief for cerebral palsy pathology after certain gait patterns are observed. Bayesian Networks are commonly considered for providing a degree of belief of a subject having a condition given certain observations. Van Gestel et al. [125] use the Bayesian Network to decide on whether a subject has a gait pattern characteristic of cerebral palsy.

2.1.4 Validation

Proposed techniques for human motion analysis for rehabilitation and medical diagnosis are often hard to validate because the medical science cannot be used to validate the results of some works, e.g., [125], expert labels are often based on subjective tests and evaluation, e.g., [117], expert labels are often not available, e.g., [115], there is limited access to clinical data, e.g., [146] [117]. Due to such limitations, most works validate their results utilizing cross validation, generalization to new subjects, simulation, figures and charts, and comparison with other techniques.

For validation, most papers chose 10-fold cross validation [117] [53]; However, Van Gestel et al. [125] use 80% of the data for training and 20% for testing. Dao and Tho [28] choose within individual cross validation, between individual cross validation, and leave one out as their testing methods. Furthermore, this work checks the generalization success of the classifier when facing a new subject. Jung et al. [53] use real patients for their validation; Taylor et al. [117] use weights and wearable joint supporters as a means to enforce faulty motion.

2.2 Rehabilitation Motion Analysis

Automatic human movement analysis of rehabilitation exercises analyses the quality of motion to discriminate between movements performed by healthy and patient populations [113] and to perform illness diagnosis [120].

Typically, for a set of movements, key elements of human movement performance are extracted. The important features are used to separate the unhealthy population from the healthy population. Most studies base their methods on classifiers that can discriminate between the two populations, e.g., [125] [144] [132] [28]. These studies rely on a patient database for training such a classifier [28] [132]. For example, *Van Gestel et al.* train a BN classifier based on clinically relevant three dimensional gait analysis (3D gait) parameters extracted from motion data collected from patients with cerebral palsy and the healthy

population, and use it to distinguish between the two. In addition to both healthy population data and patient data, the BN classifier requires prior knowledge, which in this work is obtained from clinical experts.

There are also studies that focus on monitoring features that change when a certain medication or treatment is applied to a group of patients [115] [103], or focus on detecting features that are specific to the patient population [64]. Switonski et al. [115], for example, use a set of discriminating frequency based features in a classifier to discriminate between Parkinson's patient data before and after treatments such as medication and electrical stimulation is applied.

Unlike classification methods which distinguish only between two classes (healthy vs. patient), the focus of this thesis is on patient monitoring and the detection of gradual changes in patient performance due to rehabilitation. To date, only a few studies have focused on assessing the correctness of exercises performed [117] and analysing continuous changes in the movement performed [53] [146].

Post stroke upper body functionality of patients is considered in [53]. The data is collected using a robotic exoskeleton, and sixteen neural subnetworks provide the feature vector. The data is collected from 77 healthy control subjects and 46 stroke patients and both data sets are used in feature selection and classifier training. A multi-layered neural network classifier is used to distinguish between healthy and patient populations. The summation of outputs in the last layer is used to estimate the continuous measure of progress for patients. The analysis is performed on sessions up to 50 days apart with 46 patients performing one exercise [53].

Taylor et al. [117] focus on multiple knee osteoarthritis rehabilitation exercises and record the movements with wearable accelerometers. The exercises included in their study are the standing hamstring curl, reverse hip abduction, and lying straight leg raise, which are commonly prescribed to patients with knee osteoarthritis. They implement a classifier to distinguish the correct performance of an exercise from several types of incorrect ones. Descriptive features such as the mean, minimum, and maximum are calculated from the sensor readings. No intermediate feature selection is performed, and the features are directly used in the classifier. Only data collected from a healthy population is used in the analysis. They use the Adaboost algorithm applied to linear classifiers for each feature. The healthy population data is labelled using expert opinion, and analysis is performed on motions that have recognizable differences. They classify the remaining healthy data into the predefined labels, including a correct class and several incorrect classes.

Zhang et al. [146] focus on post stroke rehabilitation. Motion data is collected with IMUs, and raw sensor output is used for feature extraction after basic filtering. The time-series data for each sensor is partitioned for different exercises. Partitions that correlate

least with corresponding partitions of other exercises are considered as motion templates. The patient data is then cross-correlated with the templates, and the peak values of the cross-correlation are considered as the features. Data is collected from rehabilitation professionals and a single patient is used for testing. K- Nearest Neighbours is used to classify patient motions, and the distance from the center of the cluster is the estimate of continuous progress.

The three studies consider different types of data, i.e., position, velocity, and acceleration, and concatenate the numerical values in a common feature vector [53] [117] [146]. All three studies use classifiers as the basis of their analysis [53] [117] [146]. When continuous labeling is desired, it is based on the decision criteria of the classifiers used [53] [146]. The classifiers are utilized to label the motions performed by patients [117], separate the healthy population from the unhealthy population and monitor progress [53], and detect whether patients perform motions correctly [146]. Neural network classifiers are the basis of analysis in [53], and often need fine tuning or are hard to replicate or extend because their structure makes clinical interpretation difficult.

Patient data is often hard to collect. Therefore in many cases patient data is not available and the validation is based on a healthy population imitating unhealthy movements [117]. There are also cases were a small number of patients (less than 5) is used for validation, e.g., one patient in [146]. Generalizability to new patients can be evaluated when the classifier is trained on the data of a subset of patients and tested on the data of the remaining subjects. Only a few studies consider generalization to new patients, e.g., *Jung et al.* [53] use 46 patients performing one exercise in their validation.

Validation of studies considering continuous labelling (e.g., [53]) is a difficult task because an objective quantitative ground truth of continuous progress is rarely available. Quantitative assessment scores are often not collected for each physiotherapy session of a patient due to the limited time in each session. Therefore, visual graphs and cross validations (e.g., comparison to other classifiers' performance) are common methods for validation [53] [117] [146].

2.3 Summary

Analysing human motion for rehabilitation purposes is a challenging task due to the multiple sources of variability in human motion data, limitations of data collection, different exercise regimens for each session and each patient, and within and between subject variabilities. The current state of the art develops models for healthy and patient populations

[117] [53] [146] and therefore is capable of assigning the class labels "healthy" or "unhealthy" to captured movement sequences. The disadvantage of classification techniques is that they cannot explicitly model the continuous progress in the patient motion and therefore are not suitable for continuous monitoring purposes. Many of the works in the state of the art focus on classifying healthy or unhealthy movements based on one specific exercise [146] [53]. This is a limitation for monitoring patients over the course of rehabilitation because the exercise regimen consists of more than one exercise, and the set of exercises performed changes over the course of the treatment. Furthermore, many of the current works [53] [117] validate their methods based on synthetic and simulated data due to lack of patient data.

In this work, we propose a technique that estimates the continuous measure of patient improvement, and is capable of handling a variety of exercises. We validate our proposed approaches based on both synthetic and clinical data. Of the challenges summarized above, we address capturing the variability caused by improvement in human motion, validating the proposed approaches based on clinical data, and handling different exercise regimens for each patient and each session. We do not address factors such as pain and fatigue that affect human motion, and we do not address the identification of causes for improvement or degradation in performance.

Chapter 3

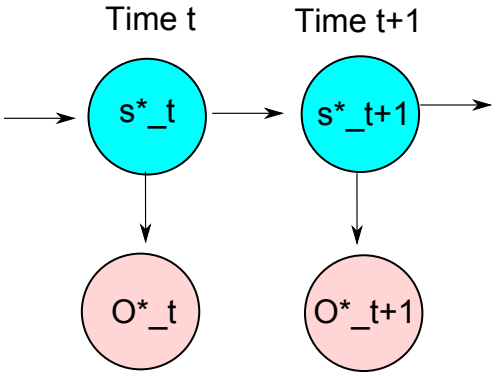
Background

The progress estimation algorithms to be discussed in subsequent chapters involve modelling human motion, and utilizing classification and regression techniques to distinguish the healthy population from the patient population and capture patient progress. In this chapter, the computational techniques used for developing the motion analysis techniques are reviewed, including Hidden Markov Models (HMM), Support Vector Machines (SVM), Naive Bayes (NB), and HMM and Kullback-Leibler (KL) divergence, Least Absolute Shrinkage and Selection Operator (LASSO), kernel based approach for a two-sample problem, and Dynamical Movement Primitives (DMP).

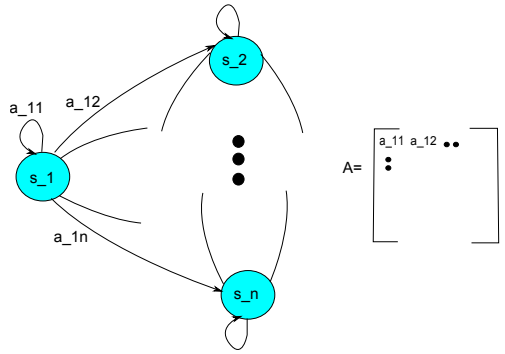
3.1 Hidden Markov Models

Modelling time series data (*signals*) allows capturing the main attributes and information of a signal and representing them in a meaningful and compact manner. HMMs [98] are a stochastic modelling approach that characterize the statistical properties of the signal. This modelling technique captures both temporal and spatial variability of the timeseries, and has been therefore applied in modelling human motion [67] [129].

The HMM represents the timeseries signal by an evolving unobservable state variable (see Fig. 3.1a). The state variable transitions through a discrete set of values, and the probability of these transitions is determined by the state transition matrix. At each time step the system either remains in its current state or transitions to another state (see Fig. 3.1b). The state follows the Markov property, i.e., at each time step the state only depends on its previous value. When modelling a timeseries using an HMM, the state is not directly observable. The hidden state is inferred by the probabilistic relation between the state,



(a) HMM representation in the time domain



(b) Graphical representation of an HMM state transition matrix

Figure 3.1: (a) Depicts the HMM representation of a timeseries, at each time step the hidden state $s*_{t+1}$ can be inferred by the probabilistic relation between the state, its observable output and the previous state. (b) Depicts the state transition matrix, which is illustrated as a directed graph.

its observable output and the previous state. To characterize an HMM λ the following parameters are needed:

- **State Transition Matrix** A**:** This matrix determines the probability of transitioning from one state value to another. If a transition is possible between all the states in an HMM, called an ergodic HMM, the transition matrix will not have any zero elements. However, in cases that transitions are not possible between all states, the elements of the matrix that correspond to disconnected states are zero. In this thesis, we use a left-right HMM, in which each state can transit to one subsequent state and itself.

- **Observation Probability** b_i**:** In the continuous HMM, the probability of the observable outputs for each state is often represented using a mixture of Gaussian distributions. Human motion is a continuous phenomenon and therefore in this research we use continuous HMMs. Based on our observations from the data, a single Gaussian distribution is sufficient for modelling human motion. In this case, the mean, μ, of the output distribution represents the most probable value for the output and the covariance, Σ, of the output distribution represents the range of variability.

- **Initial State Distribution** π**:** This vector determines the probability that the signal sequence begins with each state. In a left-to-right HMM, it is often assumed that the signal begins in its initial state.

The forward algorithm calculates the likelihood that the observation data O is generated by the model λ as follows:

$$P(Q|\lambda) = \pi_1 a_{1,2} a_{2,3} \dots a_{T-1,T} \tag{3.1}$$

$$P(O|Q,\lambda) = b_1(O_1) b_2(O_2) \dots b_T(O_T) \tag{3.2}$$

$$P(O|\lambda) = P(O|Q,\lambda) P(Q|\lambda), \tag{3.3}$$

where Q is the state sequence, and, given the model λ, a is the probability of Q occurring and b is the probability of O occurring.

The Baum-Welsh algorithm is an expectation maximization algorithm utilized to train the HMM model in an iterative fashion [98]. In the first step, the expected value E is calculated for the nth iteration model λ_n by taking the expectation of the log-likelihood of the observation O, state sequence Q, and the model of the previous iteration λ_{n-1} using the following equation.

$$E(\lambda_n|\lambda_{n-1}) = \mathbb{E}[\log(P(O,Q|\lambda_n))|O,\lambda_{n-1}] \tag{3.4}$$

In the maximization step the following equation is solved for λ to update the model parameters.

$$\frac{\partial \log(O|\lambda)}{\partial \lambda} = 0 \tag{3.5}$$

The two steps are repeated until the model parameters converge or the maximum number of iterations is reached. In this research, the maximum number of iterations is 200.

3.2 Kullback-Leibler Divergence

Given two HMMs, λ_1 and λ_2, the Kullback-Leibler divergence (KL) measures the similarity of the two models. This measure of difference between the two HMMs, λ_1 and λ_2, is defined as

$$D(\lambda_1,\lambda_2) = \frac{1}{T}[\log P(O^{(2)}|\lambda_1) - \log P(O^{(2)}|\lambda_2)] \tag{3.6}$$

where $O^{(2)} = O_1 O_2 O_3 \dots O_T$ is a sequence of observations generated by model λ_2. As can be seen from the formulation, Eq. 3.6 is a measure of how well λ_1 matches the data generated from λ_2 compared to how well λ_2 matches data generated from itself [98]. The symmetric distance is:

$$D_s = \frac{D(\lambda_1,\lambda_2) + D(\lambda_2,\lambda_1)}{2} \tag{3.7}$$

3.3 Support Vector Machine

The Support Vector Machine (SVM) is a method that constructs a halfspace in any dimension n, and can be used for classification or regression tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (functional margin), because in general the larger the margin of a classifier the lower the generalization error. The SVM formulation as given in [25] [114] is obtained using the following procedure. Given a training dataset of size N i.e. $(L_k, \mathbf{v}_k)_{k=1}^N$ where $\mathbf{v}_k \in \mathbb{R}^n$ is the k-th feature vector and $L_k \in \{1, 0\}$ is the kth class output value, the SVM aims to construct a classifier of the form:

$$L(\mathbf{v}) = sign(\sum_{k=1}^{N} \alpha_k L_k \psi(\mathbf{v}, \mathbf{v}_k) + b) \tag{3.8}$$

where $\alpha_k \geq 0$ is the support vector coefficient, $b \in \mathbb{R}$ is the bias term, and $\psi(\mathbf{v}, \mathbf{v}_k)$ is the kernel function which for the case of linear SVM is the same as dot product. The slack variable ξ_k controls the allowable degree of misclassification of the data $\mathbf{v}_i$. The risk bound is minimized by formulating the optimization problem

$$\min_{w,\xi_k} J_1(w, \xi_k) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k \tag{3.9}$$

subject to:

$$L_k(w^T \phi(\mathbf{v}_k) + b) \geq 1 \quad , \quad \xi_k \geq 0 \quad , \quad k = 1, ..., N \tag{3.10}$$

where w is the vector that defines the half space constructed by the SVM, c is a constant, and N is the total number of data points. Fig. 3.2 illustrates an SVM classifier.

3.4 Naive Bayes

The Naive Bayes (NB) algorithm is a classification algorithm which is based on the Bayes rule, and assumes that given the label of the classes, Y, the probability of the attributes, $v1 \ldots v_n$, are independent of each other (see Fig. 3.3) [116]. The NB is a simplified version of the Bayesian Network, where the probability of each node affects all the nodes connected to it directly or indirectly. The NB representation allows for a simplified representation of the probability of the input attributes given the label, i.e., $P(V|Y)$, and therefore simplifies the problem of estimating this probability from the training data. In general, when V contains

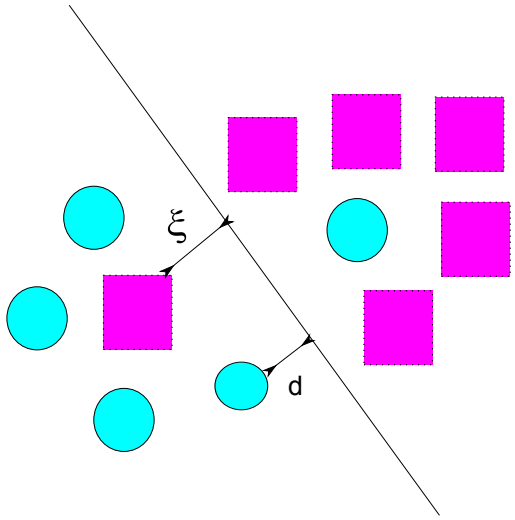


Figure 3.2: SVM model, illustrating the decision boundary of SVM, ξ is the amount of misclassification, d is the distance of the support vectors from the decision boundary, which is maximized by the algorithm.

n attributes which are conditionally independent of one another given Y, the probability of attributes given the class labels is [116]

$$P(v_1 \dots v_n|Y) = \prod_{i=1}^{n} P(v_i|Y) \tag{3.11}$$

In NB classifiers, the objective is to train a classifier that will output the probability distribution over the different outputs, Y, for every new observation of the input attributes, V. The probability of a label given the new instance of the data is calculated using

$$P(Y = y_k|v_1 \dots v_n) = \frac{P(Y = y_k) \prod_i P(v_i|Y = y_k)}{\sum_j P(Y = y_j) \prod_i P(v_i|Y = y_j)} \tag{3.12}$$

The most probable class label for a new instance of data is the label which maximizes Eq. 3.6.

3.5 Dynamical Movement Primitives

Many fields of science such as robotics, human motion analysis, and neuroscience use nonlinear dynamical systems to model complex behaviours. It is difficult to model and

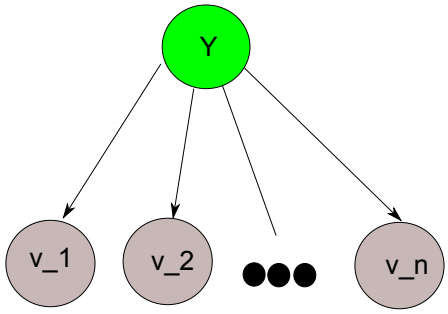


Figure 3.3: Naive Bayes

control the behaviour of non linear systems due to their parameter sensitivity, complex phase transitions, and difficulty of analysing their long term behaviour.

The Dynamical Movement Primitives (DMP) [47], uses a set of linear differential equations as the base system. The method transforms the linear system into a weak nonlinear system with desired attractor dynamics. The desired attractor dynamics is achieved by adding a learnable nonlinear term to the the differential equations of the system. *Ijspeert et al.* [47] use a damped spring model as the simple dynamical system therefore the system model is

$$\tau\ddot{y} = \alpha_z(\beta_z(y_m - y) - \dot{y}) + f, \tag{3.13}$$

where τ is a time constant, α_z and β_z are positive constants and are set such that the linear system is critically damped, i.e., $\beta_z = \frac{\alpha_z}{4}$, f represents the attractor dynamics and is called the forcing term, y_m is the base of oscillation, y is the system's position, $\dot{y}$ is the system's velocity, and $\ddot{y}$ is the system's acceleration [47].

The forcing term, f, is the term that adds the weak nonlinearity to the simple dynamical system. f can be chosen such that the system has weak nonlinearities, and periodicity is introduced to the system, i.e., limit cycle attractor [47].

In the limit cycle case, the forcing term f is obtained using the following equations

$$f(\phi, r) = \frac{\sum_{i=1}^{N} \Psi_i(\phi) w_i}{\sum_{i=1}^{N} \Psi_i(\phi)} r, \tag{3.14}$$

$$\Psi_i(\phi) = e^{((\cos(\phi - c_i) - 1))}, \tag{3.15}$$

$$\tau\dot{\phi} = 1, \tag{3.16}$$

where Eq. (3.16) is a phase oscillator utilized for learning the limit cycle attractor, r is the amplitude of the oscillator and in this thesis $r = 1$, $\phi \in [0, 2\pi]$ is the phase, Ψ_i are Gaussian like periodic basis functions, and c_i are the center of each basis function [47]. The weights of each basis function, w_i, describe the shape of the original timeseries in the effective region of each basis function. Therefore this w_i can be utilized to determine the difference between two timeseries.

3.6 A Kernel Based Approach for a Two-sample Problem

Statistical tests designed to distinguish between two sample distributions have a wide range of applications. In bioinformatics, they are used to detect the effects of different experimental procedures on micro arrays generated from the same tissue [39]. In database attribute matching, where merging multiple fields are desirable, they are used to detect the fields that correspond to each other [39].

Gretton et al. [39] propose a statistical test that determines whether two samples are drawn from similar distributions, i.e., given that samples from two distributions are available the test determines how similar the two distributions are. The statistical test is defined as the largest difference in expectation of the functions in the unit ball of a reproducing kernel Hilbert space. The statistical test as formulated in [39] becomes

$$MMD\left[P, R\right] = \left[\frac{1}{m^2}\sum_{i,j=1}^{m} k(p_i, p_j) - \frac{2}{mn}\sum_{i,j=1}^{m,n} k(p_i, r_j) + \frac{1}{n^2}\sum_{i,j=1}^{n} k(r_i, r_j)\right]^{\frac{1}{2}}, \tag{3.17}$$

where p_i is the ith sample drawn from distribution P, r_i is the ith sample drawn from distribution R, $k(.,.)$ is the kernel function, m is the number of samples available from distribution P, n is the number of samples available from distribution R, and MMD is the measure of distribution similarity. For a set of samples drawn from two different distributions, a large MMD value indicates that the samples are generated by different distributions. As the distributions become more similar the value of MMD decreases. The test value for samples of two distributions is then compared to a certain threshold. If the MMD statistic falls above the threshold, the null hypothesis is rejected, and therefore the difference of the distributions is statistically significant.

3.7 Least Absolute Shrinkage and Selection Operator

Least Absolute Shrinkage and Selection Operator (LASSO) is a regression tool which is often used for estimating linear models [118]. LASSO minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a tuning parameter. This formulation of LASSO results in a zero value for some coefficients. The number of zero coefficients depends on the tuning parameter. This property of LASSO makes it useful for feature selection.

When applying LASSO for feature selection, for a set of inputs $f_1, f_2, \ldots, f_k$, an output y and the following linear model

$$\hat{y} = w_0 + w_1 f_1 + w_2 f_2 + w_3 f_3 + \ldots + w_k f_k \ , \tag{3.18}$$

LASSO adjusts the weights $w_0, \ldots, w_k$ such that $\sum(\hat{y} - y)^2$ is minimized and $\sum_{i=0}^{k} w_i < t$ where $t \geq 0$ is a tuning parameter [118]. The parameter t is selected so that the weights w are larger than zero for only a specific number of features. When w_i becomes zero, the input f_i does not contribute to minimizing $\sum(\hat{y} - y)^2$, i.e., f_i is either uninformative or its information is redundant. Therefore the features with zero coefficients are not considered as informative features, and only features with non zero coefficients are selected.

3.8 Kruskal Wallis

The Kruskal Wallis (KW) [65] one-way analysis of variance is a non parametric approach that determines how likely it is for samples of multiple datasets to be from the same distribution. For an input vector $\mathbf{f_i} = \left[trial_1, \ldots, trial_N\right]$ and label set $\mathbf{Y} = \left[y_1, \ldots, y_N\right]$, the test ranks the data from 1 to N in an ascending order without considering the labels. The test statistic is then calculated using

$$K = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1) \ ,$$

where R_i is the sum of the ranks in group i, and n_i is the number of values in group i. The probability of being from the same distribution (p) is approximated by $Pr(\chi^2_{g-1} \geq K)$. A lower value of p indicates that it is more likely for the datasets to have different underlying distributions and that they are more likely to be separable. In our study, $\mathbf{F_i}$ contains the values of the ith feature over the different trials of the healthy and patient populations. The label set $\mathbf{Y}$ contains label values $y \in \{0, 1\}$. y is 1 if the feature value is from the healthy population and 0 if the feature value is from the patient population. The features with the lowest value of p are considered as the most informative features.

Chapter 4

Patient Progress Estimation from Rehabilitation Exercises

In this chapter, the proposed approaches for formulating the measures of performance for repetitions of one or multiple exercises are presented. Furthermore, the application of existing approaches to estimating measures of performance for repetitions of one exercise is provided.

During each physiotherapy session, the patient is instructed to perform repetitions of different exercises. The physiotherapist assesses the patient's performance based on his or her performance during each repetition of each exercise. Therefore, analysing patient progress during physiotherapy requires answering the following questions:

1- How to assess one repetition of one exercise performed by a patient?

2- How to assess multiple repetitions of one exercise performed by a patient?

3- How to combine the evaluations from different exercises and obtain a score that denotes the overall performance of a patient in a single session?

In this thesis, we aim to identify a composite index of movement features which can quantify exercise *performance*. To address the above questions, we propose two approaches. Each of the approaches proposed utilizes the difference between the patient's motion and the healthy population's motion as the measure of performance for one or multiple repetitions of an exercise. The exercise difficulty is then estimated based on the healthy population's performance, by assuming that as the exercises become harder, the variance

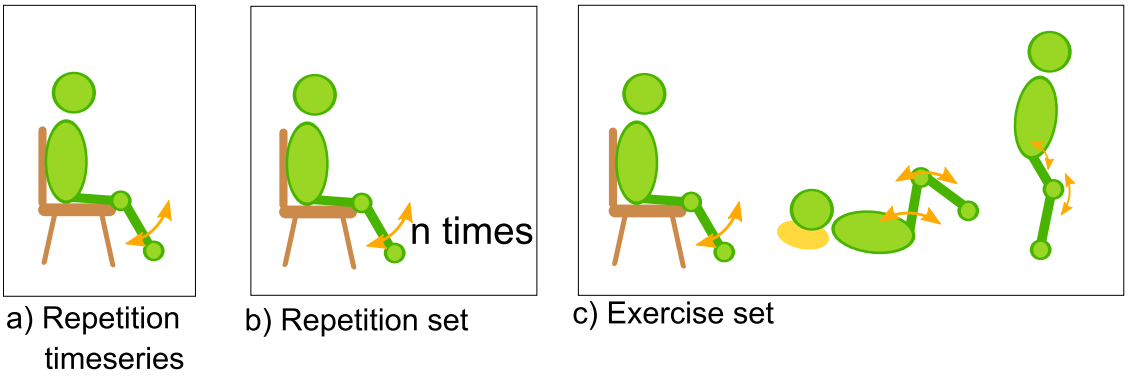


Figure 4.1: The *repetition timeseries* and the *repetition set* provide the performance measures δ, and Δ. The overall score S assesses an *exercise set* of a session. During the knee extension exercise, the subject performs a full knee extension/flexion without moving any other joints. During the knee hip extension exercise the subject lies on the ground and performs a full knee hip extension/flexion. During the squat the subject bends his knees and hips while standing.

of the measure of performance between the healthy population increases. Lastly, utilizing the measure of performance for each exercise, and the exercise difficulty, the overall performance score is calculated.

For both approaches, we assume that the motion data is available in the form of joint angle positions, velocities, and accelerations. We also assume that motion data is available from a healthy population performing the same set of exercises. We make the assumption that, at the time of the analysis, we know which exercise is being performed, that the data is segmented such that one single repetition of a certain exercise is a *repetition timeseries* $\omega = \begin{bmatrix}\boldsymbol{\gamma}(1) & \boldsymbol{\gamma}(2) & ... & \boldsymbol{\gamma}(T)\end{bmatrix}$, where T is the duration of the repetition for that exercise, and $\boldsymbol{\gamma}$ is a vector of joint kinematics $\boldsymbol{\gamma} = \begin{bmatrix}q_1 & q_2 & \dots & \dot{q}_1 & \dot{q}_2 & \dots & \ddot{q}_1 & \ddot{q}_2 & \dots\end{bmatrix}$. Multiple repetitions of the same exercise performed in the same session are the *repetition set* for that exercise $\Omega = \{\omega_1, \dots \omega_n\}$ where n is the number of repetitions. The set of multiple exercises performed in the same session are the *exercise set* of that session $\Gamma = \{\Omega_1, \dots, \Omega_m\}$ where m is the number of different exercises performed in the session (See Fig. 4.1).

4.1 Proposed Measures of Performance for a Repetition Timeseries and a Repetition Set

In this section, we propose two approaches for estimating the measure of performance for a single *repetition timeseries* of an exercise, and multiple *repetition timeseries* of the same exercise. We also describe the procedure for obtaining the measure of performance for a single *repetition timeseries* and a *repetition set* based on the related works. The proposed approaches are based on comparing the patient's motion to that of the healthy population.

4.1.1 Feature-Based Approach

In the feature-based method, the mean, minimum, maximum, skew, and range of motion of the joint angle positions, velocities, and accelerations plus the duration of each *repetition timeseries* are considered as the feature vector.

$$\mathbf{v} = [\mu_{q_1} \quad min_{q_1} \quad max_{q_1} \quad skew_{q_1} \quad rom_{q_1} \quad \mu_{q_2} \quad \dots \quad duration] \tag{4.1}$$

$$skew_{q_i} = \frac{\frac{1}{T}\sum_{j=1}^{T}(q_{i_j} - \mu_{q_i})^3}{(\sqrt{\frac{1}{T}\sum_{j=1}^{T}(q_{i_j} - \mu_{q_i})^2})^3} \tag{4.2}$$

$$rom_{q_i} = max_{q_i} - min_{q_i} \tag{4.3}$$

This definition of the feature vector is desirable because it allows modelling the timeseries of the data using statistical features. This method is fast to compute and can capture the attributes of the timeseries from one example. However, since the features are defined directly from the timeseries, the approach can be adversely affected by unwanted variabilities such as noise. The feature vectors are extracted for each repetition of every exercise performed over the course of rehabilitation. These features are also extracted from the healthy population data for the same exercises. If the number of *repetition timeseries* is larger than the number of features, features that reflect what changes most throughout the rehabilitation are chosen using Least Absolute Shrinkage and Selection Operator (LASSO) [118] (See Sec. 3.7). If the number of *repetition timeseries* is smaller than the number of features, LASSO does not have enough samples to perform the regression task correctly. In this case, we use Kruskal Wallis one-way variance analysis (KW) to select the most informative features (See Sec. 3.8).

The *distance* measure is defined as the difference between the considered set of kinematic parameters and the mean of the healthy data. The healthy data is then considered

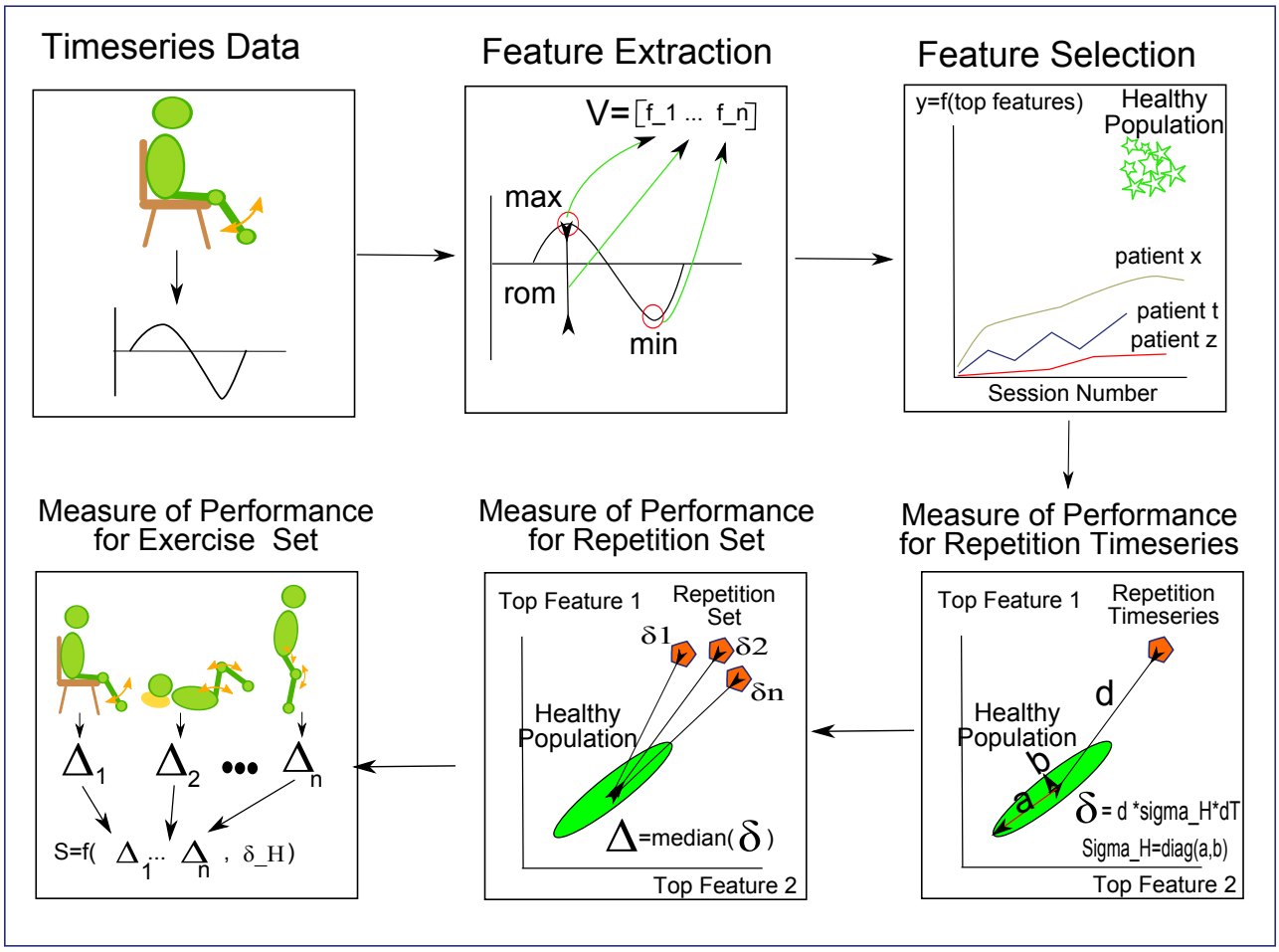


Figure 4.2: Feature-Based Approach: In the first step, data is collected from the patient using IMU sensors, and the joint angles are estimated. The data is then segmented such that each segment begins with the beginning of an exercise, and the segment ends when one repetition of the exercise is completed. Statistical features are extracted from the joint angles' timeseries, and the most descriptive features are selected using LASSO/KW as described in Sec. 4.1.1. Lastly, the measures of performance for a *repetition timeseries*, a *repetition set*, and an *exercise set* are calculated using a weighted distance between the patient population and the healthy population in the feature space.

as the reference and the *distance* considering a set of kinematic parameters to the mean of this reference dataset is the measure of progress for each exercise. The results of different exercises are normalized by the mean and variance of their corresponding healthy dataset.

The measures are then combined to obtain an overall score for each patient's performance in a given session.

Feature Selection

To assess the performance, it is essential to extract the informative features from the feature vector and exclude those that are uninformative or redundant. The informative features extracted from the feature vector are the top features. Different features may be informative for different exercises, therefore the top features are selected automatically from the data by looking for those features which show the most variation over the course of treatment and are most different from the healthy population. As there are multiple sources of variations in human motion, we cannot assume that there is a linear relationship between the number of days in treatment and motion features. We use a linear model only for feature selection, i.e., for identifying which features change during the course of treatment. When including healthy population data, a linear model can be used to estimate the suitability of features to discriminate between the healthy population and patient data while considering also a possible linear dependence between feature and session number.

In our study, we use LASSO or KW for feature selection (See Secs. 3.7 and 3.8). For LASSO the inputs $f_1, f_2, \ldots, f_k$ are the features of the *repetition timeseries* and the output y is the corresponding session number. The session numbers are normalized between 0 and 1, such that 0 corresponds to a patient's first session, and 1 corresponds to a patient's last session. We select features that change with every session and among these features the ones that most correspond to a linear relationship. This selection allows us to find the features that are changing as patients progress through the sessions. For the purposes of feature selection, we also consider the healthy population data in this regression. For the healthy population, the label y is set to be 100 times larger than the patients' last session. Introducing this outlier forces the regression to be in the direction of the healthy population data and helps to detect the features that not only change with the progress of the patients but also separate the healthy population from patients. The value of y for the healthy population directly affects the value of the weights, but the chosen features are not changed as long as y is sufficiently large. We do not use the values of the weights in our analysis and only use the features selected by this method.

The tuning parameter t is set such that five non zero coefficients remain. The remaining 5 features, which correspond to the five non zero coefficients, are considered as the top features in the subsequent analysis. Preliminary experiments showed that the algorithm is not sensitive to this value. Any value of t resulting in a range of 5-25 features results in the same performance measures.

For KW, for each feature, the inputs are the different trials $f_i = [trial_1, \ldots, trial_N]$, and the output y is the label of the classes, healthy or patient, for each trial. The five features with the smallest probability statistic p are considered as the top features.

Measure of Performance

To obtain a measurement for the performance of one exercise, the feature vectors are extracted for the patient (V_P) and healthy population (V_H) data as described in Sec. 4.1.1.

$$V_H^{'} = V_H(top_{features}) \tag{4.4}$$

$$V_P^{'} = V_P(top_{features}) \tag{4.5}$$

Based on our observations, to a smaller degree, healthy individuals employ the same compensation strategies that patients use when performing an exercise, i.e., healthy subjects show the same compensation strategies due to mental and physical fatigue, lack of physical readiness, and misunderstanding the exercise instructions. For example, in the knee hip extension exercise, the correct form of the exercise is to perform a full range of knee extension while minimizing motion in the other joints. Based on our observations, the healthy population often compensates during this motion by also rotating the extension/flexion hip joint. We assume that among the features chosen by LASSO or KW, the ones with higher variance in the healthy population are more informative, because the highly variant features are either features of the moving joint or are the features describing the compensation strategy. Therefore more weight is given to the more variant features in defining the distance measure. The distance δ between the patient repetition and the healthy population data evaluates each repetition:

$$\mu_H = mean(V_H^{'}) \tag{4.6}$$

$$\Sigma_H = diag(std(V_H^{'})) \tag{4.7}$$

$$\delta_i = (V_{P_i}^{'} - \mu_H)^T \Sigma_H (V_{P_i}^{'} - \mu_H) \tag{4.8}$$

Where $V_H^{'}$ is the healthy population top feature vector, $V_{P_i}^{'}$ is the patient's top feature vector for the ith *repetition timeseries*, μ_H is the mean of the healthy population's top feature vectors, Σ_H is the diagonal matrix of standard deviations for the healthy population top feature vectors, and δ_i is the distance between one repetition of the exercise performed and the healthy group's performance. We assume that as patients improve they get closer to the healthy data and therefore a decrease in the value of δ over the course of rehabilitation indicates improvement.

Measure of Performance for Multiple Repetitions of the Same Exercise

For each patient, δ represents the measure of performance for one repetition of an exercise. The median of the distance measures (δ) calculated for one exercise over the set is considered as the overall distance measure for the *repetition set* of that exercise

$$\Delta_\Omega = median(\boldsymbol{\delta}_\Omega) \ , \tag{4.9}$$

where Δ_Ω is the overall performance of one exercise in one session and $\boldsymbol{\delta}_\Omega$ is the vector of distance measures calculated for every *repetition timeseries data* ω_i in the *repetition set* Ω. The median is used to lessen the sensitivity of the approach to outliers. The process of calculating the measures of performance using the feature-based approach is illustrated in Fig. 4.2.

4.1.2 HMM-Based Approach

The Hidden Markov Model (HMM) [98] based approach relies on features extracted from HMM modelling of the joint angle timeseries. HMMs are trained on the *repetition set* of each exercise for the healthy and patient populations.

Individual HMMs are learned for each member of the healthy population and for each session of each patient; each *repetition set* is modelled using a 3 state, left-to-right model. States 1, 2, and 3 correspond to attempting to reach the desired posture, reaching the desired posture and pausing, and returning to the starting posture. The observations of the HMMs are the position, velocity, and acceleration of the joint angles. The mean and variance of the observation distributions in each state are considered as the feature vector.

$$\mathbf{v} = \begin{bmatrix} \mu_{state1_{q_1}} & \sigma_{state1_{q_1}} & \mu_{state2_{q_1}} & \cdots & \mu_{state3_{\ddot{q}5}} & \sigma_{state3_{\ddot{q}5}} \end{bmatrix} \tag{4.10}$$

The feature selection is performed using LASSO or KW as described in section 4.1.1.

Measure of Performance

If the number of features is smaller than the number of the *repetition sets*, the top features are selected using LASSO. If the number of features is larger than the number of the *repetition sets*, the top features are selected using KW. We assume that among the top features chosen by LASSO or KW those that are highly variant in the healthy population indicate the rehabilitation process better. Therefore, we use the same procedure as section

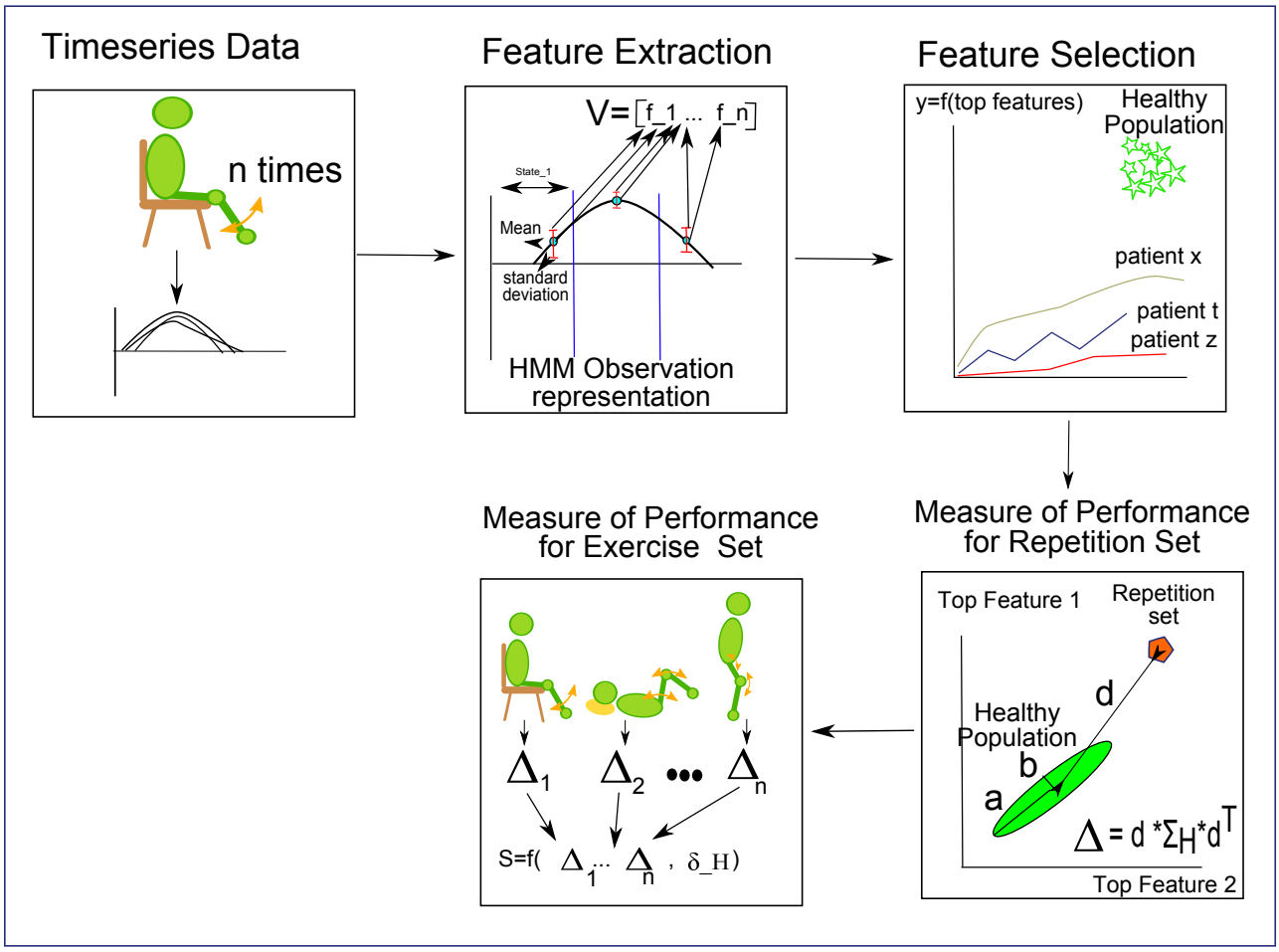


Figure 4.3: HMM-Based Approach: After motion recording and segmentation, a three-state HMM is trained for each *repetition set*. The mean and standard deviation of the observation for each state are considered as features. The most informative features are selected using LASSO or KW, and the most informative features are selected using KW, and the measures of performance for the *repetition set*s, and the overall *exercise set* are calculated.

4.1.1 to calculate distance measures. The five features chosen by LASSO or KW are chosen as the top features. The distance measure Δ_{HMM} for a *repetition set* is obtained using the HMM-based feature vectors and Eqs. 4.5-4.8.

The HMM is capable of capturing the statistical essence of a dynamic timeseries. Such a definition of the feature vector is beneficial, because it models the most likely timeseries and

the probability of variations. Furthermore, the feature-based approach requires expertise in predefining the features whereas the HMM captures the features that describe the pattern of the data automatically. The HMM is computationally more expensive than the feature-based approach and to be effective, it requires multiple samples of the timeseries data which may not be available. When the availability of data is not a concern, an advantage of the HMM is that it can be trained using several timeseries and that it represents the most likely timeseries. For this reason, training HMMs for each *repetition timeseries* is not necessary to assess the performance for an exercise set or to assess an overall score. Fig. 4.3 illustrates the required steps for calculating measures of performance utilizing the HMM-based approach.

4.2 Comparison Approaches for Determining a Measure of Performance for a Repetition Timeseries

4.2.1 Classifier-Based Approaches

Classifiers are a common technique often used in the human motion analysis literature for separating healthy population from patients [125] or to detect the changes in motion for one exercise [53]. In this thesis, we compare the performance of the classifier-based approaches in estimating the measures of performance for one exercise with the proposed approaches and the remaining related works. The features are extracted from the timeseries as described in Sec. 4.1.1. The most informative features are chosen using LASSO or KW as described in Sec. 4.1.1.

Support Vector Machine

The SVM is a common classification technique, which is often utilized in the bio-mechanical literature (See Sec. 3.3). When trained on the the entire feature set chosen by LASSO or KW, the SVM performs poorly and is not capable of capturing the trend of progress. The performance is improved when the SVM is trained using the top features chosen by LASSO or KW weighted by the variance of the features in the healthy population. In this work, the best results are reported, using the variance-weighted features.

Measure of Performance The SVM is trained using a subset of the data from both classes (patient and healthy). The label L is 1 for the healthy population data and 0 for

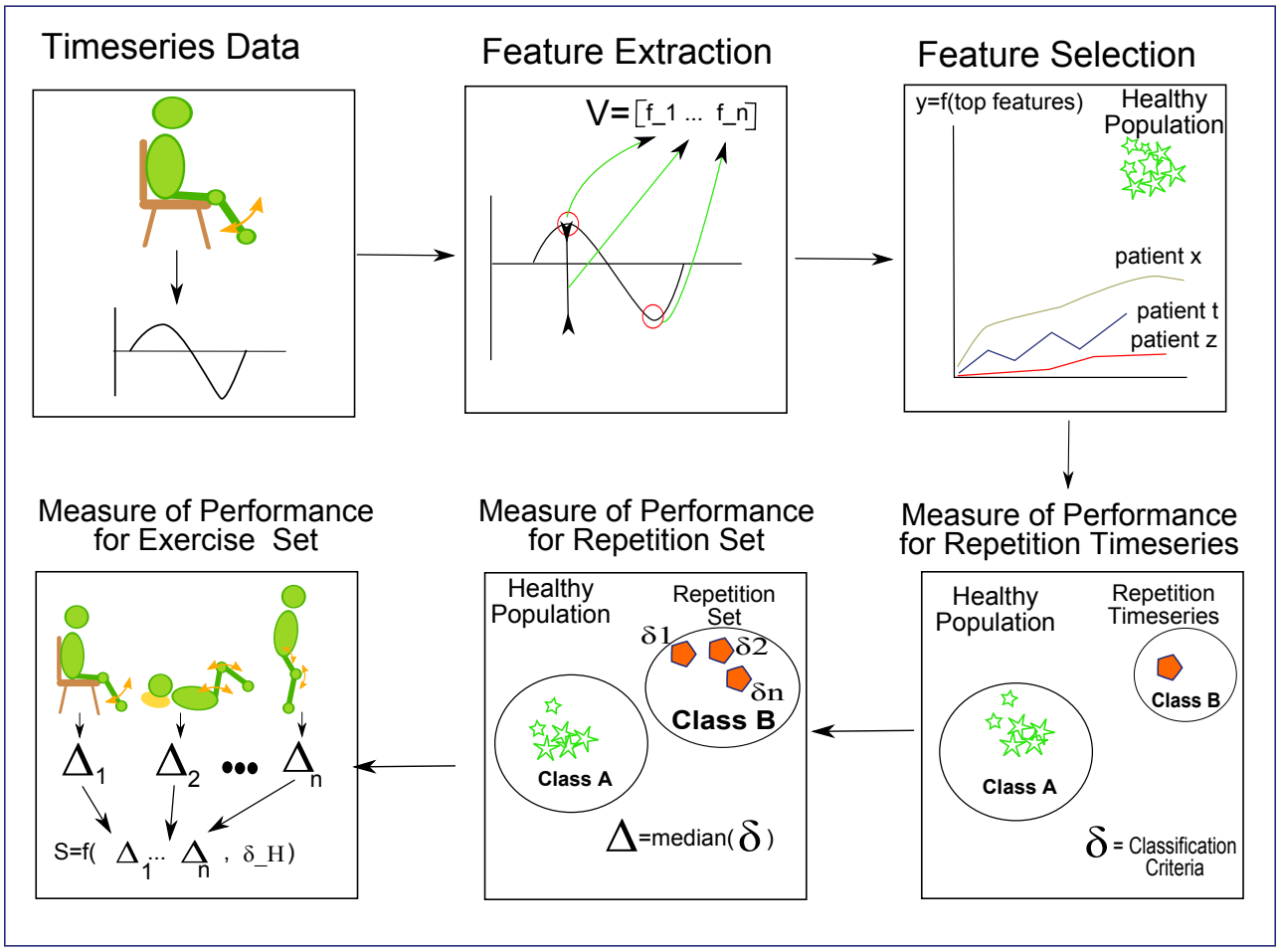


Figure 4.4: Classifier Based Approach: A classifier (NB, SVM, or KL-divergence) is trained using half of the patient data, and the healthy population data. The measure of performance for one *repetition timeseries* is calculated based on the classification criterion. The measures of performance for a *repetition set*, and *exercise set* are then calculated.

the patient data. The classification decision of the SVM is based on

$$sign(\sum_{k=1}^{N} \alpha_k L_k \psi(\mathbf{v}, \mathbf{v}_k) + b) \tag{4.11}$$

This value corresponds to each training point's distance from the decision boundary. We consider the $\sum_{k=1}^{N} \alpha_k L_k \psi(\mathbf{v}, \mathbf{v}_k) + b$ value as the distance measure δ_{SVM} for *repetition timeseries* of an exercise in each session. The median of these values is considered as the

overall distance measure Δ_{SVM} for a *repetition set*.

Naive Bayes

In this thesis, NB (See Sec. 3.4) is trained using the top features selected by LASSO or KW in 4.1.1 on data from a subset of patients and the healthy population data following the procedure described in Sec. 3.4. The probability of belonging to the healthy population class normalized by the summation of probabilities of belonging to the healthy or patient class is considered as the distance measure δ_{NB} for each *repetition timeseries*.

$$\delta_{NB} = \frac{Pr(y = healthy|\mathbf{v})}{Pr(y = healthy|\mathbf{v}) + Pr(y = patient|\mathbf{v})} \tag{4.12}$$

Kullback-Leibler Divergence

The KL divergence measures the similarity of two HMMs. In the bio-mechanics literature it is often the case that a class of data is modelled using HMMs and it is desired to detect whether the rest of the data is from the same class. In such situations the KL divergence is used as the classification measure [98].

Measure of Performance For the KL divergence, an HMM is trained for the entire healthy population data and on every *repetition set* of every exercise and each patient. The symmetric KL divergence D_{KLD} for each patient is calculated for each *repetition timeseries* using the procedure described in Sec. 3.2. The KL divergence is considered as the distance measure δ_{KLD} of each *repetition timeseries*.

$$\delta_{KLD} = D_{KLD}(\lambda_1, \lambda_2) \tag{4.13}$$

Fig. 4.4 illustrates the overall procedure of the classifier-based approaches.

4.2.2 DMP-Based Approach

In the DMP-based approach each timeseries is modelled using a non-linear dynamical system. The non-linear component of the model is learned from the data as a set of weighted basis functions. The weights of the nonlinear term are considered as the features and the measures of performance are obtained by calculating the Euclidean distance between the feature vectors.

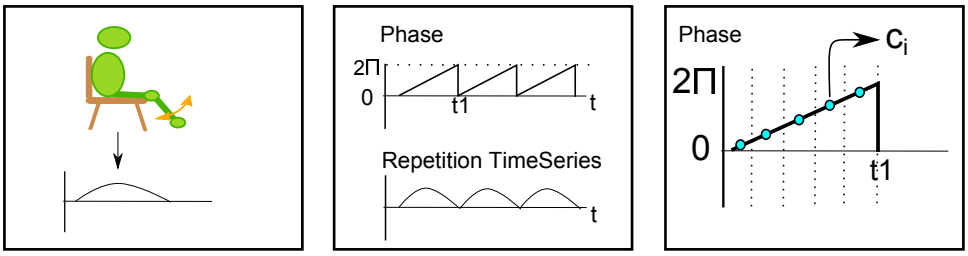


Figure 4.5: Illustrates the phase configuration and the basis configuration of the DMP-based approach

Feature Extraction

We use DMPs (See Sec. 3.5) with limit cycle attractors to model each *repetition set* of the patient population data. A DMP model is trained separately for each DoF, i.e., joint angle. Therefore in this thesis, for each *repetition set* five DMP models are trained. An overall DMP-model is trained for the healthy population, using the entire timeseries available in the healthy population. To provide enough periods for the DMP to correctly estimate the weights of the basis functions in the patient population, w_i, the data for the *repetition set* of each session is considered. The phase variable, ϕ, as given in 3.16, has a linear relationship with time and is equal to 0 at the beginning of each *repetition timeseries*, and is equal to 2π at the end of each *repetition timeseries*. To provide the synchronization between all DoFs, the same phase variable, ϕ, is used for modelling each DoF. The phase is divided into n equal segments for each *repetition timeseries*, where n is the number of the basis functions. The center of each segment is considered as center of the corresponding basis function, c_i (See Fig. 4.5). The amplitude r is set to 1. The values of the basis functions are dependant on the value of the phase and are obtained using Eq. 3.15. For each DoF, the joint angle position, velocity, and acceleration are available. The weights of the basis functions are computed by utilizing least squares for Eq. 3.13.

The weight of the basis functions, w, for all DoFs are concatenated, and are then considered as the feature vector.

$$\mathbf{v} = [w_1 \quad w_2 \quad \dots \quad w_K], \tag{4.14}$$

where K is the number of basis functions used to model the *repetition timeseries*, n, times the number of joint angles, i.e., $K = n \times num_{JointAngles}$. In this thesis, the DMP models consist of ten basis functions, $n = 10$ [1].

[1]From our observations the results do not change significantly when n is between 10 and 25

Measure of Performance

The feature vectors are calculated for the healthy population and the patient population using the procedure described in Sec. 4.2.2. The Euclidean distance is calculated between the feature vectors obtained for each *repetition set* of each patient, and the weights of the overall model of the healthy population. The calculated distance is considered as the measure of performance, Δ, for the patient's *repetition set*. The formulation becomes

$$\mathbf{v}_P = [w_1, w_2, \dots w_K], \tag{4.15}$$

$$\mathbf{v}_H = [w_1^{'}, w_2^{'}, \dots w_K^{'}], \tag{4.16}$$

$$\Delta = Norm(\mathbf{v}_P - \mathbf{v}_{H_i}), \tag{4.17}$$

where $\mathbf{v}_P$ is the feature vector calculated for one *repetition timeseries* of a specific patient P, $\mathbf{v}_H$ is the feature vector for a *repetition timeseries* in the healthy population, K is the number of elements in the feature vector, w_j is the weight of the jth basis function, Δ is the measure of performance for a *repetition set*, and is computed as the Euclidean distance between the feature vectors of the patient's *repetition timeseries* and the feature vectors of the healthy population's model. Figure 4.6 illustrates the procedure for calculating the measures of performance using the DMP-based approach.

4.2.3 Kernel-Based Approach

In the kernel-based approach, the sample distribution of the patient's motion is compared to the sample distribution of the healthy population's motion. We assume that as patients improve, the sample distribution of their motion becomes more similar to that of the healthy population. In this approach, temporal information is not considered, only spatial variabilities are used for comparison.

Measure of Performance

For estimating patient progress during physiotherapy rehabilitation we use the two sample kernel method as described in Sec. 3.6. We assume that the p samples of each *repetition timeseries* for a specific patient collected at each time step are drawn independently from the distribution P and the samples r of all *repetition timeseries* of the healthy population are drawn independently from distribution R. The temporal dependence of the samples is neglected in this approach and the timeseries data is considered as a concatenation of samples r or p. The measure of performance for each *repetition timeseries*, δ, can be obtained

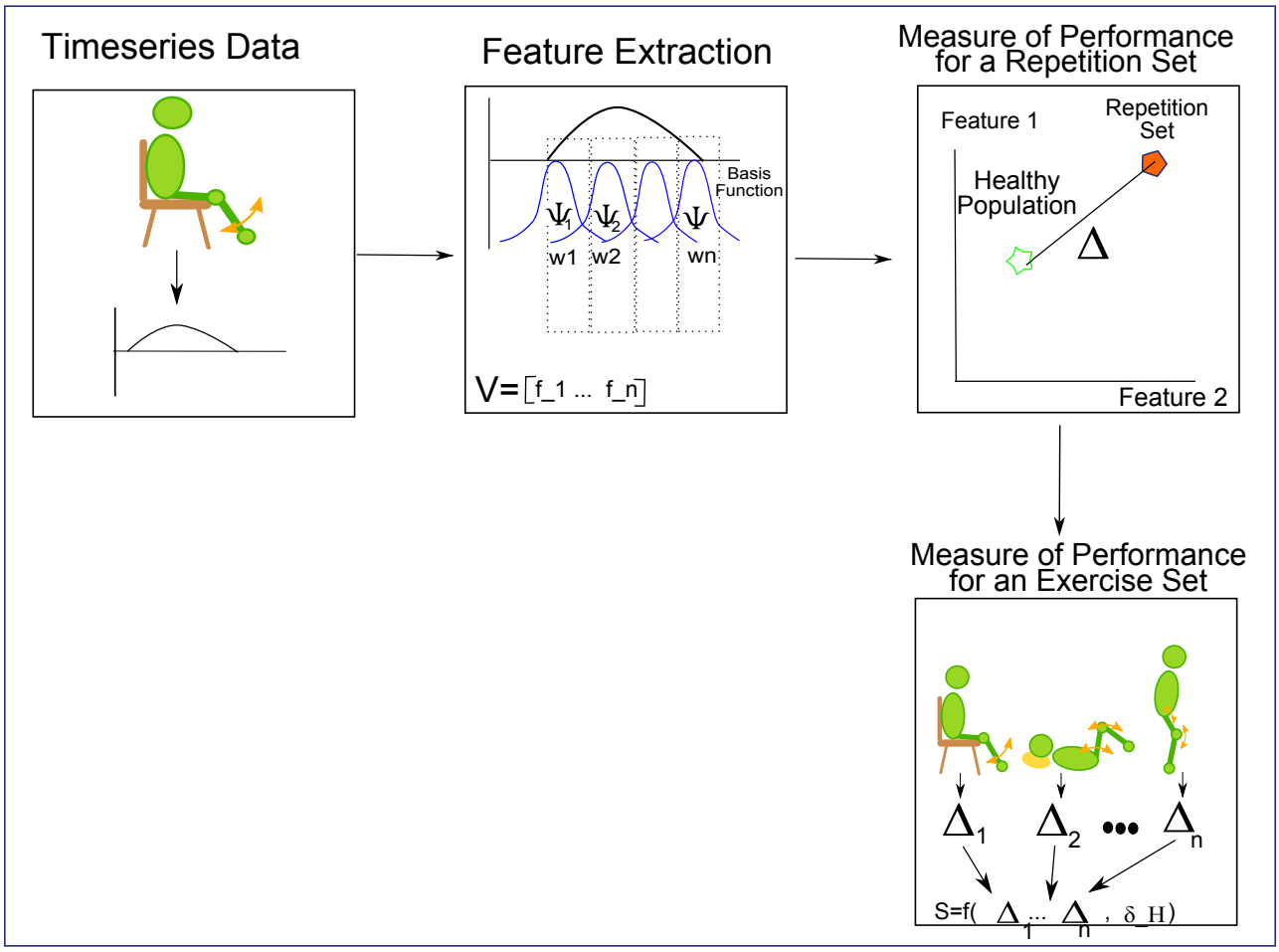


Figure 4.6: DMP Based Approach: A DMP model is trained for each *repetition set* of the patient population. The weights computed for each basis function are considered as features and the Euclidean distance between the features from the healthy population model and the features calculated for the *repetition set* is considered as the measure of performance for the *repetition set*. The measure of performance for an *exercise set* is calculated.

by measuring the similarity between distributions P and R, i.e., the similarity between the patient's performance in one *repetition timeseries* can be compared to a *repetition timeseries* from the healthy population by comparing the distributions P and R.

We define the samples for each *repetition timeseries* as the vector of joint angle positions, velocities, and accelerations at each time step.

$$p_i = \left[q_{1_p}(i), q_{2_p}(i), ..., \dot{q}_{1_p}(i), ..., \ddot{q}_{1_p}(i), ...\right]$$

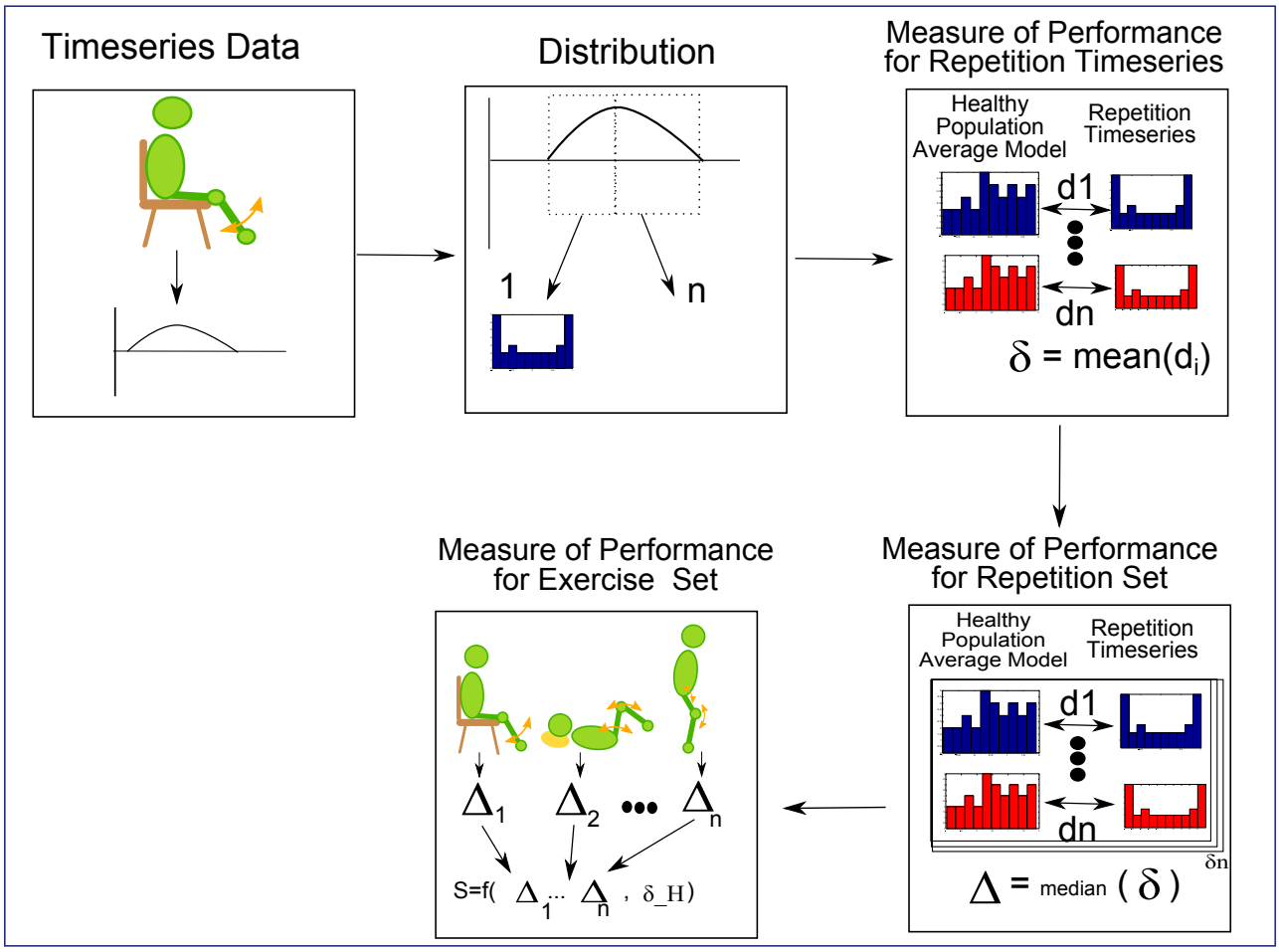


Figure 4.7: Kernel Based Approach: The timeseries is divided into n subsections, and the distribution of each subsection is calculated. The measure of performance for each *repetition timeseries* is obtained by calculating the similarity of the healthy populations' distribution and the distribution calculated for the *repetition timeseries*. Then the measures of performance for a *repetition set*, and a *exercise set* are calculated.

$$r_i = [q_{1_h}(i), q_{2_h}(i), ..., \dot{q}_{1_h}(i), ..., \ddot{q}_{1_h}(i), ...] \,,$$

where r_i is the ith sample drawn from healthy population's distribution R, p_i is the ith sample drawn from the patient's distribution P, $[q_{p_1}(i), q_{p_2}(i), ..., \dot{q_{p_1}}(i), ..., \ddot{q_{p_1}}(i), ...]$ is the patient population's vector of joint kinematics at time i for one *repetition timeseries*, and $[q_{h_1}(i), q_{h_2}(i), ..., \dot{q_{h_1}}(i), ..., \ddot{q_{h_1}}(i), ...]$ is the healthy population's vector of joint kinematics at time i for one *repetition timeseries*.

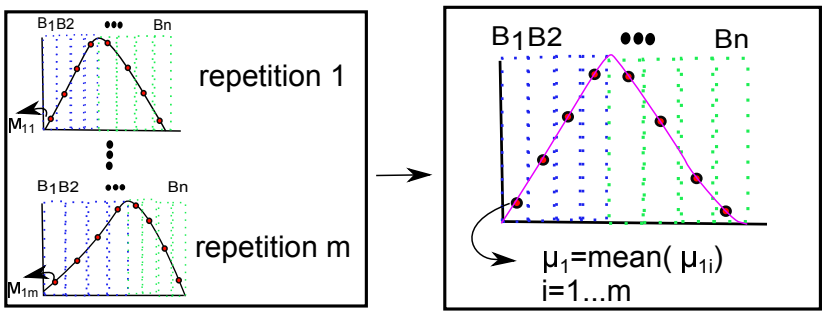


Figure 4.8: The procedure for obtaining an average model for the *repetition timeseries* of the healthy population. The first half and the second half of the motion are divided into equidistant segments. The mean of each segment is calculated for each *repetition timeseries* and the average of the segment averages over all *repetition timeseries* are interpolated to yield the average *repetition timeseries* model for the healthy population.

To reduce calculation complexity, each *repetition timeseries* of the healthy population is divided into 8[2] segments, where the first half and the second half of the motion are divided into 4 equal segments. The averages of of each segment, μ_{B_i}, over all *repetition timseries* in the healthy population data are calculated:

$$\mu_{B_{ij}} = mean(\gamma_{B_{ij}}) \tag{4.18}$$

$$\mu_{average_{B_i}} = mean(\mu_{B_{ij}})_{j=1\ldots m} \tag{4.19}$$

where $\mu_{B_{ij}}$ is the average of the ith segment for the jth *repetition timeseries*, $\gamma_{B_{ij}}$ is the kinematic vector of the timeseries of the ith segment for the jth *repetition timeseries*, and $\mu_{average_{B_i}}$ is the average of averages over all *repetition timeseries*, i.e., $1 \ldots m$.

The average of the segments are interpolated to obtain an average model of the healthy population's motion for each joint angle. This procedure results in one average *repetition timeseries* for each exercise in the healthy population. The interpolation reduces the calculation complexity and the time complexity of the approach considerably without significantly affecting the performance of the algorithm. Fig. 4.8 depicts the procedure for obtaining the average healthy population model.

Due to the way MMD is formulated and the way the samples are defined in this thesis, the time evolution of the motion is neglected. Therefore, certain motions that have the same joint angle position, velocity, and acceleration values with different evolution over

[2]Values between 8 and 20 result in similar average time series, 8 is chosen to reduce the calculation complexity.

time result in a small MMD measure. To avoid comparing different stages of motion, each *repetition timeseries* ω is further segmented into *subgroups* , ϕ_ω, based on the joint angle position of the most variant degree of freedom. The distance measure, δ_{k_ω}, for the kth *subgroup*, ϕ_{k_ω}, is obtained by calculating the MMD measure for the samples of the corresponding subgroup in the patient's *repetition timeseries* and the healthy population's average model.

$$\delta_{k_\omega} = MMD(P_{\phi_{k_\omega}}, R_{\phi_{k_\omega}}) \tag{4.20}$$

$$\delta_\omega = \{\delta_{1_\omega}, ...\} \tag{4.21}$$

The measure of performance, δ, for the *repetition timeseries*, ω, is calculated as the mean of the distance measures, δ_ω, calculated for all *subgroup*s.

$$\delta = mean(\delta_\omega) \tag{4.22}$$

The measure of performance for a *repetition set* is calculated using Eq. 4.9.

4.3 Measure of Performance for a Combination of Exercises

The distance Δ_Ω describes the patient's performance for one exercise (i.e. Ω_j) in each session which is obtained using one of the approaches described in Sec. 4.1. However, there are multiple exercises performed in each physiotherapy session (i.e. Γ) that need to be considered together for overall patient progress assessment. In this section, we assume that the measures of progress for each *repetition timeseries*, δ, are obtained using one of the approaches described in Sec. 4.1, and the proposed approach for estimating the overall measure of performance, S, for multiple exercises in one session is presented.

Quality and quantity are the two factors that affect scoring an exercise. We use the variance of distance measures δ calculated for the healthy population as a measure of exercise difficulty. Based on our observations, the features computed from exercises performed by the healthy population have larger variances when the exercise is more difficult. We therefore assume that the distance measures of the healthy population have larger variance for more difficult exercises.

The measure of performance for a *repetition timeseries* is calculated for every *repetition timeseries* of the healthy population data according to each method's formulation for δ (See Sec. 4.1), and is considered as the comparison reference. The healthy population distance

measure vector $\boldsymbol{\delta}_{H_j}$ is the vector of the distance measures calculated for every *repetition timeseries* of exercise Ω_j in the healthy population data. The patient distance measures (Δ_{P_j}) are calculated for the *repetition set* of every exercise Ω_j in the *exercise set* Γ. The mean and standard deviation of $\boldsymbol{\delta}_{H_j}$ are considered as the measure of exercise difficulty:

$$\mu_{\boldsymbol{\delta}_{H_j}} = mean(\boldsymbol{\delta}_{H_j}) \tag{4.23}$$

$$\sigma_{\boldsymbol{\delta}_{H_j}} = std(\boldsymbol{\delta}_{H_j}) \ , \tag{4.24}$$

where $\mu_{\boldsymbol{\delta}_{H_j}}$ is the mean of the healthy population distance measure vector $\boldsymbol{\delta}_{H_j}$ and $\sigma_{\boldsymbol{\delta}_{H_j}}$ is the standard deviation of the healthy population distance measure vector $\boldsymbol{\delta}_{H_j}$.

We define the measure of quality for a *repetition set* of an exercise j performed by the patient as

$$Q_j = \frac{(\Delta_{P_j} - \mu_{\boldsymbol{\delta}_{H_j}})}{\sigma^a_{\boldsymbol{\delta}_{H_j}}} \ , \tag{4.25}$$

where Δ_{P_j} is the patient's distance measure for the *repetition set* of exercise j, and a is the index that penalizes Q based on the exercise difficulty, i.e., the larger a is the more important the exercise difficulty becomes. The best value for a was found to be 2, which can be interpreted as the inverted dispersion index [26]. A perfect performance over any *repetition set* Ω results in a value of zero for the overall distance measure ($\Delta_\Omega = 0$). The overall score for the patient in a specific session is calculated as the difference between the norm of the score resulting from a perfect performance and the norm of the weighted quality measures. The quality measure Q_j of an exercise j is weighted by its number of repetitions. The score of the patient for a given session is calculated using the following equation

$$S = \sqrt{\sum_{\Omega\in\Gamma}(\frac{n_\Omega}{\sum_{\Omega\in\Gamma} n_\Omega}\frac{\mu_{d_{H_\Omega}}}{\sigma^a_{d_{H_\Omega}}})^2} - \sqrt{\sum_{\Omega\in\Gamma}(\frac{n_\Omega}{\sum_{\Omega\in\Gamma} n_\Omega}Q_\Omega)^2} \ , \tag{4.26}$$

where Γ is the *exercise set*, Ω is an exercise in the Γ, n_Ω is the number of repetitions for exercise Ω, and Q_Ω is the quality measure calculated for exercise Ω using (4.25). The score S is formulated such that performing a difficult exercise in a session would improve a patient's score. Furthermore, we assume that exercises with more repetitions in one session are the main focus of that session and therefore, the quality measures Q are weighted by the number of repetitions for each exercise. The score S is defined as the difference between a perfect weighted quality measure and the patient's weighted quality measure hence progress is assumed to result in smaller values for this measure.

The healthy population's distance values are often small and have a small variance compared to patient data. In some cases the healthy population's distance measure variance $\sigma_{\boldsymbol{\delta}_{H_j}}$ becomes smaller than 1. In (4.25), the quality measure is normalized according to the healthy population performance. To avoid dividing the quality measure with a value less than 1, all the δ values are scaled uniformly for the healthy population and patient data such that all variance values of the healthy population's distance measures become greater than 1. The algorithm flexibility in defining any *exercise set* allows us to calculate the overall score for any arbitrary set of exercises.

4.4 Summary

In this chapter, two approaches for estimating progress in patient performance during rehabilitation were proposed, and several comparison approaches for estimating progress based on existing methods were provided. The proposed feature-based approach uses the statistical features calculated from the joint angle timeseries as its feature vector. After feature selection by LASSO or KW the weighted distance between the feature vectors of the patient data and the healthy population data are considered as the measure of progress. The proposed HMM-based approach uses the mean and variance of the HMM trained for each *repetition set* as its feature vector and after feature selection by LASSO or KW, the weighted distance of the feature vectors for the healthy population and the patients are considered as the measure of performance. Three comparison methods based on existing approaches were also formulated. In the DMP-based approach, the weights of the basis functions are used as features and the Euclidean distances between the patient data and the healthy population data are considered as the measure of progress. The kernel-based approach uses the timeseries of the data as distribution samples and calculates the difference between the healthy population and the patient population distributions as the measure of progress. This method does not consider the time evolution of the timeseries. The classifier-based approach uses the statistical features, selects features by LASSO or KW, and the classification criterion is considered as the measure of progress. Table 4.1 summarizes the attributes of each method. To allow an estimate of progress across differing sets of exercises, the overall measure of progress, S is proposed. The measure of progress is computed for an *exercise set* based on the distance measures computed from one of the above methods, and formulated such that more weight is given to the exercise with more repetitions in each session. Furthermore, the more difficult exercises have a larger effect on the overall score.

Table 4.1: Approach Comparison

Method	**Feature Extraction**	**Feature Selection**	**Advantages**	**Disadvantages**
Feature-Based Approach	Statistical Features	LASSO/KW	Calculation simplicity	Highly affected by variables that affect features
HMM-Based Approach	HMM-based Features	LASSO/KW	Generates a model from multiple repetitions	Requires multiple samples for training, can have a large calculation complexity
Kernel-Based Approach	—	—	No training required, capable of detecting small perturbations	Does not consider time evolution
DMP-Based Approach	DMP-based Features	—	Generates a model of the timeseries	Highly affected by variability in segments
Classifier-Based Approach	Statistical Features	LASSO/KW	Common approach	Training data required from on both healthy and patient population

Chapter 5

Synthetic Data Experiments

Clinical data may contain multiple sources of variability , e.g., noise, sensor motion, patient progress, and etc. Furthermore, there are often no expert labels or quantified scores of performance available for each exercise repetition. The above-mentioned reasons make it difficult to evaluate the performance of the proposed approaches under controlled conditions and with known ground truth. Therefore, in this chapter, we evaluate the results of the different approaches using a synthetic dataset. The approaches will be subsequently evaluated on a clinical dataset in Chapter 6.

5.1 Generation of the Synthetic Dataset

The synthetic data set is designed to evaluate the proposed approaches with a known ground truth signal, i.e., the proposed approaches are evaluated based on a dataset with a known rate of progress and variability. The synthetic data is generated for the position, velocity, and acceleration of one joint angle and for two exercises. The joint angle position is generated as an upward bell curve for the first exercise, and as a downward bell curve for the second exercise. The joint angle velocity and acceleration are calculated by differentiation.

We assume that the healthy range of motion is always above 90% of the possible range of motion and that the execution time is between $1 - 1.5$ seconds. We allow random temporal and spatial variability between the ranges specified. We assume that the second exercise is the harder exercise, thus we allow a wider range of temporal and spatial variability for the data generated based on this exercise. The range of motion considered for the hard exercise for the healthy population is above 70% of the complete range of motion and completion time for this exercise is between $1 - 2.5$ seconds.

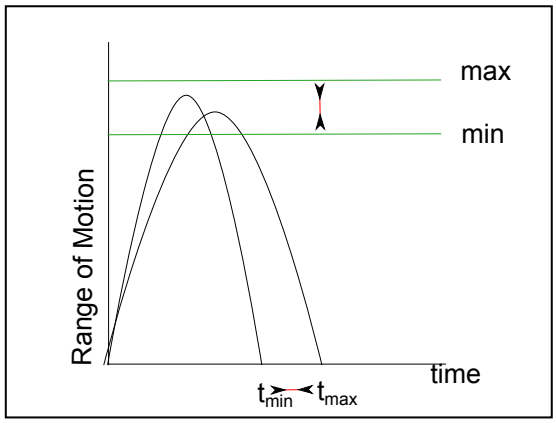


Figure 5.1: The synthetic timeseries of the healthy population and the patient population are constructed to model human motion variability.

We assume that the patients start with a small range of motion on their first session, $20\% - 25\%$ of the full range of motion for the easy exercise and $10\% - 15\%$ of the full range of motion for the hard exercise, and gradually improve through five sessions until they reach $80\% - 85\%$ of the full range for the easy exercise and $50\% - 55\%$ of the full range of motion for the hard exercise. The patients' execution times considered for the exercises are high for the first session, $5 - 6$ seconds for the hard exercise and $4 - 5$ seconds for the easy exercise, and gradually decrease to smaller values during the five sessions, where on the fifth session the execution time is between $2 - 3$ seconds for the easy exercise and $3 - 4$ seconds for the hard exercise. For different repetitions, we allow random spatial and temporal variability in the generated data within the specified ranges. Figs. 5.1 and 5.2 depict the design of the synthetic data.

The synthetic healthy dataset consists of 20 healthy members with 10 *repetition timeseries* each. The synthetic patient dataset consists of 20 patients where the number of *repetition timeseries* for each individual is chosen randomly between five and eight. The variabilities caused by the individual differences between different subjects are neglected. We consider this data as our ground truth, i.e., the data contains only random spatial and temporal variabilties and variabilities caused by the progress.

For each approach, the overall score, S, is calculated for all the patients for every session. Since the 20 generated patients are not different, their scores on each session are averaged to obtain an overall score ξ which is used for comparing the approaches with each other.

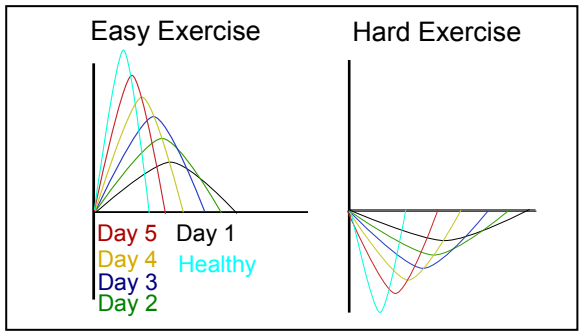


Figure 5.2: Synthetic data modelling patient progress from day 1 to day 5. When the motion performance of the patients improves, the range of motion and the speed increase. In the synthetic dataset, the healthy population always has a larger range of motion and speed compared to the patient population.

$$\begin{aligned} \xi_{i_{method}} &= mean(\{S_{i_{k\,method}}\}) \\ \Xi_{method} &= [\xi_{1_{method}}, ..., \xi_{l_{method}}] \\ k &= 1, ..., N \\ i &= 1, ..., l, \end{aligned} \tag{5.1}$$

where $S_{i_{k\,method}}$ is the overall score calculated using approach *method* for the kth patient on the ith session, N is the number of patients in the dataset, l is the number of sessions for the patients, $\xi_{i_{method}}$ is the overall performance of approach *method* for session i, Ξ_{method} is the vector of score averages over all sessions for approach *method* and is used for comparing the three approaches with each other.

5.2 Results

The HMM-based approach, feature-based approach, kernel-based approach, DMP-based approach, and SVM based approach [1] are applied on the synthetic data. When all the data is available to all methods, the correlation in average patient scores between the three methods is over 97%. Therefore, we consider this value of Ξ for each method as the ground truth, Ξ_{GT}, and to investigate the effects of other sources of variabilities we always compare

[1] The SVM is chosen for comparison with the proposed approaches because of the results obtained with the clinical data this information can be found in chapter 6

the average results with the ground truth, Ξ_{GT}. We consider the variability-free, full data case as the ground truth and investigate how correlations between this ground truth and the results from the various approaches are affected under the following conditions: (1) noisy data, (2) poor segmentation, (3) temporal variability, and (4) incomplete data. We provide both Normalized Mean Square Error (NMSE) analysis and correlation analysis. We formulate the NMSE as:

$$NMSE = \frac{Norm(\xi_{method}, \Xi_{GT})}{max(\Xi_{GT}) - min(\Xi_{GT})} \tag{5.2}$$

We depict $\log(NMSE)$ in the NMSE analysis. If the value of the *log* becomes zero it means that there is no difference between the obtained measure and the ground truth. If the value of *log* becomes smaller than zero, it indicates that the difference between the obtained score and the ground truth is less than range of the trend of progress in the ground truth. If the value of the *log* becomes greater than 0 it means that the difference between the obtained score and the ground truth is greater than the trend of progress in the ground truth.

In the first set of tests, the effect of noise on the the overall scores generated by each method is investigated. A noise signal with a standard uniform distribution is added to the joint angle position of the patient dataset. The range of the noise is between 1% to 80% of the range of the corresponding joint angle position. The effect of noise on the overall score of all the approaches using the correlation and NMSE analysis is depicted in Figs. 5.3a and 5.4a. As can be seen from the figures, the HMM-based approach is the least affected by the noise in the correlation analysis and MSE analysis. The feature-based and the SVM-based approaches are highly affected by the noise based on the results of the correlation and the NMSE analysis. The DMP-based approach is less affected compared to the SVM-based approach. The DMP based approach is less affected than the HMM-based approach in the MSE analysis. The reason for the superior performance of the HMM-based approach and the DMP-based approaches is that the HMM-based approach generates a model based on a *repetition set* and therefore it considers the average of all *repetition timeseries*. The DMP-based approach also models the timeseries using Gaussian kernels, and therefore is less affected by the noise compared to the feature based approach and the SVM-based approach. The kernel-based approach performs better than the SVM-based approach, the DMP-based approach, and the feature-based approach in both the correlation analysis and the NMSE analysus. The kernel-based approach is less affected by the noise particularly when the noise added is up to 30%. The reason for this observation is that adding noise with zero mean and a small range of percentage does not affect the sample distributions of the patient population and the healthy population significantly. The feature-based approach and the SVM have the poorest performance because the features of the feature

vector are directly affected by the noise and therefore the noise directly affects the results of the measures of performance. The SVM based approach is highly affected in the NMSE analysis because the captured trend of progress in its ground truth is very small.

In the second set of tests, the effect of poor segmentation on the results of the five approaches is investigated. The poor segmentation is modelled by adding points with a constant value to the end of the joint angle position in the patient dataset. The number of points added to the end of the time series varies from 1 to the length of the corresponding *repetition timeseries*, which by design is 30 points on average. The effects of poor segmentation are illustrated in Figs. 5.3b and 5.4b. As can be seen from the figures, the feature-based approach and the kernel-based approach are the least affected, while the HMM-based approach is affected when the number of points are significant enough to alter the states (middle points in Figs. 5.3b and 5.4b). For this test we have assumed that the segmentation happens after the patient has returned to the initial posture. Therefore, the value of the inaccurately segmented point is close to zero. Values close to zero do not change the sample distribution due to the formulation of MMD. However if the inaccurate segmentation adds points with values other than zero to the timeseries, the kernel-based approach will be more affected. The SVM is also affected in the NMSE analysis because the changes in the obtained score are much larger than the trend of progress in the ground truth. The DMP-based approach is affected the most, this happens because the DMP-based approach divides the timeseries into equal segments and trains the weight of the basis function for each segment. Adding extra points to the end of the times series influences these segments significantly and changes the stage of the motion each segment corresponds to and therefore the estimated measure of performance is highly affected.

In the third set of tests, the effect of scaling the length of the timeseries in the patient dataset is analysed. The duration of the *repetition timeseries* for the patient data is increased from 10% to 200% of its original duration. The effect of time scaling is shown in Figs. 5.3d and 5.4d. As can be seen from the results, most approaches are not significantly affected by this variability except for SVM in the NMSE analysis where the range of the trend of progress is small and any change in the score results in large differences.

In the fourth set of tests, the effects of an incomplete *repetition timeseries* on the overall results are investigated. In this test, the percentage of the available time series is changed from 10% to 200%, where the percentages more than 100% correspond to the poor segmentation of two timeseries and the percentages which are less than 100% correspond to incomplete motion or early segmentation. The results of this test are depicted in Figs. 5.3c and 5.4c. Not being affected by this variability is a limitation because the patient could perform only half of the exercise and receive the same score as when she/he performs the entire exercise. The feature-based and SVM based approaches are not much affected

Table 5.1: Synthetic Analysis Summary

Method	Observation	Discussion
Feature-Based Approach	Highly affected by noise in both MSE and correlation analysis. Is not affected by other variabilities. Cannot detect incomplete motion.	Noise affects features directly. As long as statistical features remain the same can't detect the difference in the timsereis.
HMM-Based Approach	Affected by inaccurate segmentation in both MSE and correlation analysis if the inaccurate segmentation points are more than 60% of the length of the data. Detects incomplete motions.	The large segmentation error causes a change in states. The HMM models the timeseries and can detect incomplete timeseries.
Kernel-Based Approach	Cannot detect incomplete timeseries.	Does not consider time evolution.
DMP-Based Approach	Is highly affected by inaccurate segmentation test in both MSE and correlation analysis. Is Affected by noise in the correlation analysis. Can detect incomplete timeseries	The weights correspond to a different stage of motion. Is modelling timeseries so can detect incomplete timeseries.
SVM-Based Approach [2]	Is highly affected by all variabilities in the MSE analysis. Is highly affected by noise in the correlation analysis.	The range of the trend of progress detected is very small.

[1] The KL and NB based approaches work very poorly on clinical data and therefore there are not considered in our analysis.

by this variability and cannot capture the temporal information of the timeseries. This is a limitation of the predefined features since they do not consider the timing and do not include enough information to capture the difference between a complete timeseries and an incomplete timeseries. The kernel-based approach is also incapable of capturing such a change in the timeseries very well since the temporal information does not affect its formulation, and therefore is not affected by this test. However, the HMM constructs a statistical model of the timeseries and is significantly affected by the incomplete data. Therefore, the HMM is superior to the other approaches in this test, because the significant difference in the shape of the timeseries affects its result. The DMP-based approach also models the timeseries and can capture the difference when the difference between the complete timeseries and the incomplete time series becomes significant. Table 5.1 summarizes the observations for each approach in the synthetic analysis.

As mentioned earlier, the exercise regimen differs from one session to the other. The fifth test is designed to investigate the effects of set sparsity when data of hard or easy

exercises are not available in one session. Fig. 5.5 shows the values of the average overall score, Ξ, for each exercise when the hard or the easy exercises are not available. It can be seen from the results that set sparsity results in jumps and inconsistencies in the overall score. As can be seen from the figures, when only the hard exercise is available the scores are higher compared to when only the easy exercise is available in 4 out of 5 cases. If on any day the easy or the hard exercise is not available, this would result in a jump compared to when all the data is available. This is a limitation of our exercise difficulty formulation.

As the patients improve they perform more difficult exercises. However, since these exercises are difficult, patients often have a poorer performance compared to when performing the easy exercises. If we do not consider the exercise difficulty, the harder exercises are only assessed based on the performance and therefore will result in a poorer score. The sixth test is designed to investigate the effects not considering the exercise difficulty (See Fig. 5.6). As can be seen from the figure, when the quality measure, Q, is not normalized by the exercise difficulty, all the approaches assign a poorer score to the harder exercises.

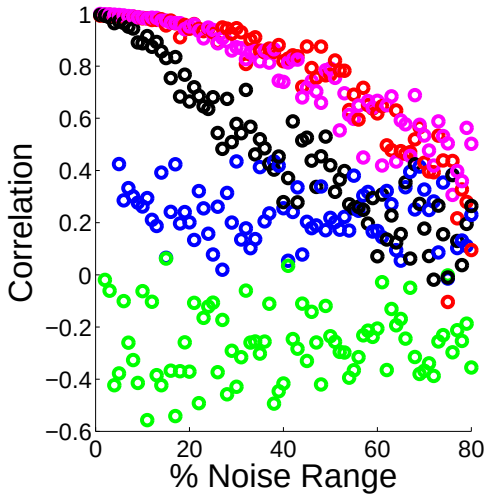
The eighth test is designed to investigate whether the proposed and the existing approaches can detect progress due to changes in the smoothness of the motion. We generate data for two synthetic patients, one with pauses during their motion and one with a shaky motion (See Fig. 5.7).

We assume that each patient starts their rehabilitation with 80% shaky motions and 20% smooth motions, and in their last rehabilitation day the patients only perform smooth motions. Figs. 5.9 and 5.8 show that all the approaches can capture this progress. The kernel-based approach detects this type of progress better because it has a larger $\frac{max(S_{Pause})-min(S_{Pause})}{max(S_{GT})-min(S_{GT})}$ and $\frac{max(S_{Shaky})-min(S_{Shaky})}{max(S_{GT})-min(S_{GT})}$.

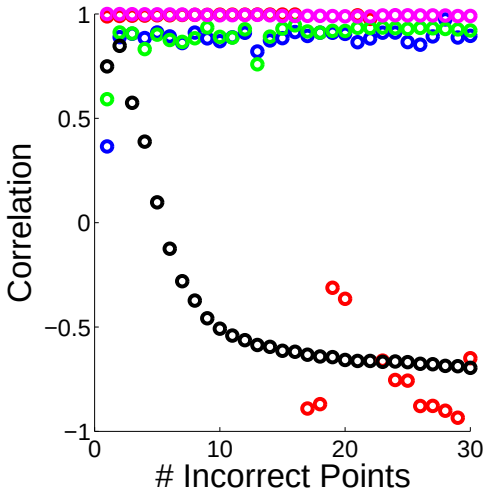
5.3 Summary

In this chapter, the proposed approaches and the existing approaches were applied to a synthetic dataset. The HMM-based approach is superior to the rest of the approaches in presence of noise, and segmentation inaccuracy. The SVM-based approach and the feature-based approach, which use statistical features obtained directly from the time series, are highly affected by the noise. The DMP-based approach and the HMM-based approach can detect when the motion timeseries is incomplete and therefore are superior to the rest of the approaches in terms of capturing the temporal information of a timeseries. The time scaling test reveals that none of the approaches are dependant on temporal scaling. The kernel-based approach performs moderately well in presence of noise and performs well in the presence of inaccurate segmentation. The kernel-based approach does not include time

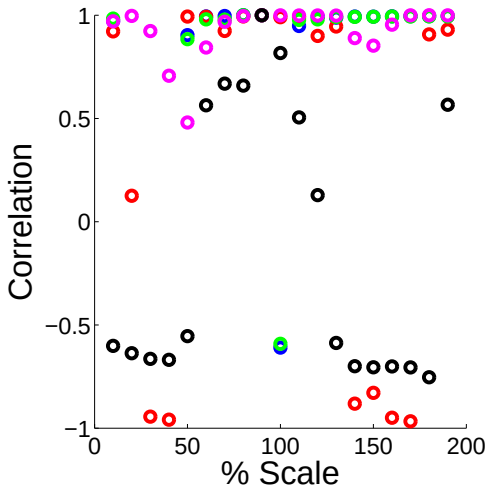
evolution of the timeseries, and therefore does not perform well in the incomplete timeseries test. The set sparsity test shows that a sparse *exercise set* results in inconsistencies and jumps in the overall score S. Normalizing the quality measure by the exercise difficulty is important to consider since otherwise harder exercises will be given a poorer score. All the approaches can capture the progress due to motion smoothness, with the kernel-based approach being superior in this test.



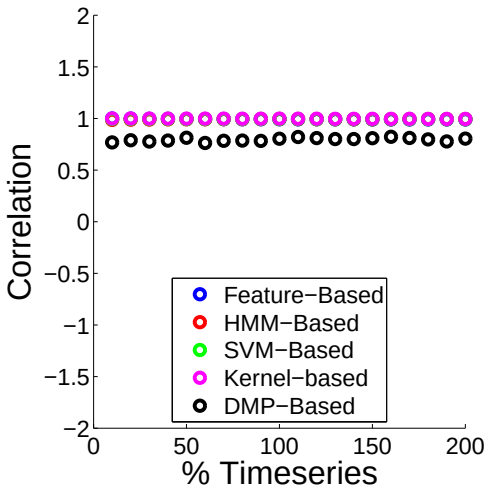
(a) Noise



(b) Poor Segmentation, HMM is affected in the middle points where the number of incorrect segmented points affect the states.

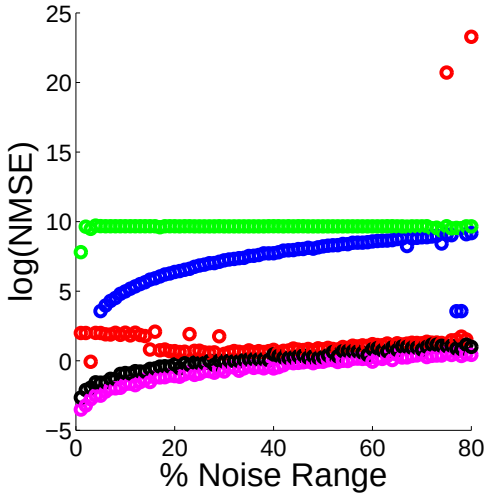


(c) Incomplete Timeseries, percentages less than 100% correspond to an incomplete motion or early segmentation. Percentages over 100% correspond to incorrect segmentation of two timeseries.

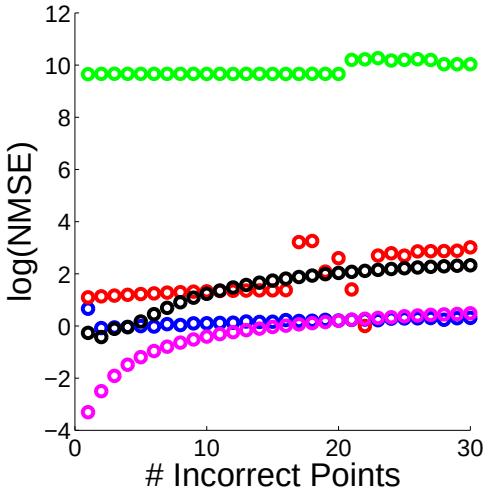


(d) Time Scaling

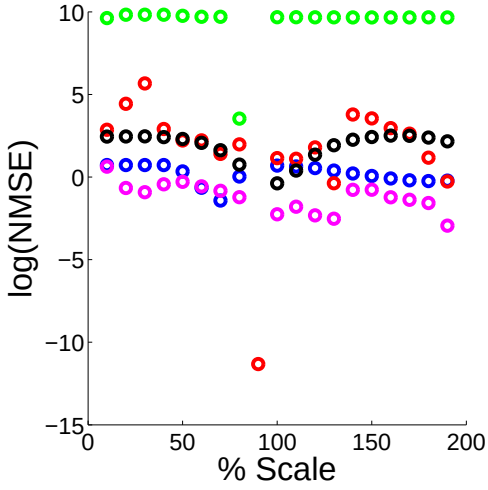
Figure 5.3: The effect of each source of variability on the correlation index between the average score, Ξ, of each method and the ground truth average score Ξ_{GT}.



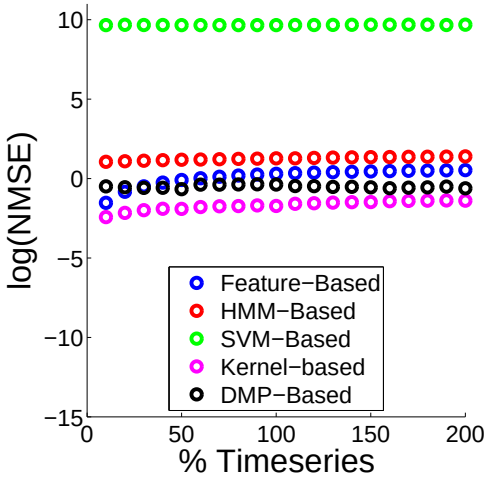
(a) Noise



(b) Poor Segmentation, HMM is affected in the middle points where the number of incorrect segmented points affect the states.

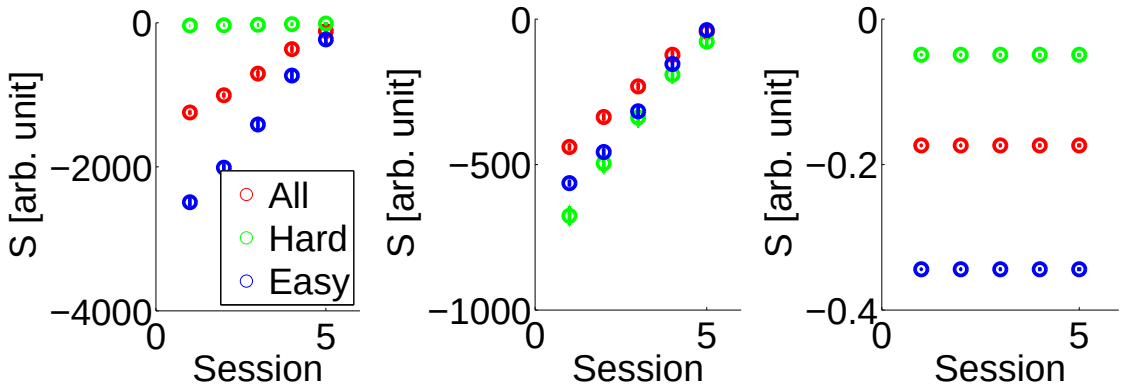


(c) Incomplete Timeseries, percentages less than 100% correspond to an incomplete motion or early segmentation. Percentages over 100% correspond to incorrect segmentation of two timeseries.

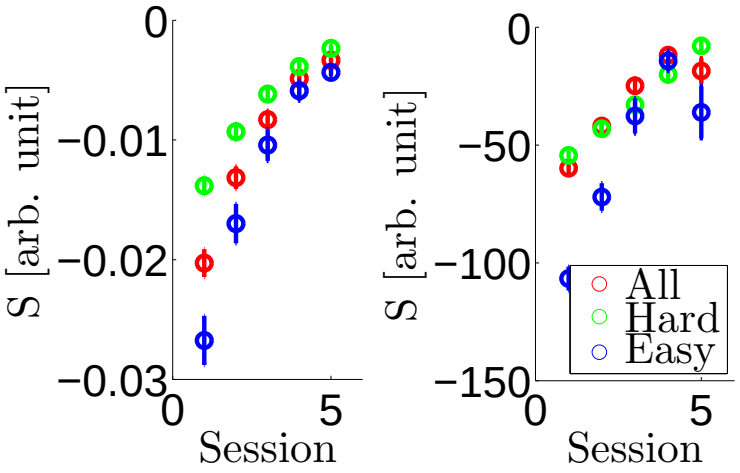


(d) Time Scaling

Figure 5.4: The effect of each source of variability on the Normalized Mean Square Error (NMSE) between the average score, Ξ, of each method and the ground truth average score Ξ_{GT}.

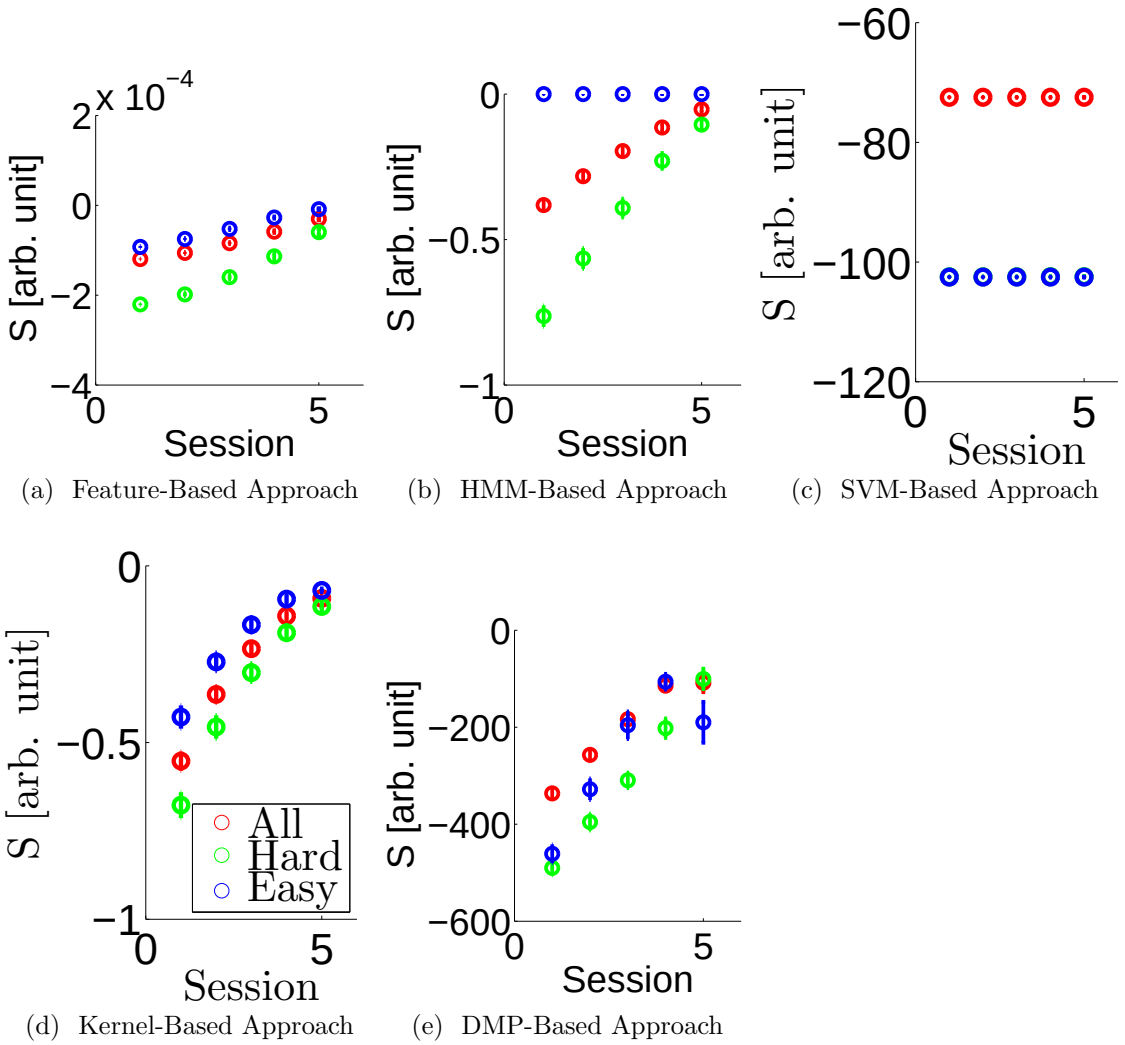


(a) Feature-Based Approach (b) HMM-Based Approach (c) SVM-Based Approach



(d) Kernel-Based Approach (e) DMP-Based Approach

Figure 5.5: The results of removing each set of exercise on the average overall score for each of the approaches. Note that scales are not comparable between the different approaches.



(a) Feature-Based Approach (b) HMM-Based Approach (c) SVM-Based Approach

(d) Kernel-Based Approach (e) DMP-Based Approach

Figure 5.6: The results of not including exercise difficulty in the score when the exercise contains hard and/or easy exercises.

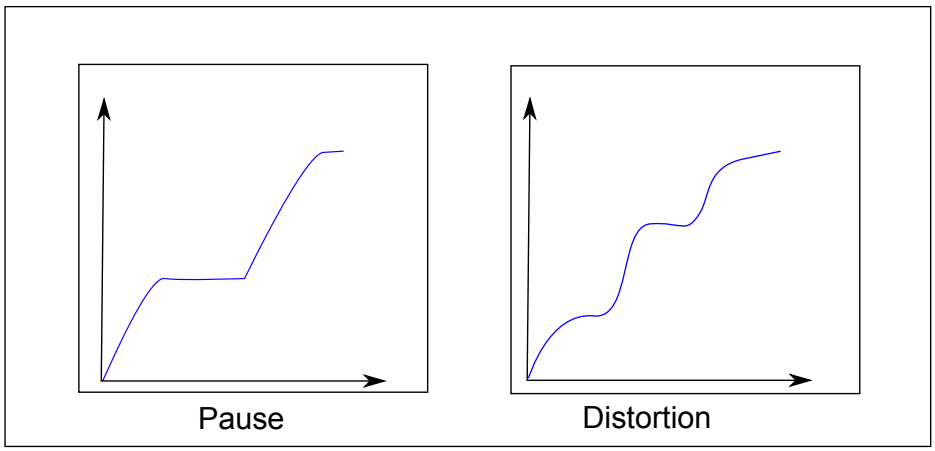
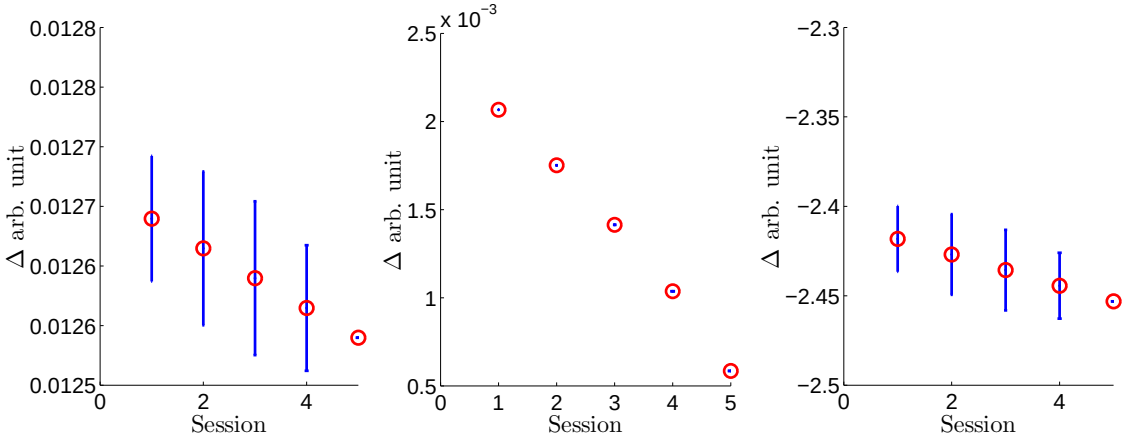
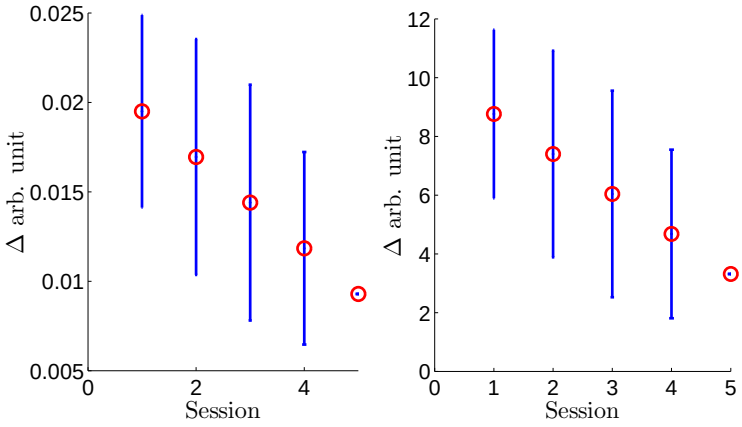


Figure 5.7: Instances of distorted motions.

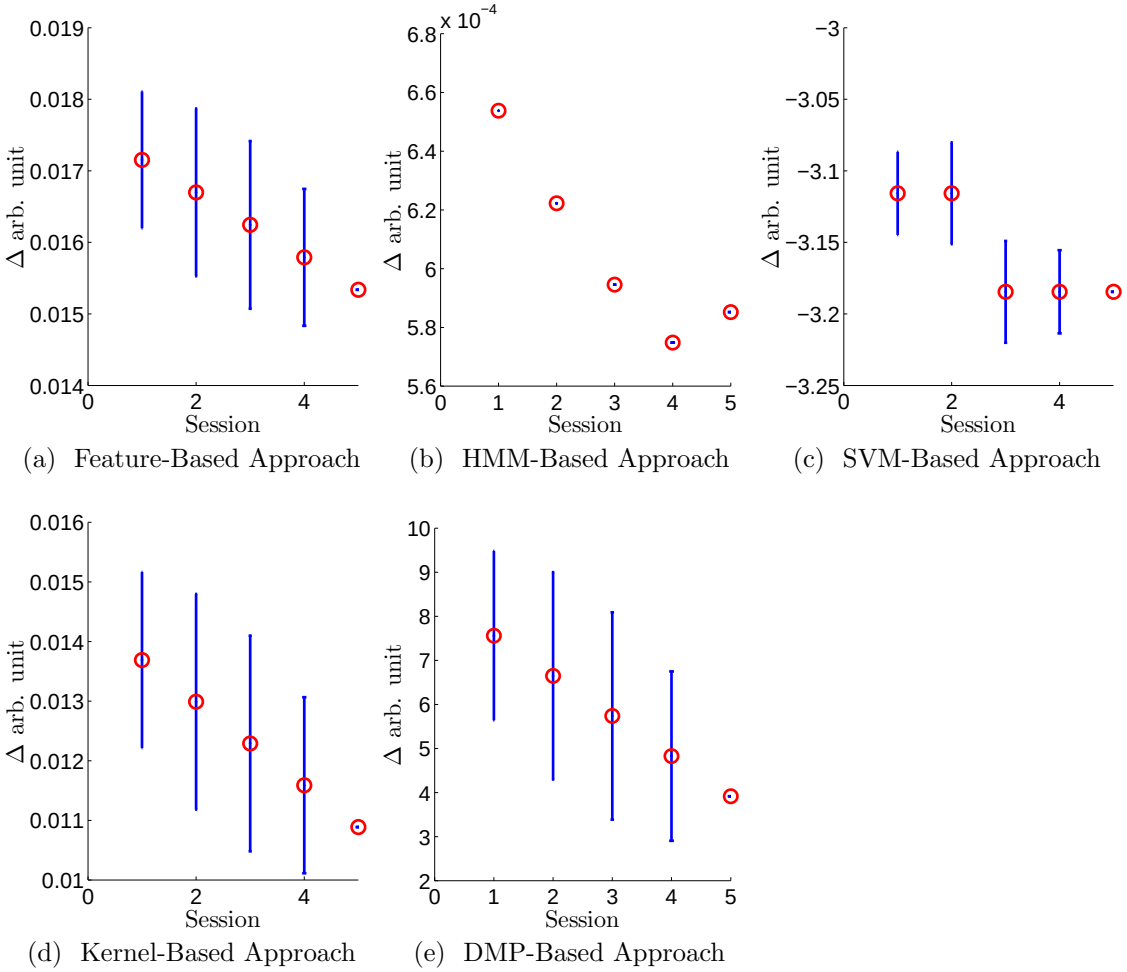


(a) Feature-Based Approach (b) HMM-Based Approach (c) SVM-Based Approach



(d) Kernel-Based Approach (e) DMP-Based Approach

Figure 5.8: The results of the scores for the synthetic patient with distortion in his/her motion.



(a) Feature-Based Approach (b) HMM-Based Approach (c) SVM-Based Approach

(d) Kernel-Based Approach (e) DMP-Based Approach

Figure 5.9: The results of the scores for the synthetic patient with pause in his/her motion.

Chapter 6

Experimental Evaluation

The feature-based approach, the HMM-based approach, the Kernel-based approach, and the DMP-based approach are evaluated on patient data collected at the Toronto Rehabilitation Institute. We focus on patients recovering from knee or hip replacement surgery. In this type of rehabilitation, the following exercises are commonly performed: knee extension/flexion while seated, knee and hip extension/flexion while supine, and squat[1]. Data of these exercises was also recorded for a healthy population. Furthermore, the feature-based and the HMM-based approaches are evaluated for two cases: 1) healthy population and a subset of patient data is available for training, and 2) only healthy population data is available for training. The healthy population data is only used to learn a reference model, and results are presented for the patient data. Moreover, each method's performance in capturing the patient progress is analysed and the advantages and the disadvantages of each method are discussed. A qualitative comparison between the results of the approaches and patients' medical chart is provided, and a quantitative comparison between the result of the approaches and the score measure obtained from the exercise regimen is presented. Furthermore, the results of the regression techniques are compared with the classifier-based approaches.

6.1 Data Collection and Pre-processing

Patient data was collected from eighteen inpatients during their rehabilitation at the Toronto Rehabilitation Institute [76]. Each patient performs one session per day and

[1]The patients do not perform the full squat but lower their torso only slightly (i.e., knee bend of 15 degs)

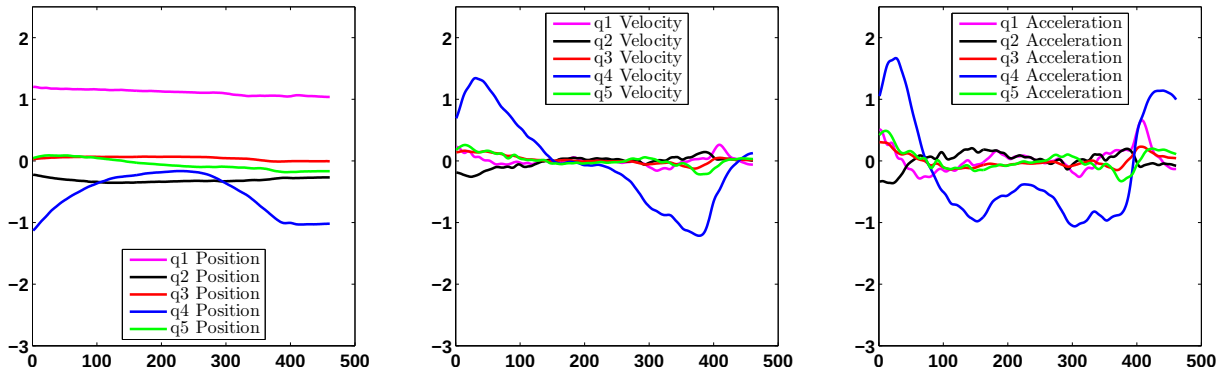
the session typically lasts between 45-60 minutes. The number of days a patient stays in the hospital varies from $4-12$ days and depends on the patient's needs and health status. The set of exercises specified by the therapist in each session differs between patients and sessions. Therefore the *repetition set* of one specific exercise is not available for every session. The healthy population data consists of 10 subjects (age: 23 ± 4.5) performing each exercise 10-20 times. The patient population tends to be elderly. Therefore, the healthy population performed the exercises slowly to consider the speed difference between the two populations.

The following results are discussed for the data of all patients. Plots and information for all 18 patients can be found in appendices D and E. A randomly selected subset of the patient data is used for feature selection (patient 1,2,3,5,6,7) and the remaining patient data is used for testing. We selected patients 2, 8, and 18 to graphically illustrate the results within this thesis. Patient 2 is an example of the patients included in the feature selection algorithms and shows gradual improvement during the rehabilitation process. Patients 8 and 18 are examples of patients whose data is not used for feature selection and plots of subjects 8 and 18 illustrate the performance of estimating progress for subjects whose data is unseen during training. Patient 18 is an example of a patient who shows a rapid progress in the course of their rehabilitation, while patient 8 is an example of a patient with a common duration of recovery. General information for patients 1-18 is summarized in Table 6.1 including any unique circumstances as recorded in the patient's medical chart. All patients were admitted to the in-patient program within a day of their surgery and commenced rehabilitation the following day.

Motion data is collected from the healthy and the patient populations using Shimmer sensors [15] mounted at the knee and ankle providing angular velocity and linear acceleration data (128 Hz). Position q, velocity $\dot{q}$, and acceleration $\ddot{q}$ of five joint angles consisting of knee flexion, knee rotation, hip flexion, hip abduction, and hip rotation are estimated from the sensor data based on a kinematic model and an Extended Kalman Filter [76], see Fig. 1.1. The Extended Kalman Filter results in a small phase lag between the velocity and the acceleration timeseries which is problematic for the DMP-based approach. Therefore, for the DMP-based approach the accelerations $\ddot{q}$ of the five joint angles of knee flexion, knee rotation, hip flexion, hip abduction, and hip rotation are calculated by differentiation. The data is filtered using a low-pass Butterworth filter with a cut-off frequency of 10Hz. For the HMM-based approach the data is downsampled to include a 100 points for each *repetition timeseries*. Fig. 6.1 illustrates the joint angles' position, velocity, and acceleration for one *repetition timeseries.*

Table 6.1: Patient Information. Patients in bold are used to illustrate the results in this chapter.

Patient ID (Age)	**Sex**	**BMI**	**Type of Replacement Surgery**	**Discharge Session**	**Special Conditions**
1 (62)	F	33.5	Hip Joint	8	Colostomy
2 (80)	**F**	**29.1**	**Hip Joint**	**11**	**Discharged and Readmitted**
3 (84)	F	28.2	Hip Joint	7	None
4 (62)	M	28.4	Both Knee Joints	3	None
5 (80)	M	24.9	Knee Joint	14	None
6 (68)	F	–	Knee Joint	12	Chronic Pain
7 (70)	M	28.1	Hip Joint	4	None
8 (59)	**F**	**42.1**	**Hip Joint**	**9**	**None**
9 (76)	F	31.2	Hip Joint	8	None
10 (81)	F	38.6	Hip Joint	7	None
11 (83)	F	25.5	Hip Joint	8	None
13 (86)	F	20	Knee Joint	10	Hip Fracture
14(48)	F	40.8	Knee Joint	11	None
15(48)	F	32.5	Knee Joint	3	None
16(85)	F	24.7	Knee Joint	6	None
17(83)	M	25.3	Hip Joint	8	Aortic Valve Replacement
18(86)	**F**	**32.9**	**Knee Joint**	**9**	**Pain**



(a) Joint Angle Position (b) Joint Angle Velocity (c) Joint Angle Acceleration

Figure 6.1: The joint angle positions, velocities, and accelerations for one *repetition time-series*.

6.2 Feature-Based Approach

In the following analysis, we first illustrate the distribution of the most informative kinematic parameters for the patient data and the healthy population data. We then discuss the results for training on both a subset of the patient data and healthy population data and finally investigate the performance of the proposed approaches when only healthy population data is available for training.

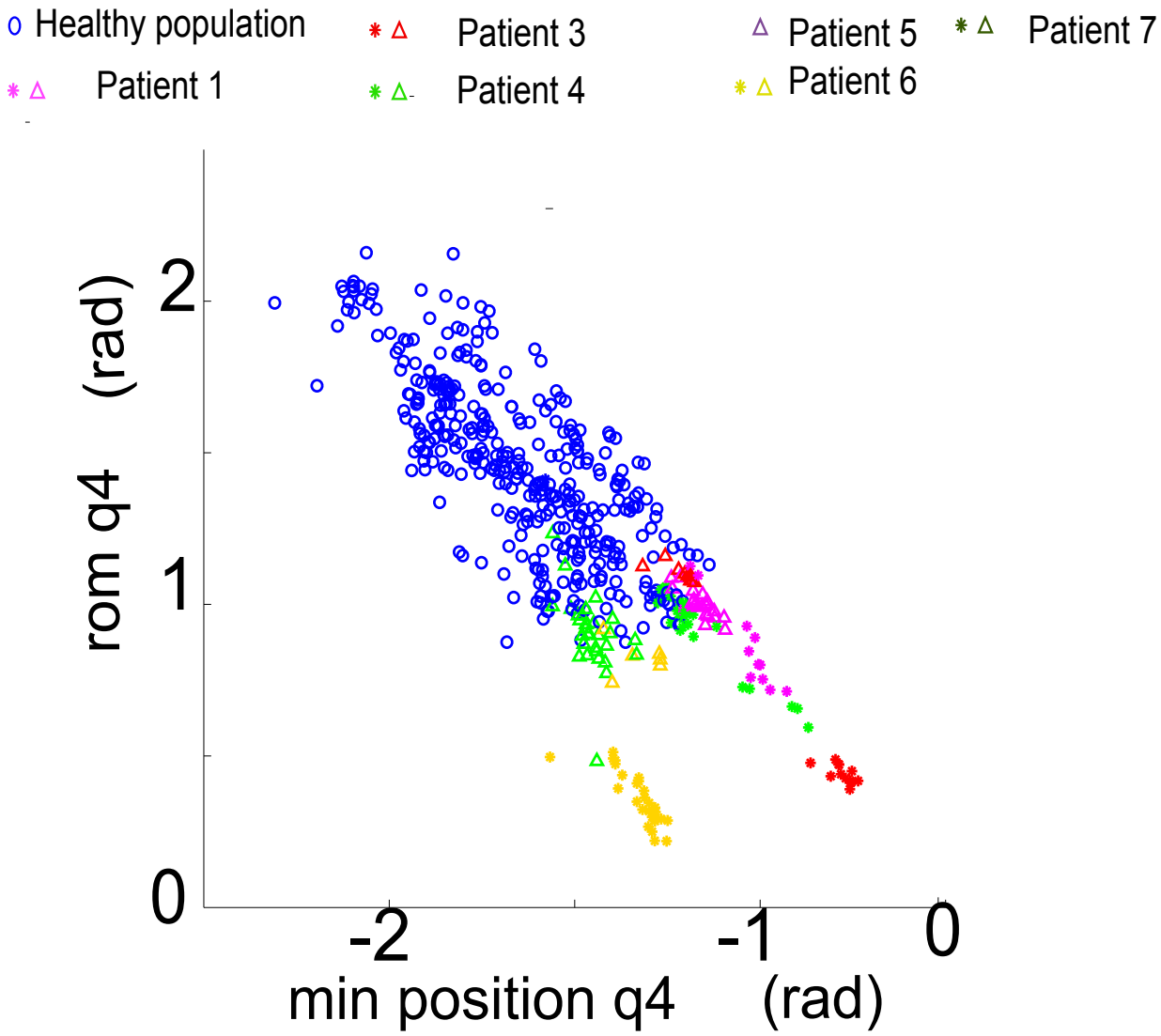


Figure 6.2: Knee Extension Flexion Exercise: Patient data and healthy population data are separable for the most informative features. The star indicates the first day the patients have performed the exercises and the triangle indicates the last day the patients have performed the exercises (only 1 session available for patient 5). q_4 is the joint angle corresponding to knee extension.

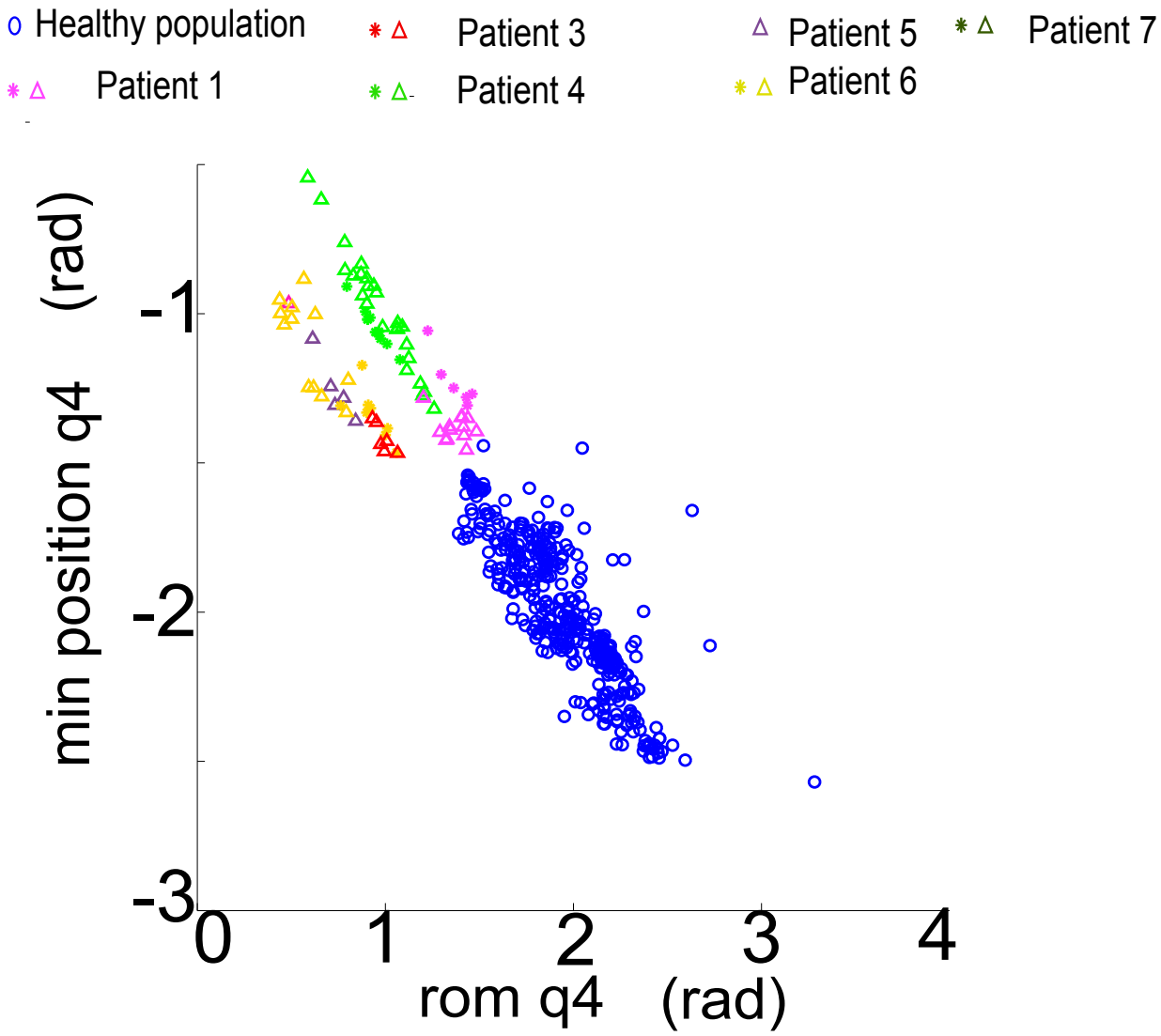


Figure 6.3: Knee Hip Extension Flexion Exercise: Patient data and healthy population data are separable for the most informative features. The star indicates the first day the patients have performed the exercises and the triangle indicates the last day the patients have performed the exercises (only 1 session available for patient 5). q_4 is the joint angle corresponding to knee extension.

6.2.1 Feature Selection

The number of *repetition timeseries* is greater than the number of features. Therefore, the LASSO technique described in section 4.1.1 is used for feature selection. Figs. 6.2, 6.3,

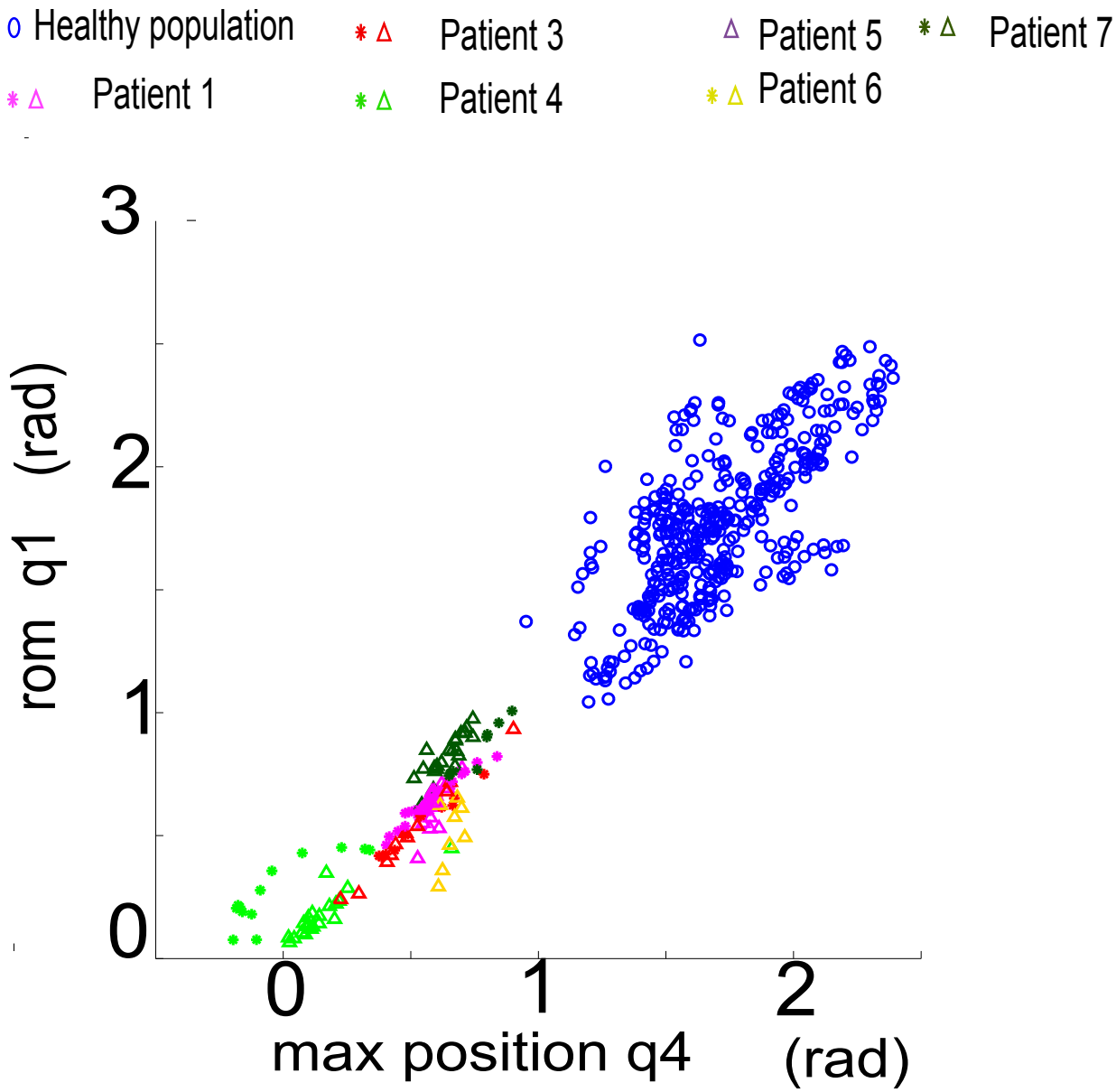


Figure 6.4: Patient data and healthy population data are separable for the most informative features. The star indicates the first day the patients have performed the exercises and the triangle indicates the last day the patients have performed the exercises (only 1 session available for patient 5). q_1 is the joint angle corresponding to hip extension and q_4 is the joint angle corresponding to knee extension.

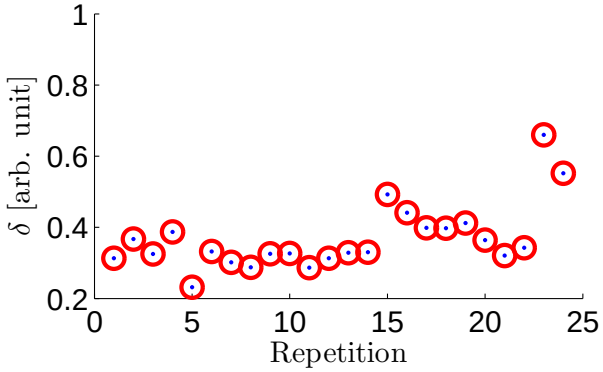
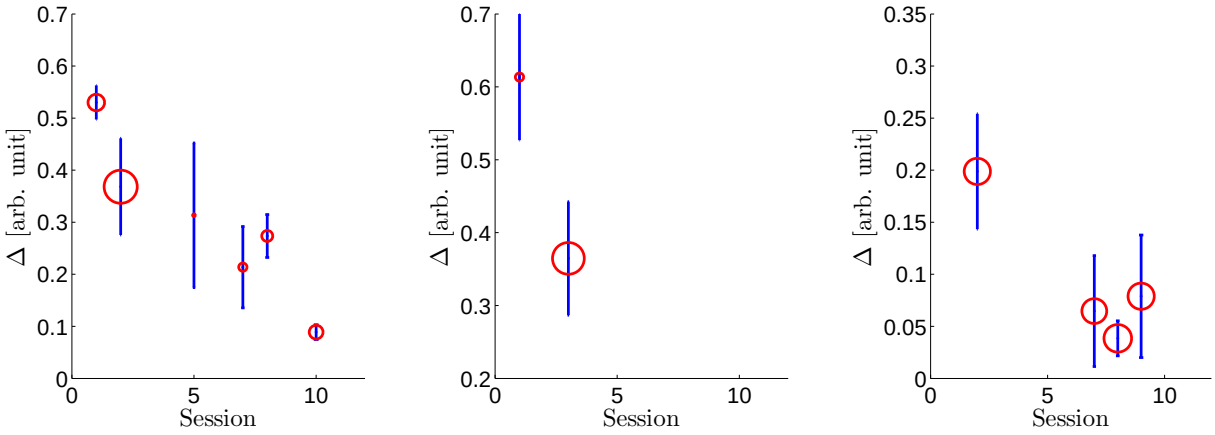


Figure 6.5: Measure of performance for each *repetition timeseries* δ is illustrated for session 2 of patient 2 performing the knee extension/flexion exercise.

and 6.4 show the distribution of the *repetition timeseries* of the healthy population and the training subset of the patient data over the two features selected by LASSO that have the largest variance in the healthy population. The clusters of the healthy population data and the patient data are separable. Furthermore, Figs. 6.2, 6.3, and 6.4 show that a patient's progress is in the direction of the variance of the healthy population data and moves towards the mean of the healthy population as the patients improve their performance during rehabilitation.

6.2.2 Measure of Performance for Repetition Set

Fig. 6.5 illustrates the values of δ for the second session of patient 2. Fig. 6.6 shows the calculated distance measure Δ and the distribution of δ for the 3 example patients. The exercise regimen is specific to each patient. The exercises are performed in a subset of the sessions, e.g., patient 2 performs knee extensions in session 1, 2, 7, 8 and 10. Furthermore factors such as pain, fatigue, psychological status, and environmental conditions contribute to patients' performance and it can not be expected that the patient progress increases monotonically. For all three patients an overall improvement over the course of the physiotherapy treatment can be observed. Some patients show rapid progress and are discharged early, e.g., patient 18 (in Fig. 6.6b). The distance measure for a *repetition set* is more reliable when the number of repetitions available for that exercise is larger. The feature-based approach generalizes to unseen patient data, e.g., the data of patients 8 and 18 was not used for the feature selection and the distance measure Δ shows the patient's progress throughout the rehabilitation as demonstrated in Fig. 6.6c.



(a) P2 knee extension (b) P18 knee extension (c) P8 knee extension

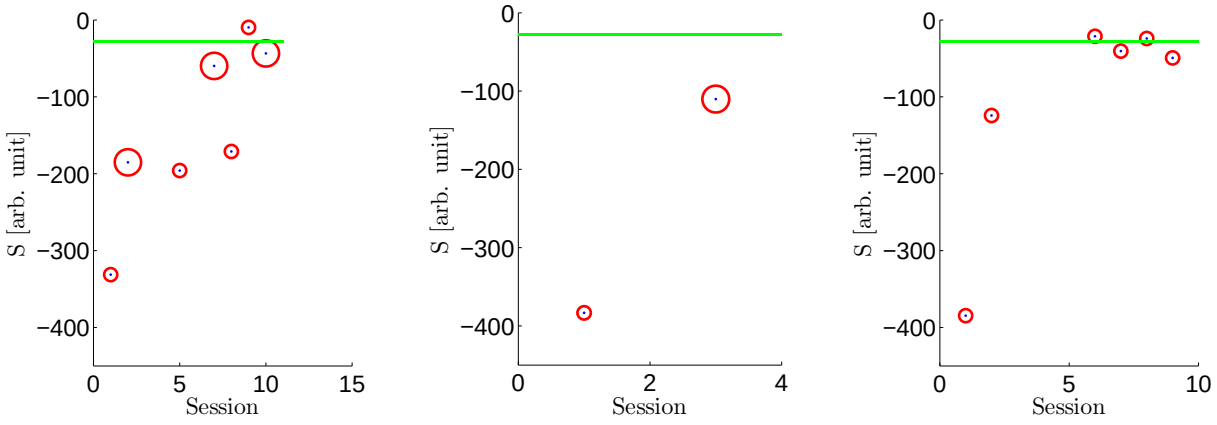
Figure 6.6: Results for the distance measure Δ calculated using the feature-based approach are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*. For knee extension, the top features are $min_{q_4}, mean_{q_1}, mean_{\ddot{q}_4}, rom_{q_4}, time$

6.2.3 Measure of Performance for Exercise Set

The quality measure Q and the overall score S for each session are obtained according to (4.25) and (4.26) using the overall distance measure Δ calculated for every *repetition set* of each session. We assume that as patients improve, the overall score increases from negative values towards zero.

Fig. 6.7 shows the score measures for each patient. It can be seen from the figures that the trend of the score shows progress but there are some inconsistencies in patient 2 session 8 (in Fig. 6.7a). These inconsistencies could be partially due to the small number of performed exercise repetitions. The other reason for these inconsistencies could be that the patient progress is not monotonic, and a decrease in the overall score could mean a decrease in the patient's performance.

Due to differences in health status, the exercise regimen of each session is different from one patient to the other. Among the three exercises chosen for analysis in this study, there are sessions where only one of these exercises is performed and therefore the score is based solely on the performance quality of that single exercise. This results in inconsistencies in the improvement trend of the score measure since a poor performance in one exercise



(a) Patient 2 (b) Patient 18 (c) Patient 8

Figure 6.7: An overall score S (○) is calculated for a *exercise set*, combining individual distance measures Δ of knee extension, knee-hip extension, and squat. The size of the marker indicates the number of exercises available in each session. The green line — shows the best score of the patients in their last physiotherapy session.

is not an accurate measure of the patient's overall status. The score measure estimates the patient's overall status more accurately when more exercise data from each session is available.

6.2.4 Measures of Performance with Feature Selection using Only Healthy Population Data

Visual analysis of the distribution of the most relevant features (see Figs. 6.2, 6.3, and 6.4) motivates to investigate whether only healthy population data is sufficient to select the most relevant features. Such an approach is beneficial when a physiotherapist may include a new exercise into the exercise regimen and patient data is not yet available for this exercise. Healthy population data can be easily collected by the physiotherapist him/herself performing the new exercise. We investigate this extension by using only healthy population data for feature selection. As can be seen in Figs. 6.2, 6.3, and 6.4, the top features capturing patient progress are highly variant in the healthy population. This motivates us to perform the feature-based approach using the most variant features in the healthy population. Variabilities caused by initial posture and sensor positioning are highly variant in the healthy population. Because we are only considering the variation in the healthy population, such features could get selected using our current approach.

To avoid this, features obtained from joint angle positions were removed from the feature vector. The top features are chosen from the fifteen most variant features that correlate less than .5 with each other. Equation (4.8) is used to calculate the distance measure for each *repetition timeseries*, (4.9) is used to calculate the distance measure for the *repetition set*, and (4.25) and (4.26) are used to calculate the overall score for each session.

Fig. 6.10a shows the correlation between the overall score from the feature-based approach when using healthy and patient population data for feature selection and when using only healthy population data for feature selection. As can be seen from the figure, the results correlate highly (over .65) for most patients. Even though feature selection based only on healthy population data does not take compensation strategies specific to the patient population into account, the extension using only healthy population data for feature selection shows promising results in detecting patient progress. Some inconsistencies between performing the feature selection only on the healthy population and performing the feature selection on the healthy and patient populations occur when the overall performance of a patient is constantly high (patient 7 and 15) or low (patient 5) over the course of the rehabilitation, and changes in the scores are small. In these cases, the correlation index can be low, because the two techniques differ when assessing small changes in performance.

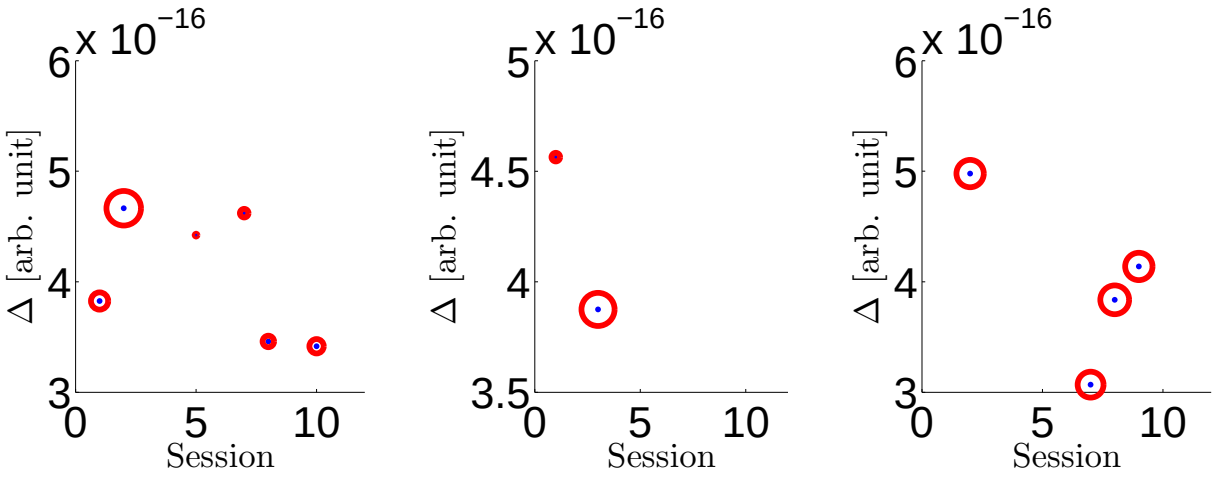
6.3 HMM-Based Approach

6.3.1 Measure of Performance for Repetition Set

The number of *repetition sets* are less than the number of features, therefore KW is chosen for feature selection. Fig. 6.8 shows the overall distance measure Δ (see Sec. 4.1.2) calculated for the sessions when patients 2, 8, and 18 performed the knee extension exercise. The features chosen by KW indicate the progress in unseen data by decreasing δ values over the course of the rehabilitation sessions, e.g., shown for patient 8 in Fig. 6.8c.

6.3.2 Measure of Performance for Exercise Set

Based on the quality measure Q (see (4.25)) of each exercise in one session the overall score S for an entire *exercise set* is calculated using (4.26). Fig. 6.9 shows the scores S_{HMM} for each patient. The scores show an overall trend of improvement for most patients. As before, the reliability of the score measure depends on the number of available exercises, i.e., outliers are usually observed when only one exercise is available to calculate the score.



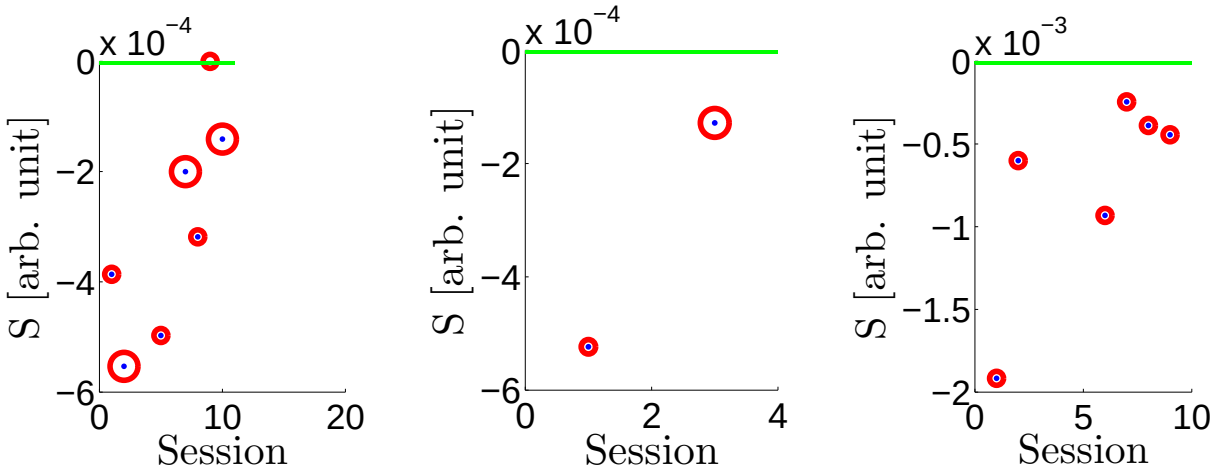
(a) Patient 2 knee extension (b) Patient 18 knee extension (c) Patient 8 knee extension

Figure 6.8: The HMM-based distance measure Δ_{HMM} (○) shows the trend in progress over the sessions, here illustrated for patients 2, 8, and 18. The marker size indicates the number of repetitions available in the *repetition set*.

The features chosen by KW generalize well to the unseen data e.g., patient 8 whose trend of improvement is captured by the approach (in Fig. 6.9c). Furthermore, the method captures the rapid improvement of patient 18 (in Fig. 6.9b), who is discharged earlier than other patients.

6.3.3 Measure of Performance with Feature Selection using Only Healthy Population

Moreover, we investigate whether healthy population data is sufficient for feature selection for the HMM-based approach. Feature extraction for the healthy population is described in section 4.1.2. Among the first fifteen most variant features in the healthy population, those that correlate least with each other (less than .5) are chosen as the top features. The distance measure Δ_{HMM} and the overall score S_{HMM} for each patient and each session are calculated following Sec. 4.1.2. The correlation between the overall score using healthy and patient population data for feature selection and using only healthy population data for feature selection is above .65 for most patients (see Fig. 6.10b). Negative correlation indices are observed when the changes in a patient's overall score are small.



(a) Patient 2 (b) Patient 18 (c) Patient 8

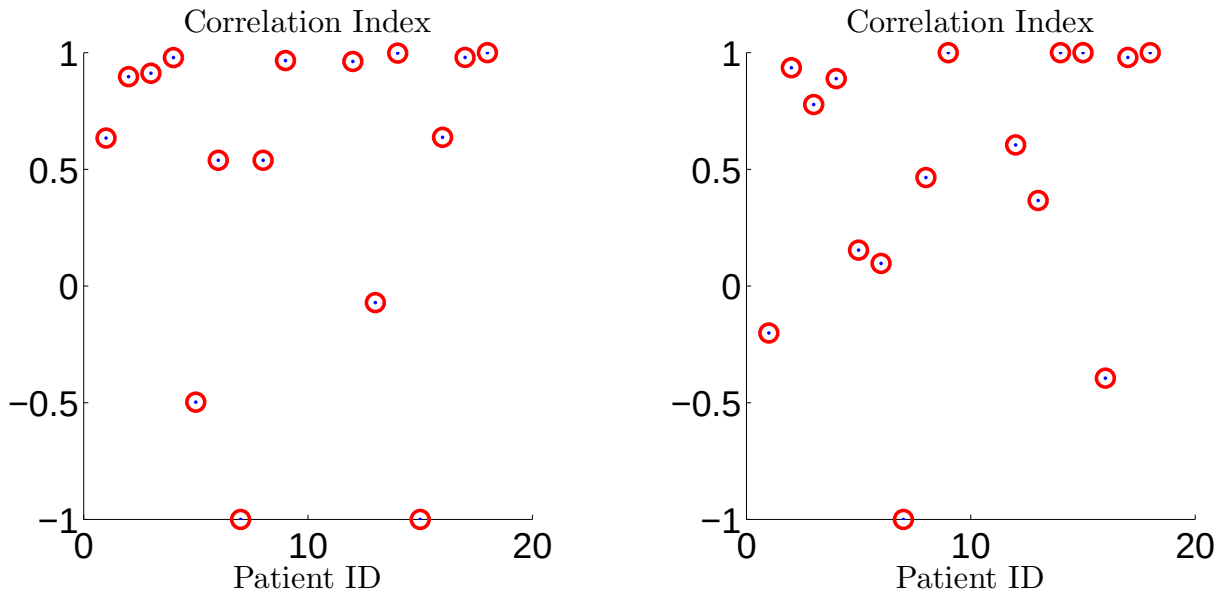
Figure 6.9: The overall score S_{HMM} (○) shows the trend of progress during rehabilitation. The marker size indicates the number of exercises available for each session. — shows the best score of the patients in their last session of performing the three exercises.

6.3.4 Effect of Number of States in the HMM

We also tried using an HMM with 5 states in our HMM based Approach. We assumed that the 5 states would correspond to: beginning of the movement, acceleration, pause, deacceleration, and ending of the movement. The models were trained on healthy and patient population for every *repetition set* and the distance measures for each exercise were calculated. The results were unable to capture the progress. Fig. 6.11 shows the the time series of three *repetition set* for three patients and the mean of states obtained from the 3-state and 5-state HMM models trained on these *repetition sets*. As can be seen from Fig.6.11, due to patients individual characteristics the same states in different patients correspond to different stages. Therefore the feature selection method cannot detect the features that correspond to variability caused by physical improvement in different sessions. On the other hand, the states of the 3 state HMM models always correspond to the same stages and are more appropriate for our purpose.

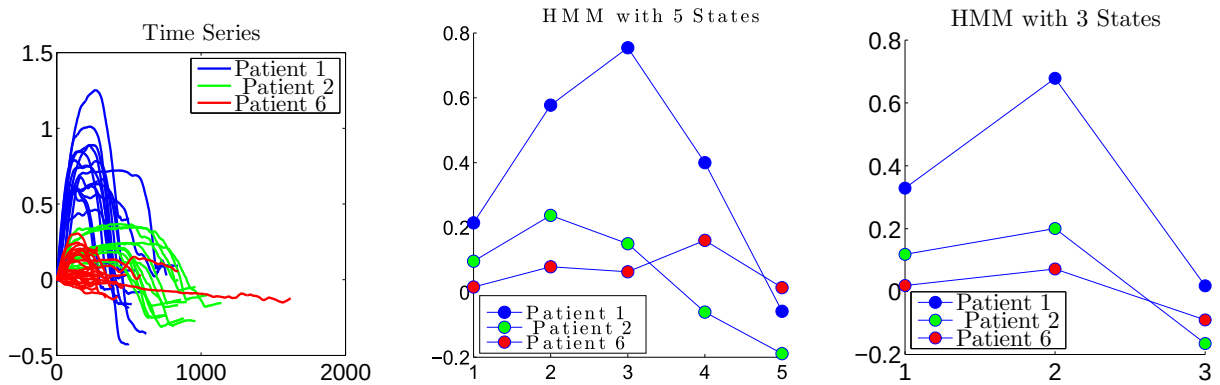
6.3.5 Comparison to Existing HMM-based Approaches

If only healthy population data is available, an intuitive distance measure δ for the HMM-based approach is the loglikelihood. Utilizing loglikelihood as a comparison tool between



(a) Feature-based approach (b) HMM-based approach

Figure 6.10: The correlation index between approaches with feature selection on healthy and unhealthy data and approaches with feature selection only on healthy data are for most patients over .65.



(a) Original Time Series (b) HMM with 5 States (c) HMM with 3 States

Figure 6.11: Effects of adding more states to the HMM

different HMMs is a popular approach in gesture and motion recognition literature [3] [12] [71]. The method is formulated such that a *repetition timeseries* of a patient is generated

by an HMM trained only on healthy population data. The median of the log likelihoods is considered as the distance measure Δ for a *repetition set*. We assume that as patients improve, their performance becomes more similar to the healthy population and the log likelihood increases towards zero and positive values through rehabilitation. This method does not capture the trend of progress for 80% of patients since distance measures Δ 2-3 times larger than their average range are observed for many repetition sets.

6.4 Comparison to Existing Approaches

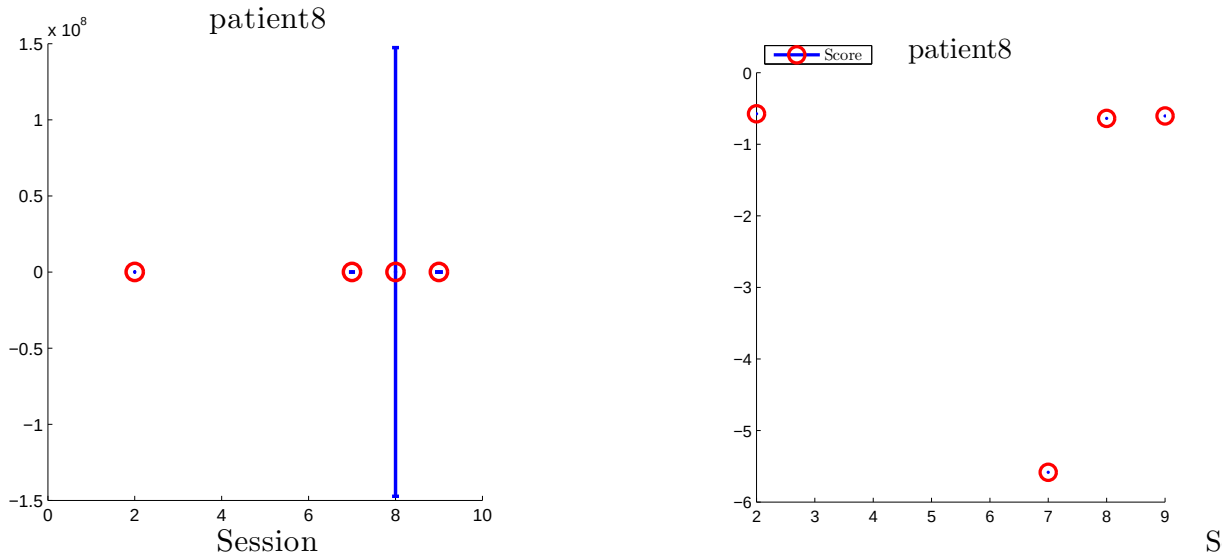
6.4.1 Comparison with Classifier-Based Approaches Based on Both Healthy and Patient Data

We provide a comparison of the proposed measures predicting patient progress to classifier-based approaches using both healthy and patient data. Popular approaches in the human motion literature are Naive Bayes (NB) [60], Kullback-Leibler (KL) divergence [66], and Support Vector Machines [105], which require training on both the patient data and healthy population data.

The measure of performance for SVM, NB, and KL divergence are computed using the procedure described in Sec. 4.2.1. The Δ values are calculated for the three approaches, i.e., Naive Bayes, KL divergence, and SVM. However, among the three methods only SVM was able to capture patient progress and therefore is selected as a comparison basis.

Fig. 6.12 depicts the of measure of performance, Δ, for patient 8 using the NB-based approach and the KL-based approach. With the NB-based approach, the measures of performance have large variances in many sessions for most patients, and are not reliable for detecting the trend of progress. With the KL-based approach, jumps 5-6 times greater than the average of the rest of the sessions can be observed in the measures of progress for most patients, which makes the result of the algorithm unreliable for detecting progress. The reason for this observation for the NB could be due to its input independence assumption which is violated by the features in this study. The reason for the jumps in the KL-based approach can be due to the attributes of KL distance, which results in small values when the models are similar and much larger values when the models are not similar.

The SVM-based approach is very dependant on the decision boundary. If the progress in a specific direction of the feature vector is not included in the training set for the SVM, the distance to the decision boundary is no longer a reliable measure of performance (See Fig. 6.13). Therefore, the SVM-based approach does not capture the trend of progress for some exercises of some patients in the dataset.



(a) NB-Based Approach

(b) KL-Based Approach

Figure 6.12: The measure of progress Δ of knee extension/flexion exercise calculated for patient 8. As can be seen from the figure, no visible trend of progressed is observed for the two methods.

6.4.2 DMP-Based Approach

In this section, the results of applying the DMP-based approach to the clinical data are discussed. A DMP model with ten basis functions is trained for each *repetition set* according to the procedure described in Sec. 4.2.2. We have also tried the approach with the number of basis functions ranging between 5-25. Ten basis functions is the smallest number of weights which results in higher correlation with the exercise regimen difficulty measure described in Sec. 6.5.1, and therefore is chosen for our analysis. The feature vectors are calculated by concatenating the weights of the basis functions computed for each *repetition set* as described in Sec. 4.2.2. The measure of performance, Δ, for each *repetition set* of each session for each patient is obtained by calculating the the euclidean distance between the corresponding feature vector and the feature vector of the healthy population according to Eq. 4.17. The measure of performance for the overall score for the *exercise set*, S, are calculated using Eq. 4.26.

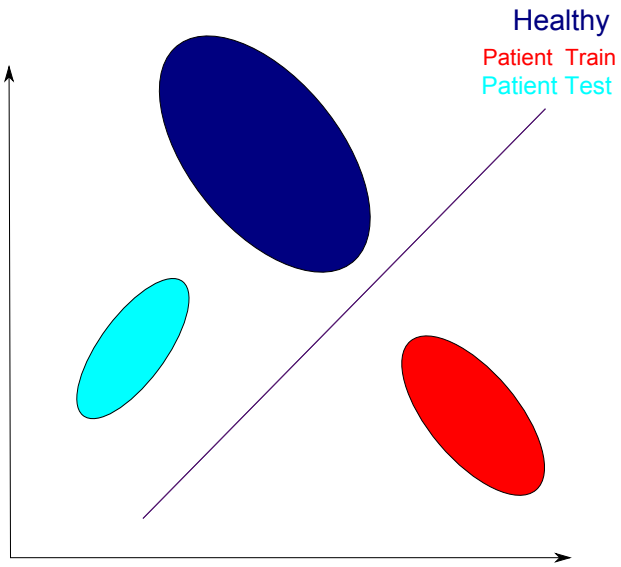


Figure 6.13: The SVM-based approach is highly affected by the training dataset. For example, if the red patient's data is the only data available during the training of SVM, the resulting SVM cannot capture the trend of progress for the cyan patient.

Measure of Performance for Repetition Set

Fig. 6.16 illustrates the results for the measures of performance, Δ, for the knee extension/flexion *repetition set*, for the exemplar patients. The approach is capable of capturing the trend of progress for patient 18 (Fig. 6.16c), and patient 8 (Fig. 6.16b); However it does not capture the trend of progress for patient 2 (Fig. 6.16a). The reason for this observation is that, as seen in Sec. 5, the DMP based approach is highly sensitive to inaccurate segmentation (See Fig. 6.14). Our clinical data is manually segmented and has small delays in the beginning and the ending of each repetition. These small delays affect the performance of the DMP-based approach. The other reason for this observation is that individuals have different temporal variability when performing the same exercise, e.g., fast or slow reaching the hold position, short or long duration of the hold position, and fast or slow reaching the rest position. Therefore the basis functions do not always correspond to the same stage of the motion across individuals, and the weights of the basis functions do not always correspond to the same stage of the motion. When the stages are misaligned, comparison of their weights is error-prone (See Fig. 6.15). As can be seen from Fig. 6.15, segments 1 and 2 correspond to reaching the desired posture in (a), whereas in (b), these segments correspond to reaching the desired posture and holding this posture. This reduces the performance of the DMP-based approach.

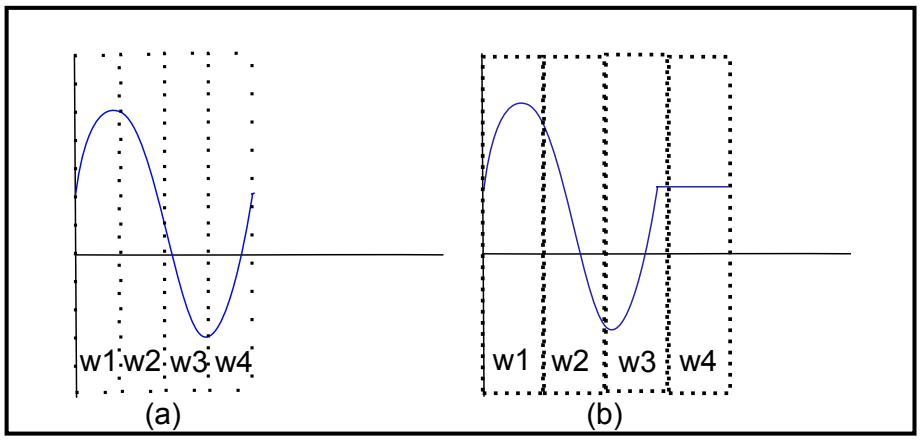


Figure 6.14: The inaccuracy in segmentation affects the DMP-based approach. Segmentation is delayed in (b) and segment w4 models the rest position whereas segment w4 models reaching the rest position in (a).

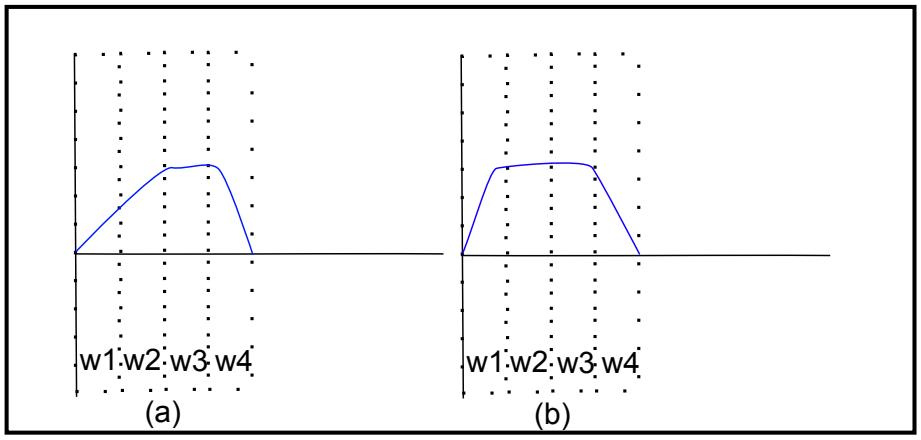
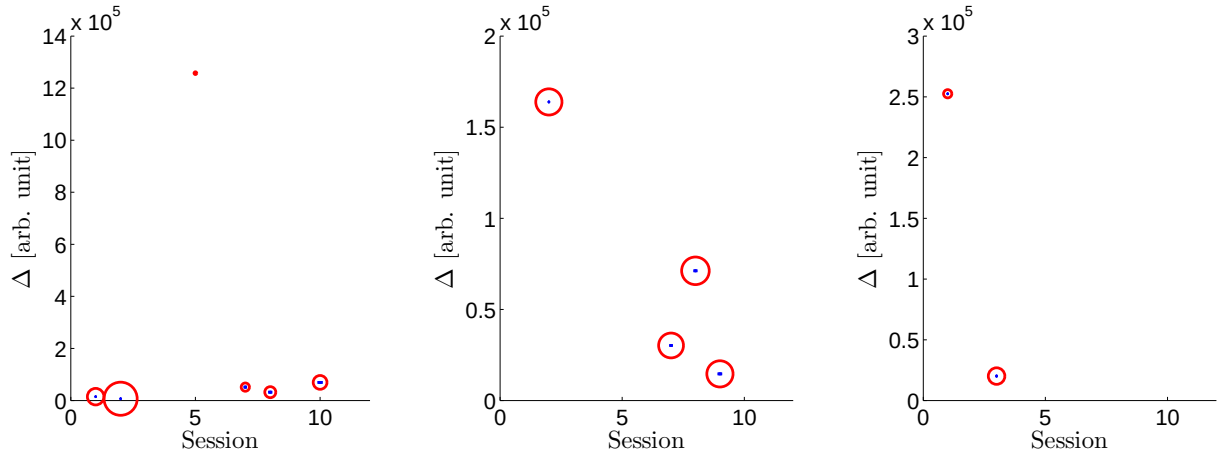


Figure 6.15: Different temporal variabilities affect the DMP-based approach since each weight corresponds to a different stage of the motion.

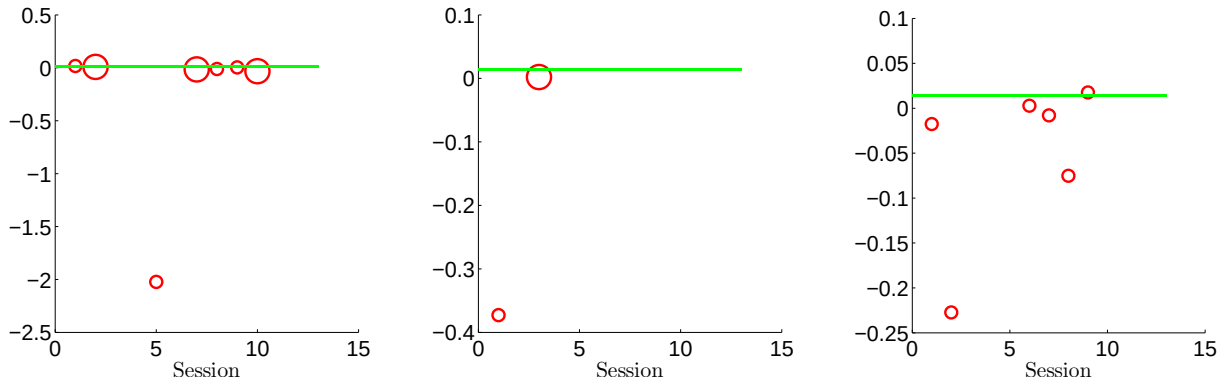
Measure of Progress for Exercise Set

Fig. 6.17 shows the results of calculating the score for the exemplar patients using the DMP-based approach. The trend of progress is captured for patient 18 as can be seen in Fig. 6.17b, however, for patients 2 and 8 the trend of progress is not visibly captured. This is partially due to the limitations of the DMP-based approach. It is also partially due to the sparsity of the *exercise set* as discussed in Sec. 5. Furthermore, the patients



(a) Patient 2 knee extension (b) Patient 8 knee extension (c) Patient 18 knee extension

Figure 6.16: The measure of performance Δ for the *repetition set* of the same exercise over multiple sessions utilizing the DMP-based approach.



(a) Patient 2 (b) Patient 18 (c) Patient 8

Figure 6.17: An overall score S (○) calculated for an *exercise set*, combining individual distance measures Δ of knee extension, knee-hip extension, and squat using the DMP-based approach. The size of the marker indicates the number of exercises available in each session. The green line — shows the best score of the patients in their last physiotherapy session.

performance improvement is not always monotonic.

Phase Adjustment in DMPs

To remove the effect of the temporal variability on the DMP-based approach, we also tried two phase adjusted DMP-based approaches, where the phase of each *repetition timeseries* in the patient population is adjusted based on the timeseries generated from the DMP model of the healthy population. In the phase adjusted DMP-based approach, the adjusted phase vectors are used to train a DMP model of each *repetition set* and the weights of the DMP model are considered as features (See Appendix A for details on the approach). The phase adjusted DMP-based approach is not capable of capturing the trend of progress in the patient population. When performing the phase adjustment, the temporal information contained in the patient motion is neglected. However, a part of patient progress is due to temporal improvement, and when this information is neglected the approach does not perform well.

6.4.3 Kernel-Based Approach

In the model-based approaches such as the HMM-based and the DMP-based approach, the information about a few distorted motions in a movement in one *repetition set* is often lost. The kernel-based method is capable of capturing the difference between the patient population and the healthy population when the patients' motions are shaky or consist of pauses (See Fig. 6.18). In this section, we discuss the results of applying the kernel-based method on the clinical data for estimating patient progress through physiotherapy rehabilitation. First, we use the entire healthy population's data to obtain the measure of progress for each *repetition timeseries* of the patients. The kernel-based approach is computationally expensive, and therefore it is preferred if the measures of performance could be obtained using the data of only one healthy member. In the following section, we discuss how the results will change if the healthy population consists of one member only compared to when all the healthy members are included, e.g., when the data is collected from the physiotherapist for a customized *exercise set*.

Measure of Performance for Repetition Set

The distance measure for a *repetition timeseries* is calculated using Eq. 4.22 following the procedure described in Sec. 4.2.3. The number of *subgroups* for the exercises included in this thesis is two. The first *monotonic subgroup* in all the exercises corresponds to reaching the desired posture, and the second *monotonic subgroup* corresponds to returning to the initial posture. All the members of the healthy population are considered for obtaining the average model of the *repetition timeseries* for the healthy population.

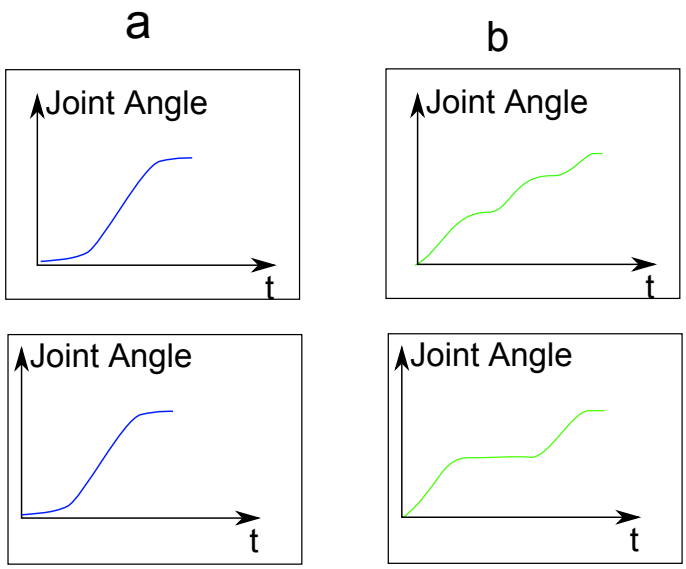


Figure 6.18: The figures in each row have different sample distribution. The first row illustrates the timeseries of a shaky motion, and the second row depicts the pauses that happen in patients' motion. (a) corresponds to the motion of the healthy population in the first *monotonic subgroup*. (b) are the examples of pauses and shakes in the motion of patients.

The performance measure for a *repetition set* in one session, Δ, is calculated using Eq. 4.9. We assume that as patients improve, their motion distribution becomes closer to the healthy population, resulting in smaller values for the performance measure. Fig. 6.20 shows the calculated distance measure, Δ, for the exemplar patients. Fig. 6.20a shows that the algorithm is capable of capturing the trend of progress for patient 2, and inconsistencies can be observed for patient 8 in Fig 6.20b. The inconsistencies are both due to the jumps caused by *exercise set* sparsity and the algorithm's limitations (See Sec. 5).

Furthermore, the kernel-based approach neglects the temporal evolution of the time-series and therefore there are different cases where the motion timeseries look very different but have a small difference based on the MMD value (See Fig. 6.19), which also contributes to the algorithm's performance. As can be seen in Fig. 6.19, using the MMD measure, motions with perturbations which are common during the first days of rehabilitation do not differ from motions with pauses in between which are more common during the last days of rehabilitation. Furthermore, scaling a motion through time does not change the MMD measure. Therefore, there are certain aspects of progress that MMD cannot capture.

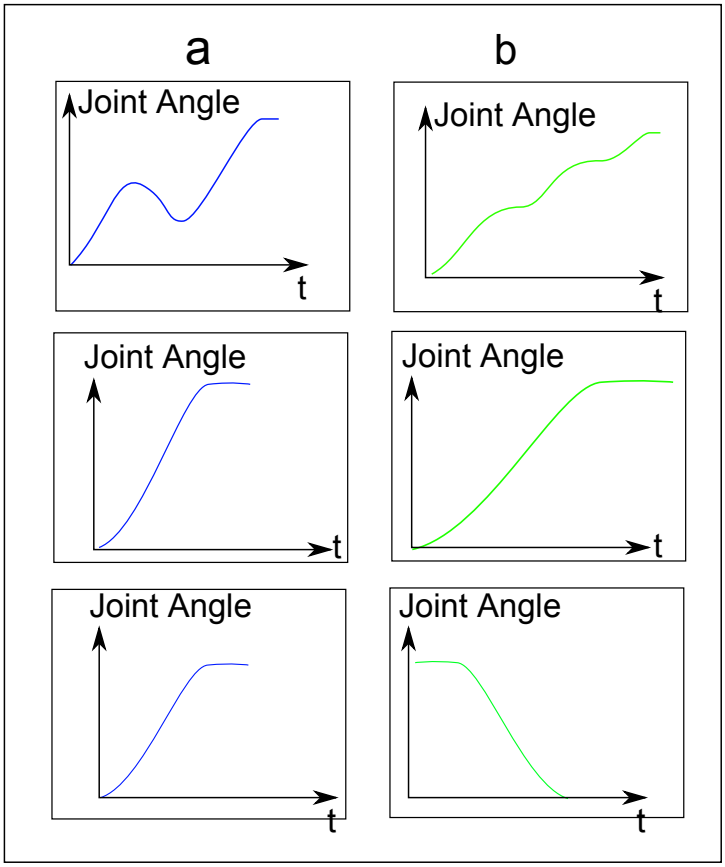
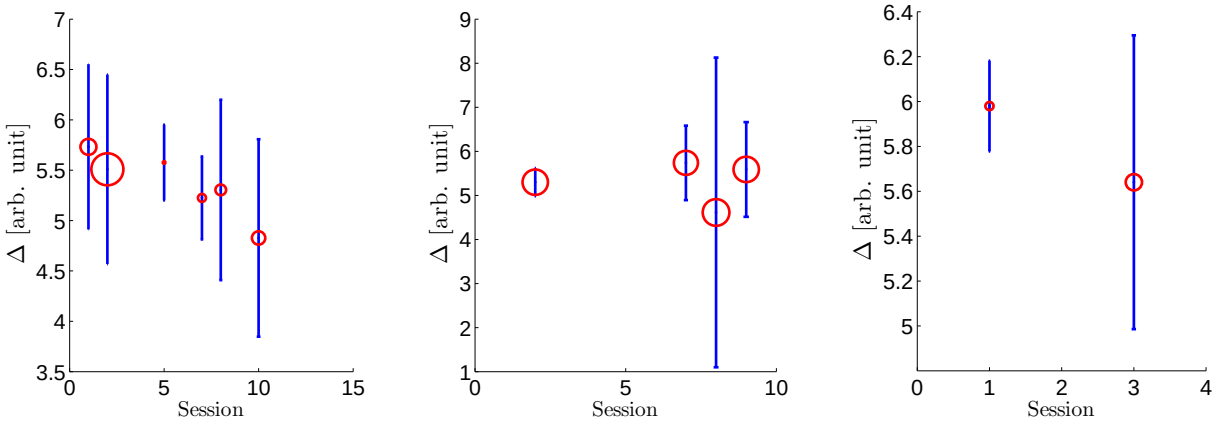


Figure 6.19: The figures in each row have the same sample distribution even though they look very different.

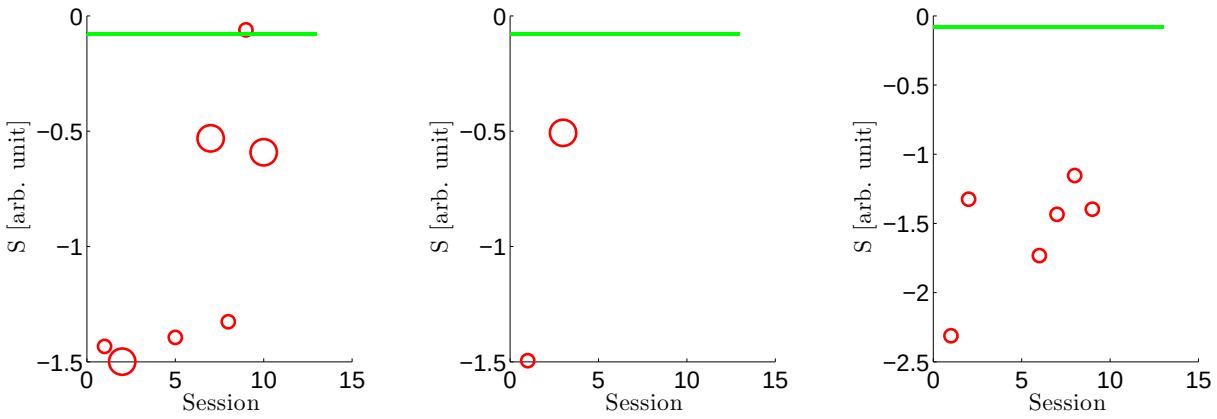
Measure of Performance for Exercise Set

The quality measure, Q, and the overall score, S, for each session are obtained using Eqs. 4.25 and 4.26. We assume that as patients improve the score value increases from negative values to zero. Fig. 6.21 illustrates the score measures for the exemplar patients. As can be seen from the figure, the trend of improvement is captured by the approach but there are some inconsistencies. The inconsistencies and jumps could be partially due to the *exercise set* sparsity and the small number of performed exercises. The patient's progress is not monotonic. As will be discussed in Sec. 6.5.2, the inconsistencies could be due to the patient's decrease of performance in some sessions. Therefore, the reliability of the score depends on the number of available exercises in the *exercise set*.



(a) Patient 2 knee extension (b) Patient 8 knee extension (c) Patient 18 knee extension

Figure 6.20: The measure of performance Δ for the *repetition set* of the same exercise over multiple sessions utilizing the kernel-based approach.



(a) Patient 2 (b) Patient 18 (c) Patient 8

Figure 6.21: An overall score S (○) is calculated for an *exercise set*, combining individual distance measures Δ of knee extension, knee-hip extension, and squat. The size of the marker indicates the number of exercises available in each session. The green line — shows the best score of the patients in their last physiotherapy session.

The kernel-based approach is based on distinguishing the distribution of the two time-series. Due to the formulation of the method, it does not require training on the patient population data and therefore is ideal for cases where the patient data is not available in advance, e.g., customized exercises. The kernel-based approach is capable of capturing the

trends of progress that are due to increase in range of motion. It can also capture the trend of progress due to the improvement of motion smoothness (See Sec. 5). This approach is not capable of capturing the trend of progress which is due to temporal improvements only.

Effect of Having one Healthy Member

As mentioned earlier, the kernel-based approach is computationally expensive. Therefore, it is beneficial to investigate whether one healthy member in the healthy population is sufficient for the analysis. The analysis is redone 20 times, each time including only one member in the healthy dataset. The average correlation index of the distance measures for the *subgroup*s, δ_{k_ω}, is calculated for the case where all the healthy members are considered in the healthy dataset vs. the case where only one healthy member is considered in the healthy population. The results indicate that having only one member in the healthy population correlates highly with the case were the all the healthy members are considered in the healthy population (over .87).

6.5 Validation

The patient's physical status is visually assessed by the physiotherapist in each rehabilitation session. The physiotherapist uses this assessment to formulate the patient's regimen and decide his or her treatment duration. This evaluation is subjective and does not have a quantified interpretation. Moreover, the evaluations of each session have not been documented regularly in the dataset used for this study.

While direct quantified expert evaluation is not available for comparison, exercise difficulty and duration can be used as an indirect measure of PT assessment. In the first sessions of rehabilitation, exercises recommended by the physiotherapist are mostly composed of supine and sitting exercises with very few repetitions. As patients improve, the recommended exercise regimen becomes harder, i.e., includes standing and gait exercises. Furthermore, the number of repetitions of the exercises increase. Therefore, the exercise regimen can be utilized to obtain an estimate of the clinical assessment of the patient's overall performance, i.e., overall score. Therefore, to validate our approaches, we perform the following experiments:

- Comparison of the score measures with an estimate of patient progress calculated based on their exercise regimen

- Qualitative comparison of the score measures with the physiotherapist notes in the patient health charts

6.5.1 Comparison of the Score Measures Based one Estimation of Patients Progress Calculated Based on Their Exercise Regimen

To estimate the true measure of patient progress from the exercise regimen, we use the complete information of all *exercise set*s available from all patients. We consider the exercises performed in the last session for patients with fewer than 4 sessions, and the exercises performed in the last two sessions for patients with more than 4 sessions as the hard exercise set. We consider the last two sessions because for most patients the last session is a last check up and contains very few exercises. The first session exercises are considered as the easy exercise set. For each exercise, the number of patients who have performed this exercise in their last day are counted and this number is divided by the total number of patients to obtain the probability that the exercise belongs to the hard exercise set, $p_H(\Omega)$. We eliminate the exercises performed by fewer than three patients, i.e., exercises with probability less than .15, from the hard exercise set. The same procedure is performed to determine the probability of belonging to the easy exercise set, $p_E(\Omega)$. If an exercise is not in the hard exercise set the probability that this exercise belongs to the hard exercise set, $p(H|\Omega)$, is assigned a value of .01:

$$p(H|\Omega) = \begin{cases} p_H(\Omega), & \text{if } \Omega \in \Gamma_H \\ .01, & \text{otherwise} \end{cases} \tag{6.1}$$

$$p(E|\Omega) = \begin{cases} p_E(\Omega), & \text{if } \Omega \in \Gamma_E \\ .01, & \text{otherwise} \end{cases}, \tag{6.2}$$

where Ω is the exercise from the *exercise set*, Γ_H is the hard *exercise set*, and Γ_E is the easy *exercise set*. The overall measure of progress for each session and each patient is calculated as

$$\begin{aligned} S_{i_{GT}} &= \frac{\sum_{\Omega \in \Gamma} log(p(H|\Omega))}{\sum_{\Omega \in \Gamma} log(p(E|\Omega))} \\ S_{GT} &= [S_{1_{GT}}, S_{2_{GT}}, ...], \end{aligned} \tag{6.3}$$

where i is the number of sessions, $S_{i_{GT}}$ is the ground truth overall score for each session, and S_{GT} is the overall score for all the sessions.

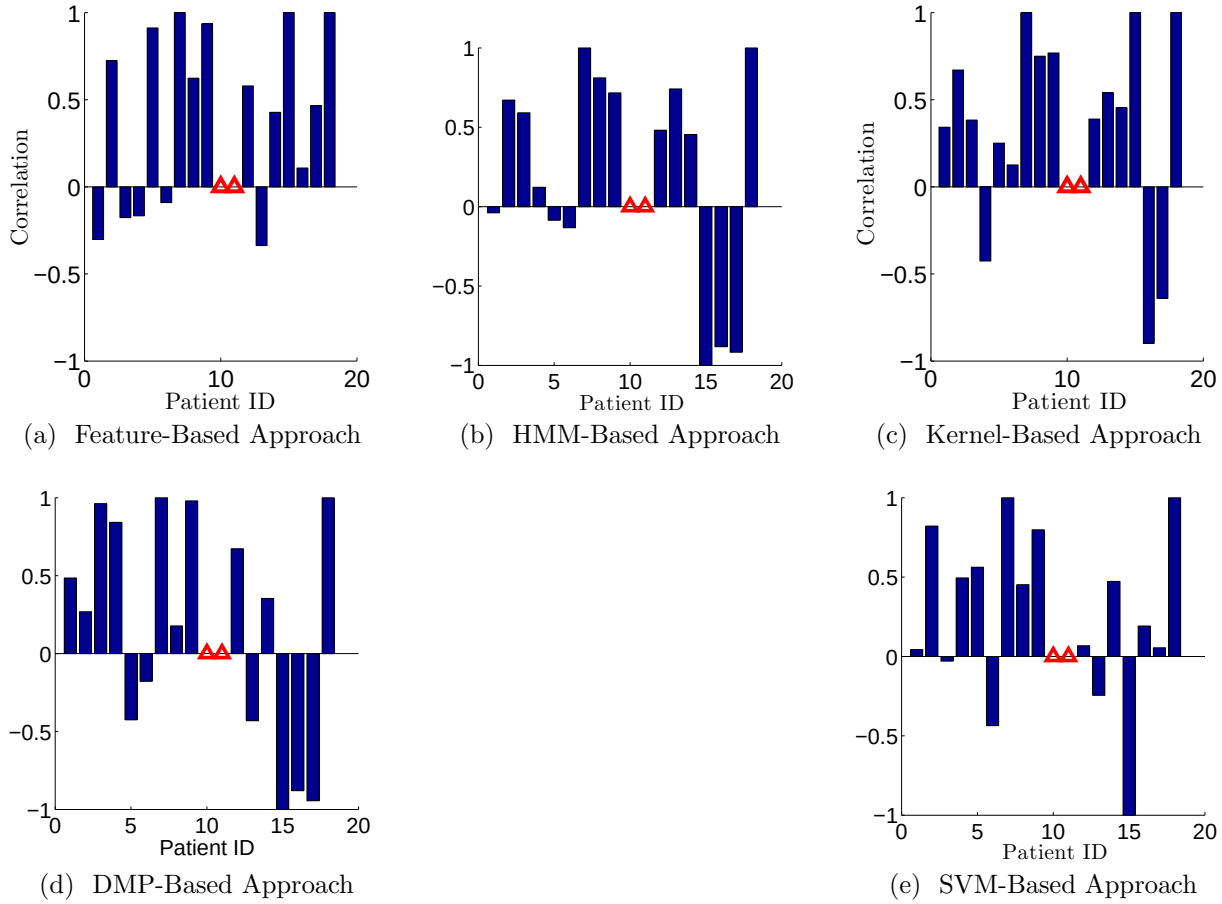
Fig. 6.22 shows the correlation index comparing each method's overall score, S, for each patient with the overall score obtained from Eq. 6.3. The feature-based and the HMM-based approaches correlate moderately (correlation over 0.4 value) in most cases (over 55% of patients for the feature-based approach and over 56% of patients for the HMM-based approach). Low correlations occur mostly in cases where for many of the sessions few exercises are available for evaluation, e.g., patient 4 and 3. As mentioned in Sec. 5 the inconsistency in the number of available exercises between different sessions can cause jumps in the overall score, which in turn results in a poor correlation with the ground truth. The cases where the clinical assessment does not correlate well with our proposed approaches are either caused by patients who do not show a visible change in their overall score or are caused by patients who have inconsistency in the number of available exercises in more than half their sessions. When these patients are excluded (9 patients out of 16 remain) the mean of the correlation becomes 67% for the feature-based approach and 72% for the HMM-based approach.

The kernel-based approach correlates moderately (correlation over 0.4 value) with the overall score from Eq. 6.3 for 43% of the cases, and the DMP-based approach correlates moderately with the overall score for 43% of the cases. The poor performance of the DMP-approach is partially due to the fact that it is sensitive to segmentation, and that the basis functions are highly affected by the motion temporal variability. The kernel-based approach has a poorer performance compared to the feature-based approach and the HMM-based approach. This is a result of the algorithm's inability to capture the temporal information in the human motion.

The SVM-based approach correlates moderately with the overall score calculated utilizing Eq. 6.3 for 43% of cases. The SVM-based approach is highly affected by the decision boundary and performs very poorly on most of the on unseen data.

6.5.2 Qualitative Comparison of the Score Measures with the Patient Health Charts

Physiotherapists assess and record patient performance and condition on admission. Even though these assessment forms often contain unfilled sections and are mostly qualitative, they include information about the initial status of the patients. Furthermore, patient performance during rehabilitation is sometimes recorded by the physiotherapists in the daily charts. We have studied these forms and charts for each patient. In this section, we provide the physiotherapist assessments for the exemplar patients and compare these evaluations to the HMM-based and feature-based approaches' proposed overall score.



(a) Feature-Based Approach (b) HMM-Based Approach (c) Kernel-Based Approach

(d) DMP-Based Approach (e) SVM-Based Approach

Figure 6.22: Correlation between the overall score calculated for each method and the ground truth for each patient. The data of patients 10 and 11 has only 1 session available and therefore correlation cannot be calculated.

Patient 2 was admitted to the hospital after a hip joint replacement surgery. Based on the first day assessment, she was forbidden to perform hip abduction due to hip precautions. She was capable of bearing her weight on her feet, but needed assistance for rolling in bed and transferring from bed to wheelchair. She used a 2 wheel walker and was capable of walking for 2 meters only. The range of motion score in the recovering leg for different DoFs (e.g., hip abduction, and etc.) was $\frac{8}{18}$ and the patient had a high risk of fall according to her stability test results. On the night before session 5, the patient fell in the bathroom causing pain in her lower extremity joints and therefore affecting her performance on session 5. This information matches the overall scores calculated for both of the proposed approaches in

Figs. 7 and 9. The patient is sent back to the ER on session 8 due to complications unrelated to her surgery; The effects of this incident on her performance are captured by both of the proposed approaches in Figs. 6.7 and 6.9. The patient comes back a week later to continue her physical rehabilitation. On session 9 the patient is able to walk 70 meters with a 2 wheel walker with supervision and her range of motion score for the rehabilitated leg's DoF becomes $\frac{18}{18}$. This progress is captured in the 9th and 10th session by both the HMM and feature-based approaches in Figs. 6.7 and 6.9.

Patient 8 was admitted to the hospital after a hip joint replacement surgery. Based on her first day assessment, she was capable of bearing her own weight but was feeling severe pain. She had high bed mobility but needed assistance in transitioning from bed to wheelchair. She had a high fall risk according to her stability test and was capable of walking for 10 meters only. Her range of motion score was $\frac{12}{18}$ and she could perform the exercises with assistance. On her second session she was capable of performing all her transfers independently and was capable of walking 50 meters independently using a 2 wheel walker. the proposed overall score for both approaches capture the progress for this patient between sessions 1 and 3 in Figs. 6.7 and 6.9. In her 9th session she performed 20 repetitions of bilateral exercises which indicates improvement in her performance. The physiotherapy did not record the range of motion score at discharge.

Patient 18 was admitted to the hospital due to knee replacement surgery. He had a high risk of fall and had a somewhat normal bed mobility. He needed supervision for bed to wheelchair transfer. He could walk 30 meters with supervision. In session 3 he had two physiotherapy sessions where he walked 40 meters supervised using a 4 wheel walker in the morning and 70 meters in the afternoon. Our scores capture this rapid progress for this patient between the first and third sessions in Figs. 6.7 and 6.9. For Patient 18, the physiotherapist did not record the motion score, either at admittance or at discharge.

The proposed approaches qualitatively correlate with the medical charts of the patients. However, the notes in the medical chart often do not contain a quantified measure of the patient's performance, and are often subjective. For example, the patient's level of pain is dependent on the patient's perception of his/her pain, and it is not obvious from the notes how this pain may have affected the patient's performance. Furthermore, while the overall quantity or duration of the exercises the patient has performed may be recorded, e.g., 70 meters walking, detailed assessment of how well the task was performed is not recorded. Therefore, in the sessions that pain is reported for the patient, we cannot be sure if this pain has affected the patient's performance, or, when the quantity or duration of the exercise that the patient performs increases, we cannot be sure that the patient's performance has also improved.

Table 6.2: Experimental Evaluation Summary

Method	Observation	Discussion [3]
Feature-Based Approach	Captures trend of progress for exemplar patients, there are jumps and some inconsistencies in the measures of performance	The approach is highly affected by noise
HMM-Based Approach	Captures trend of progress for exemplar patients, there are jumps and some inconsistencies in the measures of performance	—
Kernel-Based Approach	Does not capture trend of progress for patient 8	Does not consider time evolution
DMP-Based Approach	Does not capture trend of progress for patient 2	Highly affected by temporal variability and segmentation inaccuracy
Classifier-Based Approach	Captures trend of progress for exemplar patients, there are jumps and some inconsistencies in the measures of performance	Dependence on training data and decision boundary

[1] Common possible causes for inconsistencies in the overall score: patient progress is not always monotonic, small number of data points.

6.6 Summary

Quantified and continuous [2] measure of performance can be beneficial for monitoring patient progress during the course of physiotherapy rehabilitation. This work introduces two approaches, feature-based and HMM-based, for capturing the continuous change in patient data. A distance measure is introduced as a measure of performance for a *repetition time-series* and *repetition set*. The overall score is then calculated for the *exercise set* in each session and captures the overall performance of the patient. The proposed approaches are evaluated on data of exercises commonly performed after hip or knee replacement surgery. The results show that the proposed approach is promising for tracking patient progress over the course of treatment. The proposed approaches have a higher correlation with the score obtained from the exercise regimen compared with the existing approaches. Furthermore, the proposed approaches correspond qualitatively with the patients' medical charts. Table 6.2 summarizes the results for the experimental evaluation.

[2] The measure is provided after each repetition of an exercise.

The proposed approaches show promising results for capturing patient progress. However, the proposed approaches are not capable of detecting progress in the current data set when the changes in the patient's performance are small. Furthermore, it is difficult to estimate the patient's performance when only a few samples of the patient's motion are available. Moreover, in this study we only consider a subset of the exercises and generalization to additional exercises remains to be investigated. When considering an *exercise set* , we have used exercise difficulty to ensure that the exercises with different levels of difficulty are comparable. This consideration however is imperfect and requires further investigation and improvement.

Chapter 7

Conclusions and Future Work

Monitoring exercise performance during physiotherapy can provide an objective measure of patient progress. Movement performance shows temporal and spatial variability caused by multiple sources including the stochastic nature of muscle recruitment, as well as individual differences in height, age, pain, fatigue, and progress. The objective of the proposed distance measures and the overall score are to capture the variability caused by improvement and progress over the course of the physiotherapy treatment.

7.1 Conclusions

In this thesis, we estimate a continuous[1] measure of patient performance to capture their progress through rehabilitation, whereas most existing works [117] [146] can only separate healthy from patient data using classification. We formulate a measure of performance for an *exercise set*, whereas most current works [53] [146] consider only a single exercise. Moreover, we evaluate our approach both on synthetic and patient data, whereas many of the current works focus on synthetic analysis and simulated data only. Furthermore, the HMM-based approach and the feature-based approach can be applied when patient data for a motion is not available whereas the classification techniques require both healthy and patient data for training. Both proposed approaches achieve generalization to new patients by including healthy population data as reference. Furthermore the score measure formulations can be applied to any set of exercises as long as the corresponding healthy population data is available. This flexibility enables the physiotherapist to include patient

[1] The measure of performance is provided after each repetition of the exercise.

specific or novel exercises requiring only a healthy reference set. The score measure is formulated in a way to handle individual exercise regimens and a variable number and type of exercises.

To enable feature selection when little or no patient data is available using the feature-based approach and the HMM-based approach, we assumed that the healthy population exhibits the same compensation strategies as the patients to a smaller degree. This hypothesis is formulated on the basis that difficult motions result in compensatory strategies in human motion. This assumption is supported for the three exercises discussed in this thesis as shown in Figs. 6.2, 6.3, and 6.4. In the absence of patient data, we considered the most variant features in the healthy population as the top features. However, feature selection using both healthy and patient population data is more accurate because it allows the method to detect the compensatory strategies which are specific to the patient population.

For both the feature-based and the HMM-based approach, the distance measure Δ for a *repetition set* and the overall score S for a *exercise set* assess patient progress. The feature-based approach is faster to compute whereas the HMM-based approach provides details about each stage of the motion.

Of the comparison methods considered, the kernel-based approach and the DMP-based approach do not perform as well as the HMM-based approach and the feature-based approach. The DMP-based approach is highly affected by segmentation inaccuracy, and temporal variability of motion execution. The kernel-based approach does not include the temporal information of the motion, and therefore is not capable of capturing certain types of change in human motion.

We also compared the proposed approach to estimating patient progress based on the magnitude of the SVM-margin between the healthy and patient population data. Our proposed approach has a high degree of correlation with the SVM-based approach, while requiring less training data. SVM requires feature selection on top of our LASSO feature selection to identify the most variant features, and requires training data from both the healthy and patient population. In its current form, the SVM is not capable of capturing the progress based on different exercises. We combine the SVM approach for generating distances with our approach to generate the overall performance score for multiple exercises using the SVM. The results obtained with synthetic data illustrate that the proposed approaches are superior to this classification method in the presence of noise, inaccurate segmentation, and incomplete timeseries.

We also compared the HMM-based approach and the feature-based approach to physiotherapist evaluations by computing the correlation between progress estimate and advancement of the exercise regimen, and qualitatively. The results indicate that both proposed

approaches correlate moderately with the score obtained from the exercise regimen in 56% of the cases. Furthermore, the method qualitatively correlates with physiotherapist evaluations. The proposed approaches are superior to the existing approaches when comparing to the score obtained from the exercise regimen.

The proposed approaches are not capable of explaining the cause of the patient's progress or lack of progress. Furthermore, the proposed approaches do not include a generative model of progress incorporating physiological principles. The causes of change in the motion profile of a patient cannot be differentiated by the proposed approaches. Moreover, the validation of the proposed approaches is impeded due to lack of clinical ground truth.

7.2 Future Work

The proposed measures consider the improvement due to exercise performance; other factors such as pain and psychological status are not included in our analysis. Different pain treatments can affect motion performance, e.g., reducing pain killer medication may lead to a decrease in observed exercise performance even though the overall health status improves.

The order of exercises performed in obtaining the overall score is not considered in the proposed formulation and effects of fatigue on movement performance are not included. Exercises vary in their difficulty and the variance of the healthy population's performance is considered as an estimate of exercise difficulty. Considering patients, exercise difficulty may further depend on the type of surgery. Variance in the healthy population depends further on fitness level and familiarity with an exercise.

The proposed approaches can be used both to provide information about how well a patient performs a specific task and repetitions of that task, and also to identify what is different between the ideal motion and the patient's motion. However, since the proposed approaches calculate the performance measures based on a set of features, the information about the contribution of each feature and the reasons for the observed difference between the patient's performance and the healthy performance is not captured. To determine the cause of the difference in performance between the patient and the healthy population, either the features need to be further investigated or the hypothesized causes of the difference should be explicitly modelled, e.g. for fatigue [57].

In the current evaluation, the healthy population consists of individuals between 22-35, and the patient population consists of older population between 48-86 years old. There may be differences in performance between the two populations due to the age difference.

Furthermore, as observed from the data, the healthy population's motion is not always completely correct, and often some form of compensation is observed in their data. Moreover, in this thesis, we have only considered three exercises. Therefore, future work includes evaluating with an age matched healthy population or physiotherapist demonstrations, and evaluating on a larger set of exercises.

Patients perform different exercises during one physiotherapy session. The fatigue and/or pain after each exercise could affect the performance in the following exercises. Therefore, the proposed approach could also be enhanced by considering the order of the exercises in the formulation.

Each patient has a unique medical history, and a different physical fitness and status. Furthermore, patients could be recovering from different surgeries. Therefore a possible future direction for the measures of progress are subject independent measures of exercise difficulty, and fatigue and pain.

The classifiers considered for the existing classifier based approaches are common linear separators. More powerful classifiers may yield better performance.

Future directions may also investigate the smallest clinical important difference of the proposed score and whether physiotherapists using the new distance metrics gain additional clinically relevant information not available through visual observation alone.

APPENDICES

Appendix A

Phase Adjusted DMP

The DMP-based approach can be modified to account for the timing difference between the patients' motions and the healthy population's motion. In the basic DMP-based approach, the phase evolves linearly with time for each repetition timeseries, i.e., the phase is equal to zero at the beginning of each repetition timeseries and is equal to 2π at the end of each repetition timeseries, and has a linear relationship with time. This phase assignment causes the same value of phase to correspond to different stages of motion between different trials(See Fig. A.1). This limitation of the basic DMP-based approach results in a poor performance for some patients.

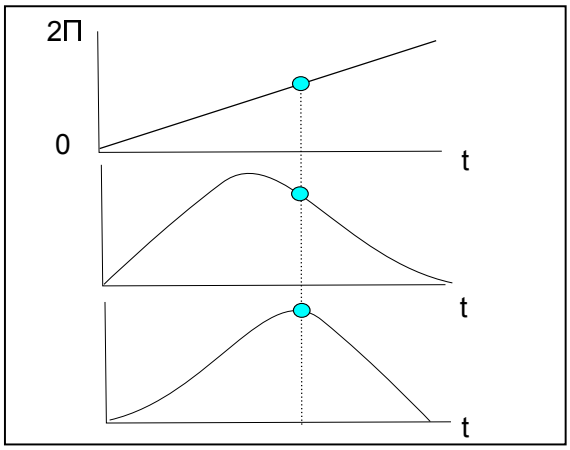


Figure A.1: The different stages of motion correspond to the same phase value.

To avoid this limitation, we use a DMP-based approach, which relies on phase adjustment between the patient population and the healthy population. In the phase-adjusted

DMP-based approach, the DMP model for the healthy population is trained on all the repetition timeseries in the healthy population. The phase of each repetition timeseries in the patient data is adjusted based on the generated timeseries from the healthy population's model. The DMP model for each *repetition set* is then trained using the adjusted phase vectors of the *repetition timeseries*. The measures of performance for each *repetition set* are the difference between the weights of the healthy population model, and the obtained weights for each *repetition set*.

In the first step of this approach a DMP model is trained for each DoF, i.e., joint angle positions, based on all repetition time series in the healthy population. For each of these repetition timeseries, the phase variable ϕ is considered to vary linearly with time with 0 at the beginning of each repetition timeseries and 2π at the end of each repetition timeseries (See Fig. A.2).

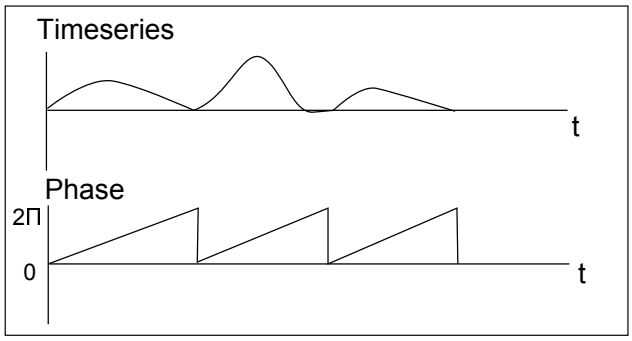


Figure A.2: The timeseries and the phase values used for training the DMP model for the healthy population.

To adjust the phase for each patient's repetition timeseries, we assume that the human motion is synchronous and we consider the most variable joint angle for phase adjustment. For the most variable joint of each repetition timeseries in the patient population, a healthy population timeseries is generated using the DMP model of the most variable joint in the healthy population. A constant value is added to the generated timeseries of the healthy population to make the initial value of the two timeseries the same. The generated timeseries of the healthy population is then scaled so that the maximum of the two timeseries become equal (See Fig. A.3).

At each time step in the generated timeseries of the healthy population, the value of the generated timeseries is compared to the values of the patient's repetition timeseries. The closest equal value in the patient's repetition timeseries is assigned the same phase as the generated timeseries at the given time step. If a value does not exist in the patient's

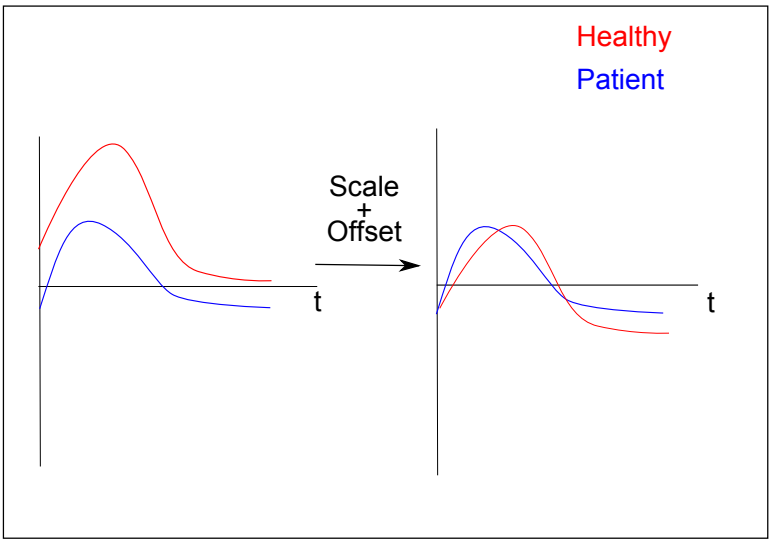


Figure A.3: The generated timeseries of the healthy population is scaled and added to an offset so that the beginning of the generated timeseries and the maximum of the generated timeseries match those of the patient's repetition timeseries.

repetition timeseries, we assume that the phase remains constant until the next equal value is found. Furthermore, we assume that the phase at the end of the patient's repetition timeseries is 2π (See Fig. A.5). Fig. A.4 illustrates the resulting phase for one repetition timeseries.

After the phase vector is generated for the patient's repetition timeseries, a DMP model is trained for the *repetition set*. The weights of the resulting DMP model are considered as the feature vector for this *repetition set*. The measure of performance for each repetition set is then calculated as the Euclidean distance between the weights of each repetition set and the weights of the healthy population model. Fig. A.6a depicts the result of this approach for knee extension/flexion of patient 8. As can be seen from the figure, the approach is not capable of capturing the trend of progress. This observation can be explained by the fact that the temporal information of patients' motions contain valuable information about their progress. By training the DMPs using the phase adjustment approach, this temporal information is neglected.

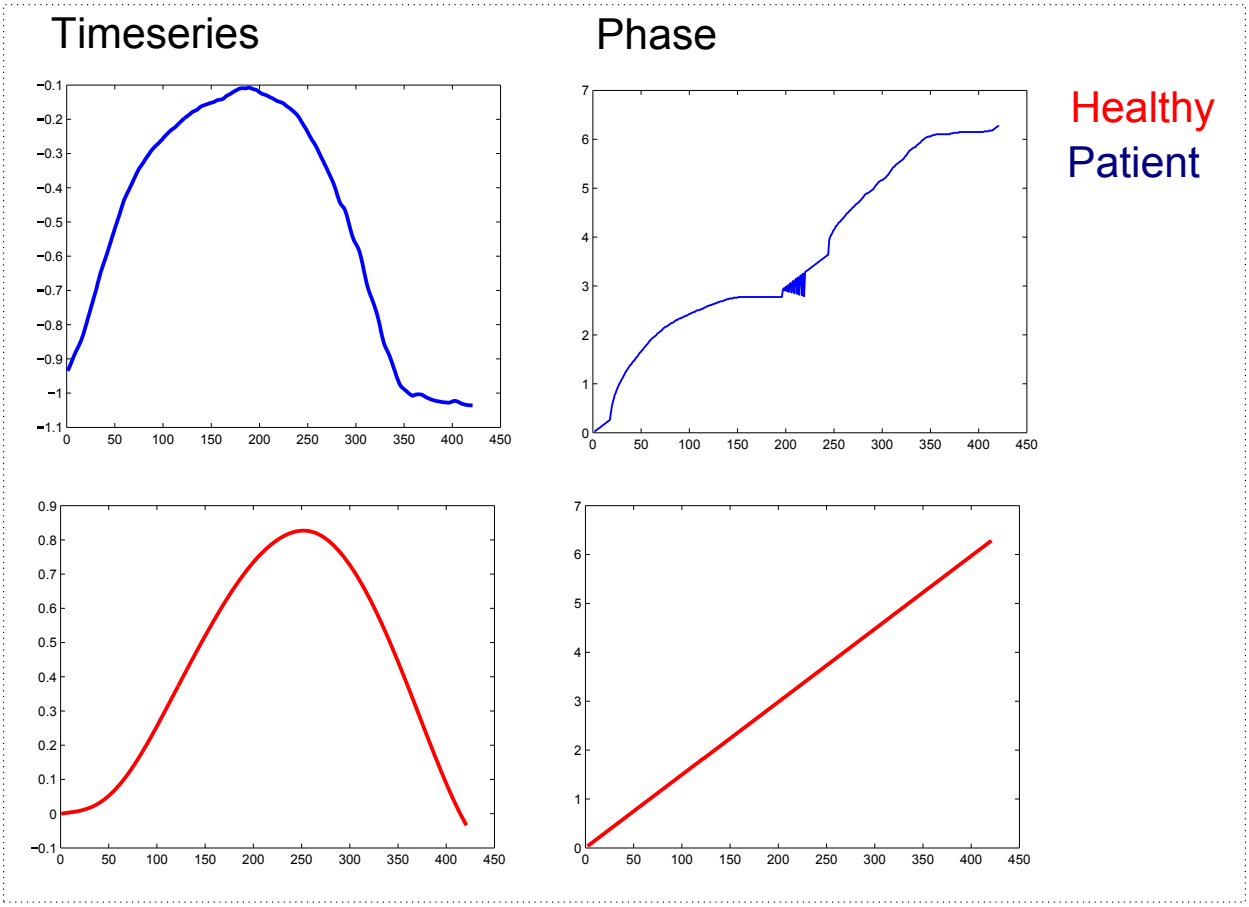


Figure A.4: The resulting phase for a repetition timeseries.

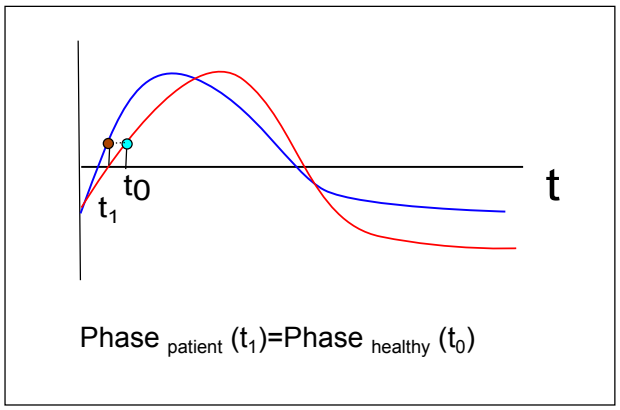
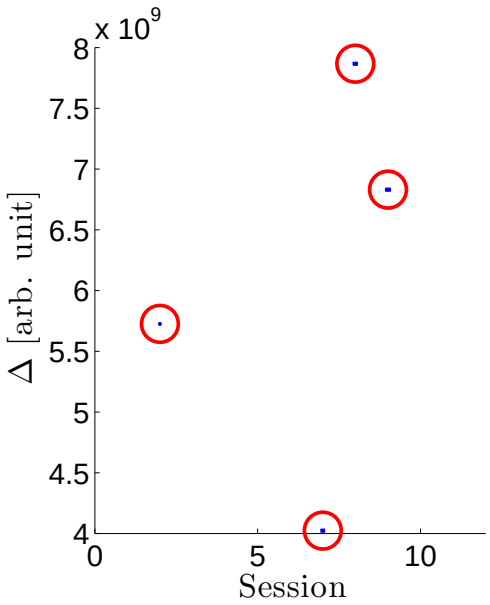
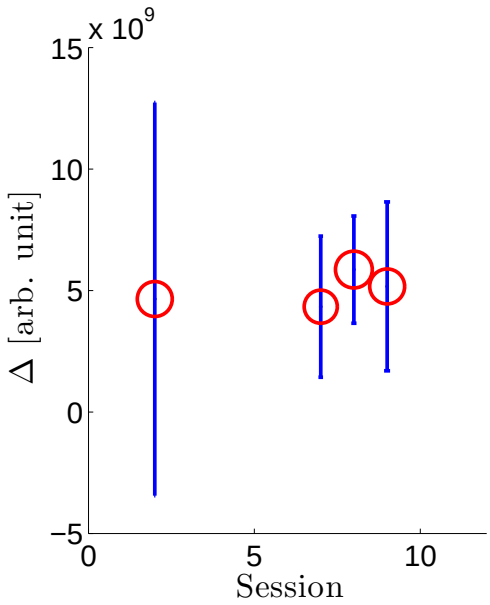


Figure A.5: The phase for the patient population is generated based on the generated timeseries of the healthy population. At each timestep, the value equal to the generated timeseries of the healthy population is found in the patient's repetition timeseries, and the phase of the point is assigned equal to that of the generated timeseries. If an equal value is not found it is assumed that the phased has remained constant. The motion of patients often contains perturbation when they reach the desired posture. This perturbation results in jumps in phase as can be seen in the figure.



(a) First Phase Adjusted DMP-Based Approach



(b) Second Phase Adjusted DMP-Based Approach

Figure A.6: The measure of progress Δ of knee extension/flexion exercise calculated for patient 8. As can be seen from the figure, no visible trend of progressed is observed for the two methods.

Appendix B

Patient Information

Table 6.1 summarizes the patient information and any confounding factors. Some patients are discharged significantly earlier than others indicating an early onset of acceptable performance through rehabilitation. Our approaches are able to identify this for patients 16 and 7. Some patients suffer from special circumstances which may have influenced their performance through rehabilitation , e.g., patient 6 suffers from chronic pain.

Appendix C

Top Features

Table 2 summarizes the feature selection results for the feature based method. Table 3 summarizes the feature selection results for the HMM based method.

Table C.1: Top features selected by Lasso for feature based method.

Knee Extension	**Knee Hip Extension**	**Squat**
min_{q_4}	min_{q_4}	$min_{\dot{q_1}}$
$mean_{q_1}$	$mean_{q_2}$	max_{q_4}
$mean\ddot{q}_4$	max_{q_2}	$max_{\dot{q_4}}$
$romq_4$	rom_{q_4}	$romq_1$
$time$	$time$	$time$

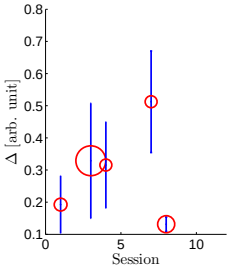
Table C.2: Top features selected by Lasso for HMM based method.

Knee Extension	**Knee Hip Extension**	**Squat**
$\sigma_{3_{\dot{q_4}}}$	$\sigma_{2_{q_4}}$	$\mu_{2_{q_1}}$
$\sigma_{1_{\dot{q_3}}}$	$\sigma_{2_{\ddot{q_3}}}$	$\sigma_{1_{q_2}}$
$\sigma_{1_{\dot{q_5}}}$	$\sigma_{2_{q_1}}$	$\mu_{1_{\ddot{q_1}}}$
$\sigma_{1_{q_4}}$	$\sigma_{3_{q_2}}$	$\sigma_{3_{\dot{q_4}}}$
$\sigma_{3_{\dot{q_5}}}$	$\sigma_{3_{q_4}}$	$\mu_{2_{q_1}}$

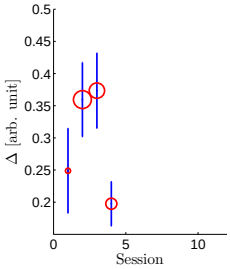
Appendix D

Measure of Performance Δ

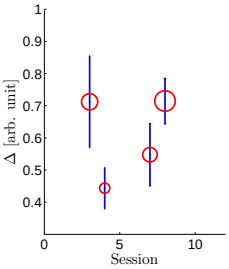
Figs. D.1, D.2, D.3, D.4, D.5, and D.6 illustrate the results for the distance measures Δ calculated for each *repetition set* using feature-based approach with feature selection on healthy and patient population (FBA FHP). Fig. D.7, D.8, D.9, D.10,D.11, and D.12 depict the results for the distance measures Δ calculated for each *repetition set* using feature-based approach with feature selection on healthy population (FBA FH).



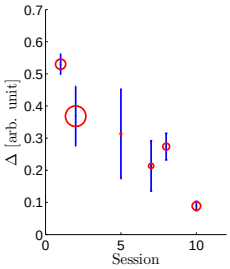
(a) P1 knee extension



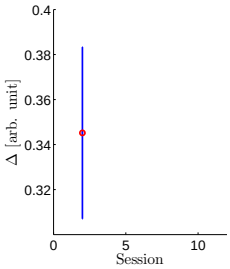
(b) P1 knee hip extension/Flexion



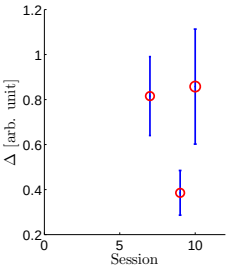
(c) P1 squat



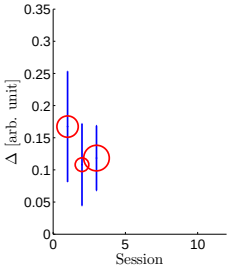
(d) P2 knee extension



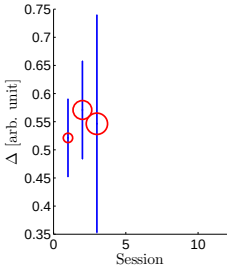
(e) P2 knee hip extension/Flexion



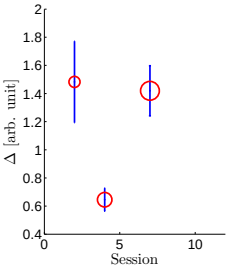
(f) P2 squat



(g) P3 knee extension

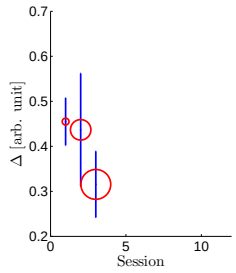


(h) P3 knee hip extension/Flexion

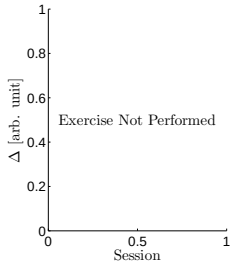


(i) P3 squat

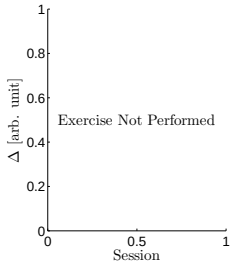
Figure D.1: Results for the distance measure δ for patients 1-3 calculated using the FBA FHP are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



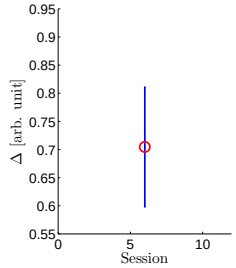
(a) P4 knee extension



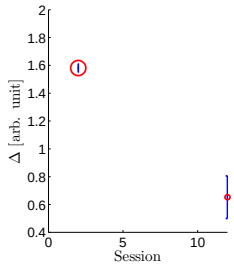
(b) P4 knee hip extension/Flexion



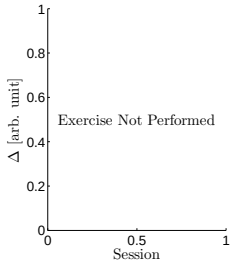
(c) P4 squat



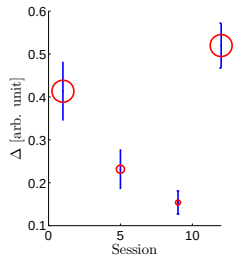
(d) P5 knee extension



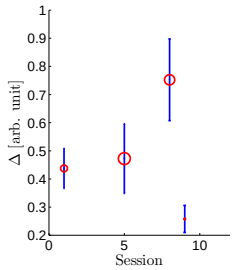
(e) P5 knee hip extension/Flexion



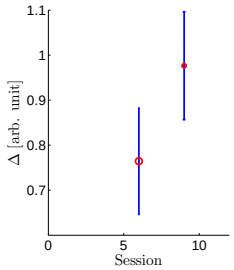
(f) P5 squat



(g) P6 knee extension

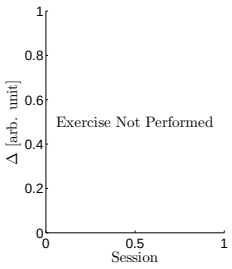


(h) P6 knee hip extension/Flexion

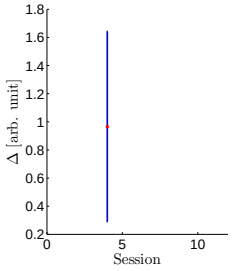


(i) P6 squat

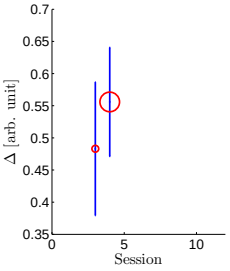
Figure D.2: Results for the distance measure δ for patients 4-6 calculated using the FBA FHP are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



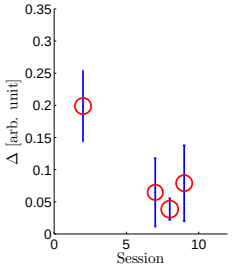
(a) P7 knee extension



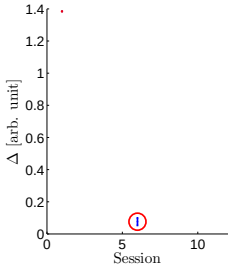
(b) P7 knee hip extension/Flexion



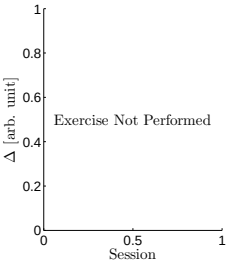
(c) P7 squat



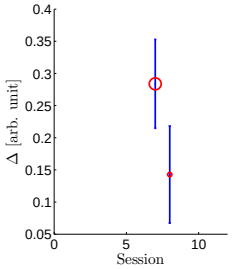
(d) P8 knee extension



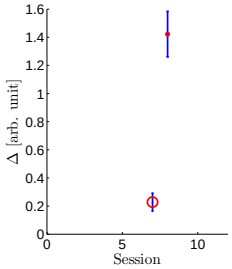
(e) P8 knee hip extension/Flexion



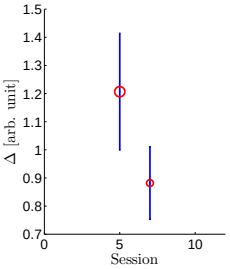
(f) P8 squat



(g) P9 knee extension

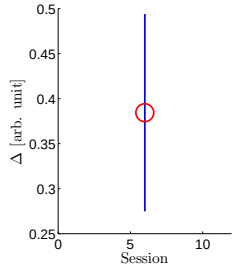


(h) P9 knee hip extension/Flexion

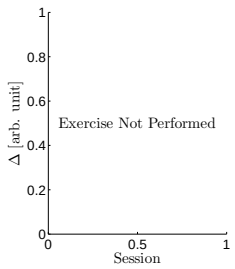


(i) P9 squat

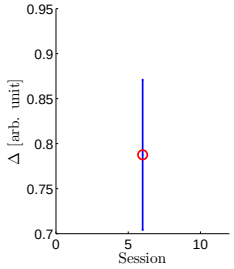
Figure D.3: Results for the distance measure δ calculated for patients 7-9 using the FBA FHP are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



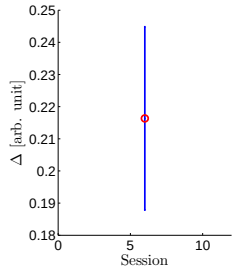
(a) P10 knee extension



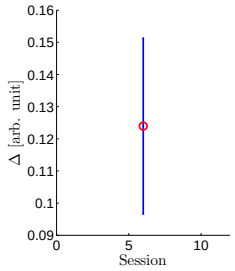
(b) P10 knee hip extension/Flexion



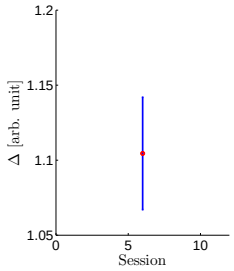
(c) P10 squat



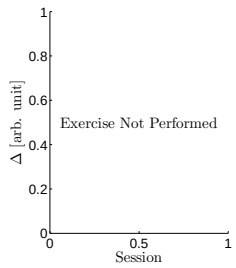
(d) P11 knee extension



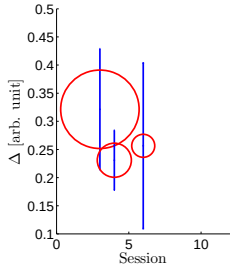
(e) P11 knee hip extension/Flexion



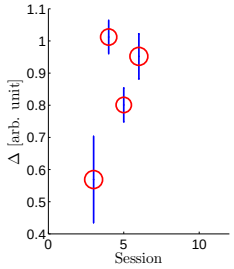
(f) P11 squat



(g) P12 knee extension

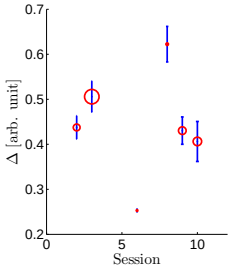


(h) P12 knee hip extension/Flexion

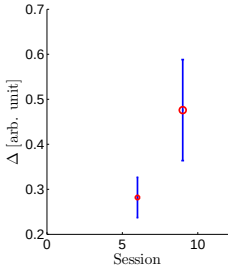


(i) P12 squat

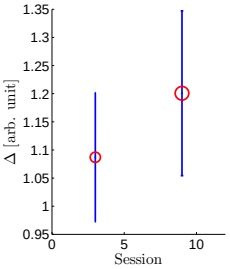
Figure D.4: Results for the distance measure δ calculated for patients 10-12 using the FBA FHP are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



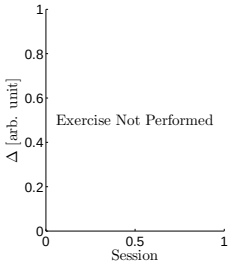
(a) P13 knee extension



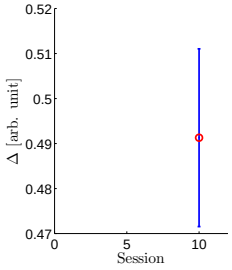
(b) P13 knee hip extension/Flexion



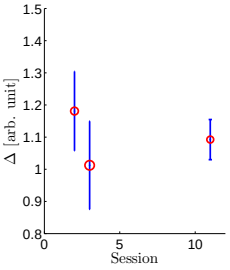
(c) P13 squat



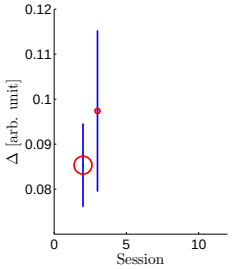
(d) P14 knee extension



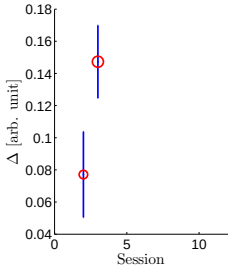
(e) P14 knee hip extension/Flexion



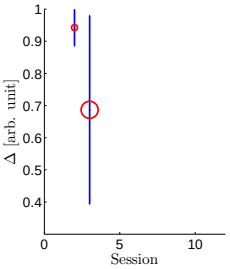
(f) P14 squat



(g) P15 knee extension

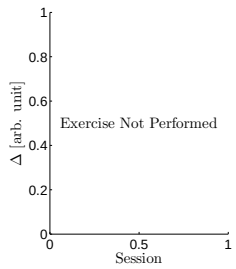


(h) P15 knee hip extension/Flexion

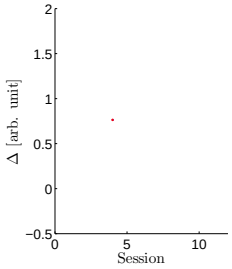


(i) P15 squat

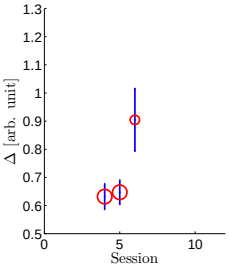
Figure D.5: Results for the distance measure δ calculated for patients 13-16 using the FBA FHP are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



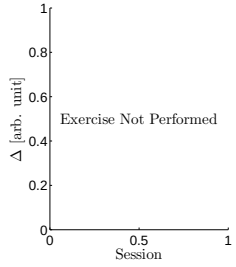
(a) P16 knee extension



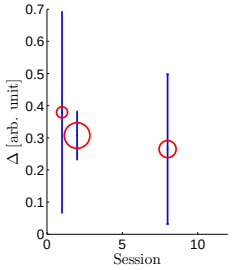
(b) P16 knee hip extension/Flexion



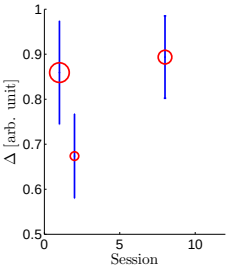
(c) P16 squat



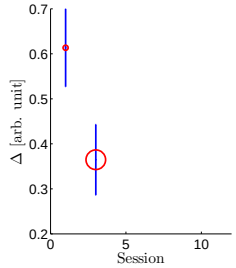
(d) P17 knee extension



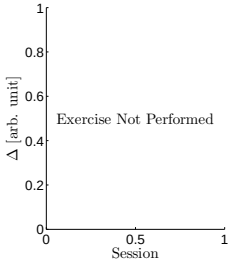
(e) P17 knee hip extension/Flexion



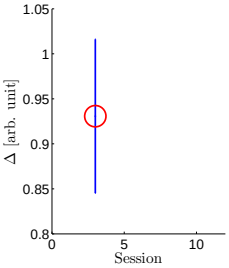
(f) P17 squat



(g) P18 knee extension

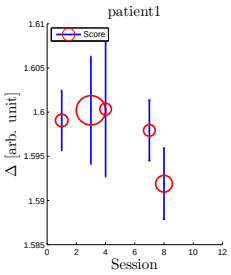


(h) P18 knee hip extension/Flexion

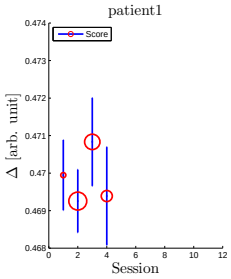


(i) P18 squat

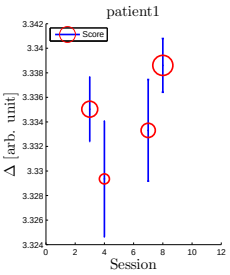
Figure D.6: Results for the distance measure δ for patients 16-18 calculated using the FBA FHP are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



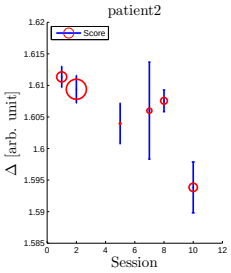
(a) P1 knee extension



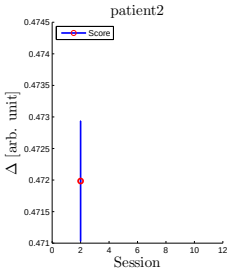
(b) P1 knee hip extension/Flexion



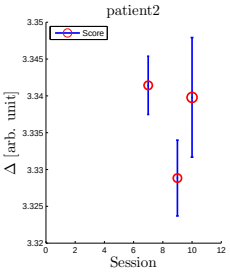
(c) P1 squat



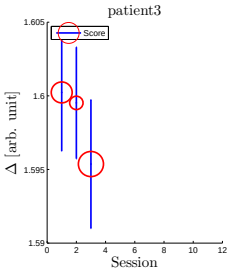
(d) P2 knee extension



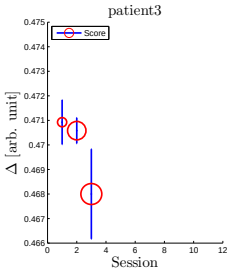
(e) P2 knee hip extension/Flexion



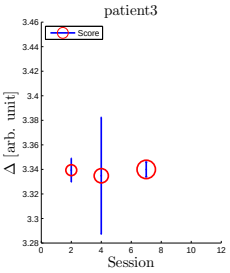
(f) P2 squat



(g) P3 knee extension

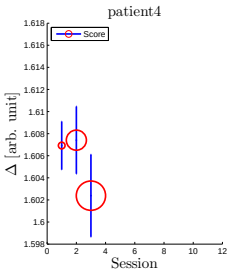


(h) P3 knee hip extension/Flexion

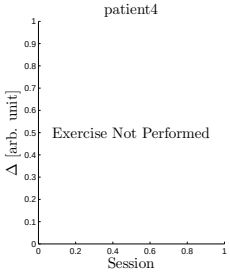


(i) P3 squat

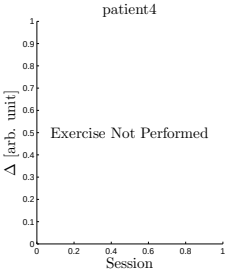
Figure D.7: Results for the distance measure δ for patients 1-3 calculated using the FBA FH are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



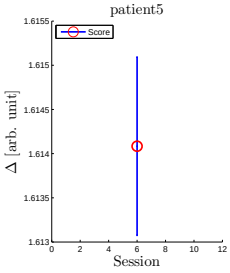
(a) P4 knee extension



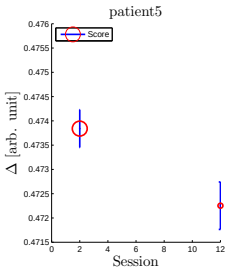
(b) P4 knee hip extension/Flexion



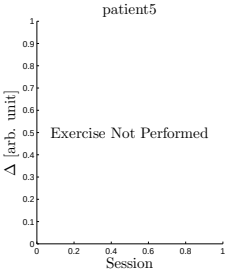
(c) P4 squat



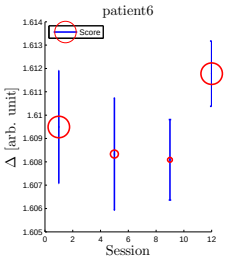
(d) P5 knee extension



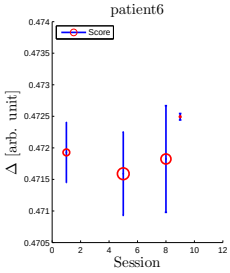
(e) P5 knee hip extension/Flexion



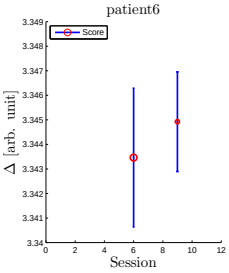
(f) P5 squat



(g) P6 knee extension

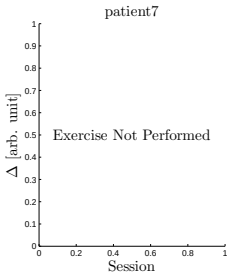


(h) P6 knee hip extension/Flexion

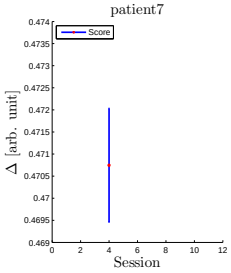


(i) P6 squat

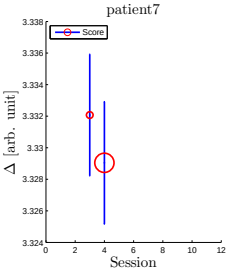
Figure D.8: Results for the distance measure δ for patients 4-6 calculated using the FBA FH are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



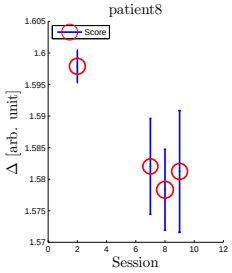
(a) P7 knee extension



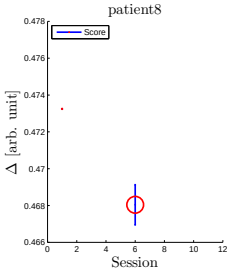
(b) P7 knee hip extension/Flexion



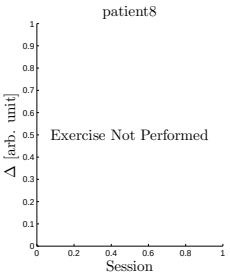
(c) P7 squat



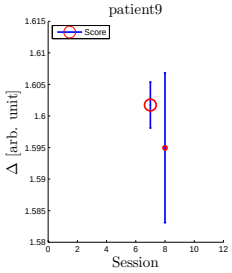
(d) P8 knee extension



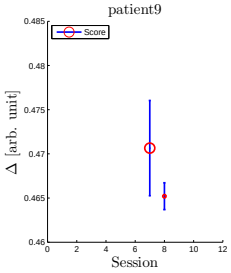
(e) P8 knee hip extension/Flexion



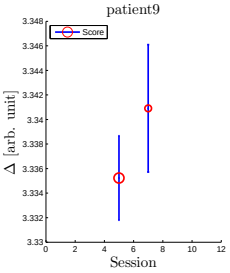
(f) P8 squat



(g) P9 knee extension

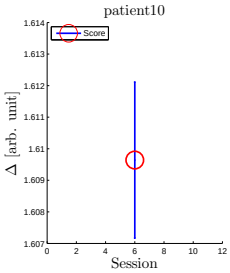


(h) P9 knee hip extension/Flexion

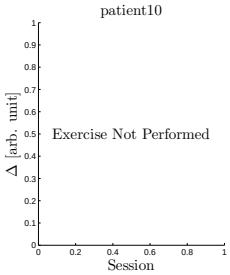


(i) P9 squat

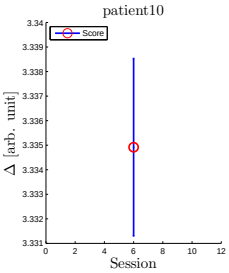
Figure D.9: Results for the distance measure δ calculated for patients 7-9 using the FBA FH are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



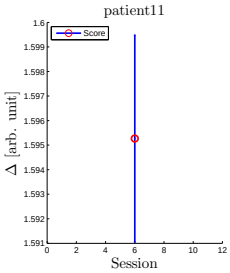
(a) P10 knee extension



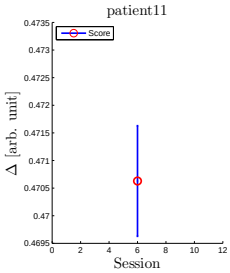
(b) P10 knee hip extension/Flexion



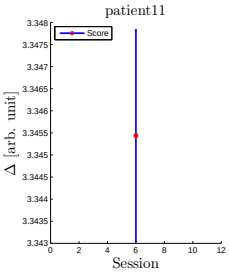
(c) P10 squat



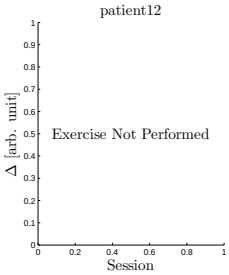
(d) P11 knee extension



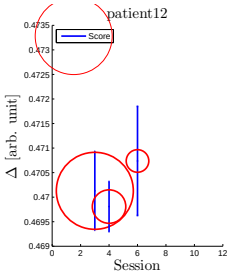
(e) P11 knee hip extension/Flexion



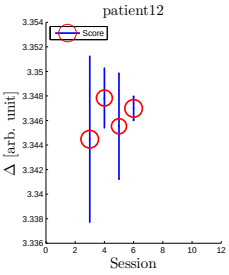
(f) P11 squat



(g) P12 knee extension

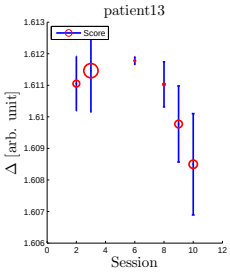


(h) P12 knee hip extension/Flexion

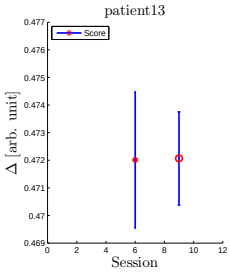


(i) P12 squat

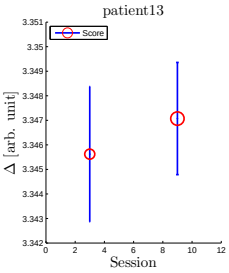
Figure D.10: Results for the distance measure δ calculated for patients 10-12 using the FBA FH are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



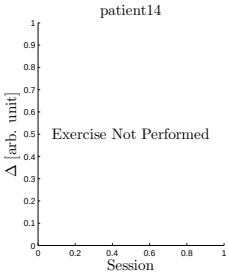
(a) P13 knee extension



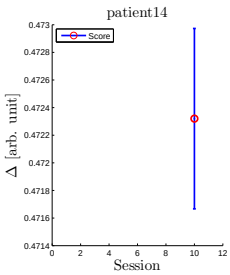
(b) P13 knee hip extension/Flexion



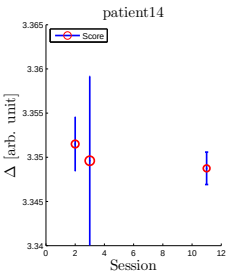
(c) P13 squat



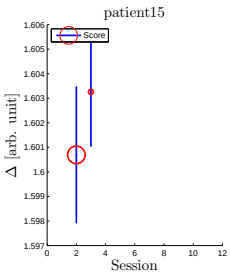
(d) P14 knee extension



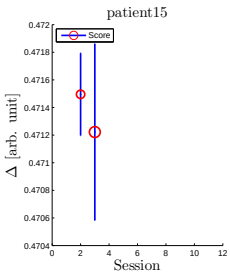
(e) P14 knee hip extension/Flexion



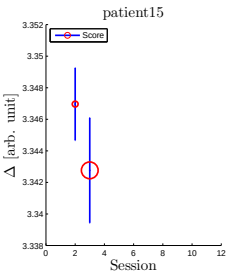
(f) P14 squat



(g) P15 knee extension

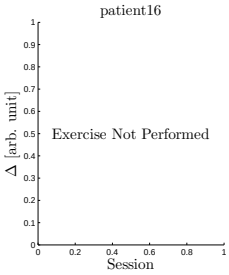


(h) P15 knee hip extension/Flexion

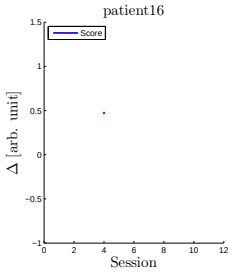


(i) P15 squat

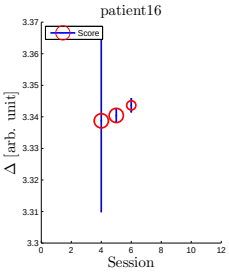
Figure D.11: Results for the distance measure δ calculated for patients 13-15 using the FBA FH are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.



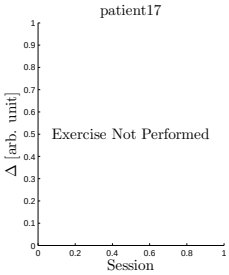
(a) P16 knee extension



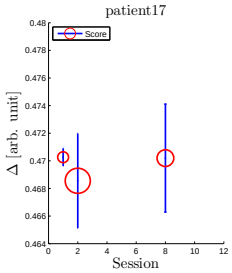
(b) P16 knee hip extension/Flexion



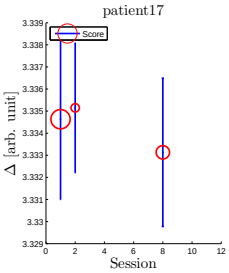
(c) P16 squat



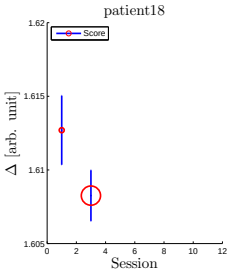
(d) P17 knee extension



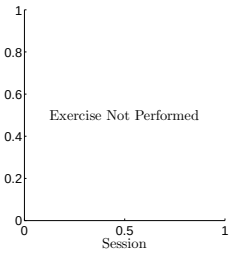
(e) P17 knee hip extension/Flexion



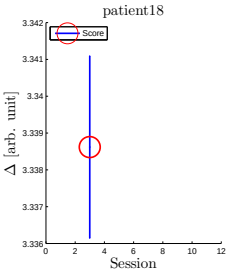
(f) P17 squat



(g) P18 knee extension



(h) P18 knee hip extension/Flexion



(i) P18 squat

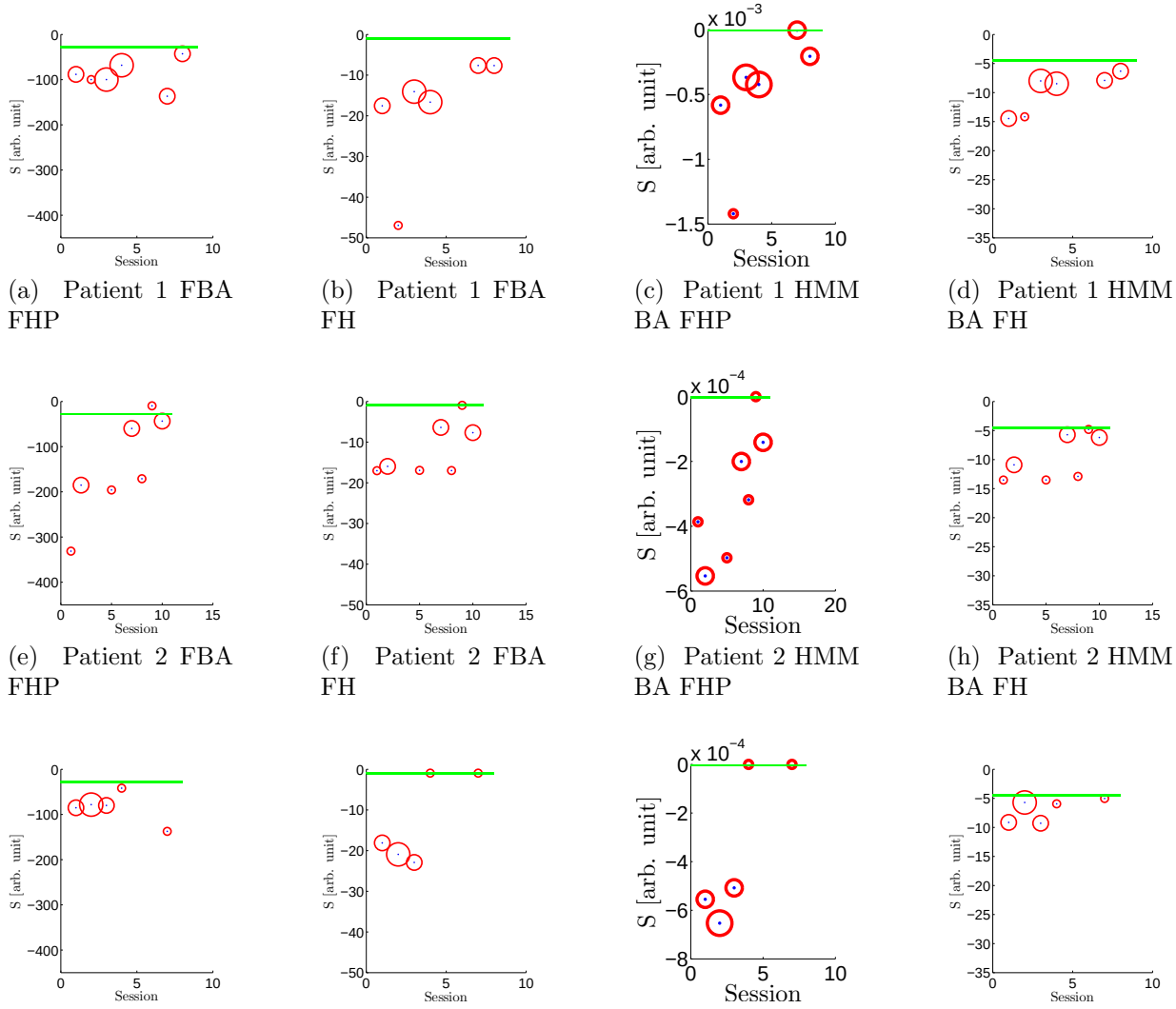
Figure D.12: Results for the distance measure δ for patients 16-18 calculated using the FBA FH are shown for three exercises. The red circle ○ illustrates the median of the distance measures (i.e. Δ) in each session and the blue bar depicts the variance of the distance measures δ in one session. The size of the circle indicates the number of repetitions available in each *repetition set*.

Appendix E

Overall Score S

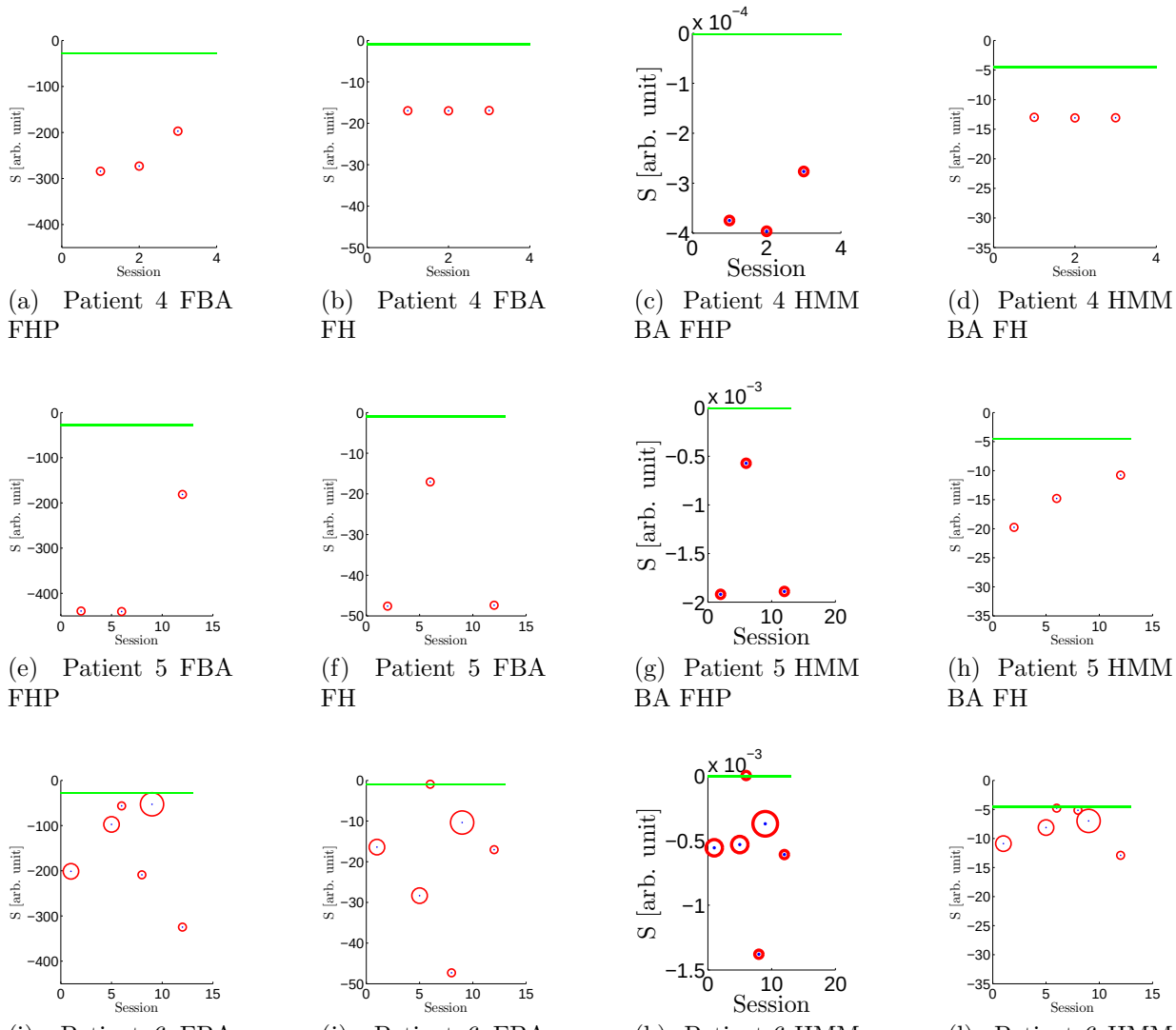
E.1 The Feature-Based Approach and The HMM-Based Approach

Figs. E.1, E.2, E.3, E.4, E.5, and E.6 show the score measure S for each patient for every session performing the three exercises. The score measure S is calculated using the feature-based approach with feature selection on healthy and patient population (FBA FHP), feature-based approach with feature selection on healthy population (FBA FH), HMM-based approach with feature selection on healthy and patient population (HMM BA FHP), and HMM-based approach with feature selection on healthy population (HMM BA FH). These four methods capture the same trend of progress for more than 85% of the cases and differences occur mostly when the patient's condition remains unchanged over the course of rehabilitation. Only patients 1, 2, 3, 5, 6, 7 are included in feature selection for FBA FHP and HMM BA FHP. The score is formulated to account for exercise difficulty. As the exercises become more difficult, the score focuses more on the fact that the patient is able to perform them rather than focusing on how well they are performed. This can be seen in the results of patient 13 where the individual exercises in Figs. D.5a, D.5b, D.5c do not show an absolute trend of progress but the score considers the difficulty of the exercises and the number of repetitions and therefore patient improvement can be seen in the score as shown in Fig. E.5a.



(a) Patient 1 FBA FHP
(b) Patient 1 FBA FH
(c) Patient 1 HMM BA FHP
(d) Patient 1 HMM BA FH
(e) Patient 2 FBA FHP
(f) Patient 2 FBA FH
(g) Patient 2 HMM BA FHP
(h) Patient 2 HMM BA FH
(i) Patient 3 FBA FHP
(j) Patient 3 FBA FH
(k) Patient 3 HMM BA FHP
(l) Patient 3 HMM BA FH

Figure E.1: Figures depict the score measure S (○) for patients 1-3. The green line — shows the best score of the patients in their last physiotherapy session.



(a) Patient 4 FBA FHP (b) Patient 4 FBA FH (c) Patient 4 HMM BA FHP (d) Patient 4 HMM BA FH

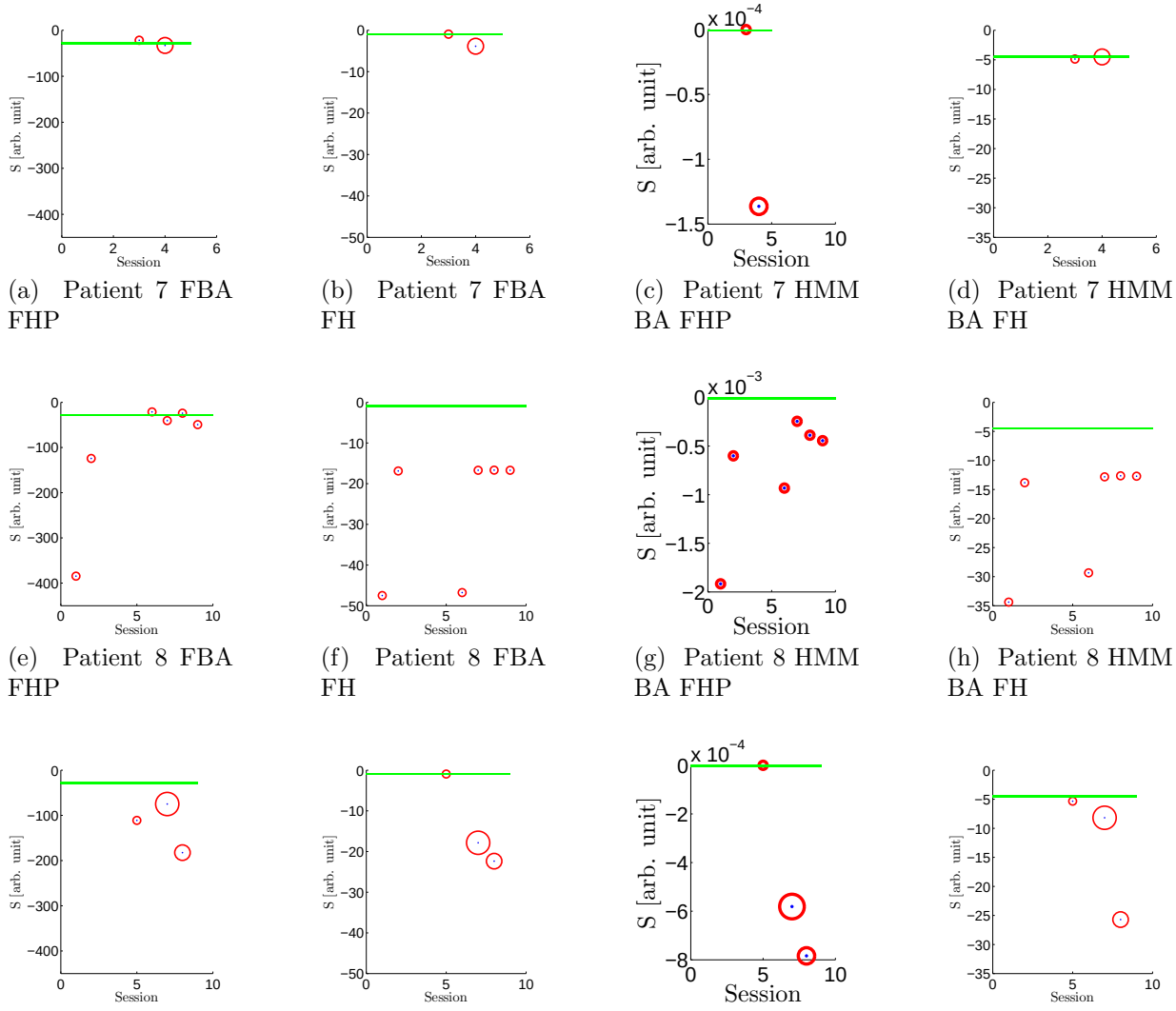
(e) Patient 5 FBA FHP (f) Patient 5 FBA FH (g) Patient 5 HMM BA FHP (h) Patient 5 HMM BA FH

(i) Patient 6 FBA FHP (j) Patient 6 FBA FH (k) Patient 6 HMM BA FHP (l) Patient 6 HMM BA FH

Figure E.2: Figures depict the score measure S (○) for patients 4-6. The green line — shows the best score of the patients in their last physiotherapy session.

E.2 The DMP-Based Approach and The Kernel-Based Approach

Figs. E.7, E.8, E.11, and E.12 show the score measure S for each patient for every session performing the three exercises. The score measure S is calculated using the kernel-based



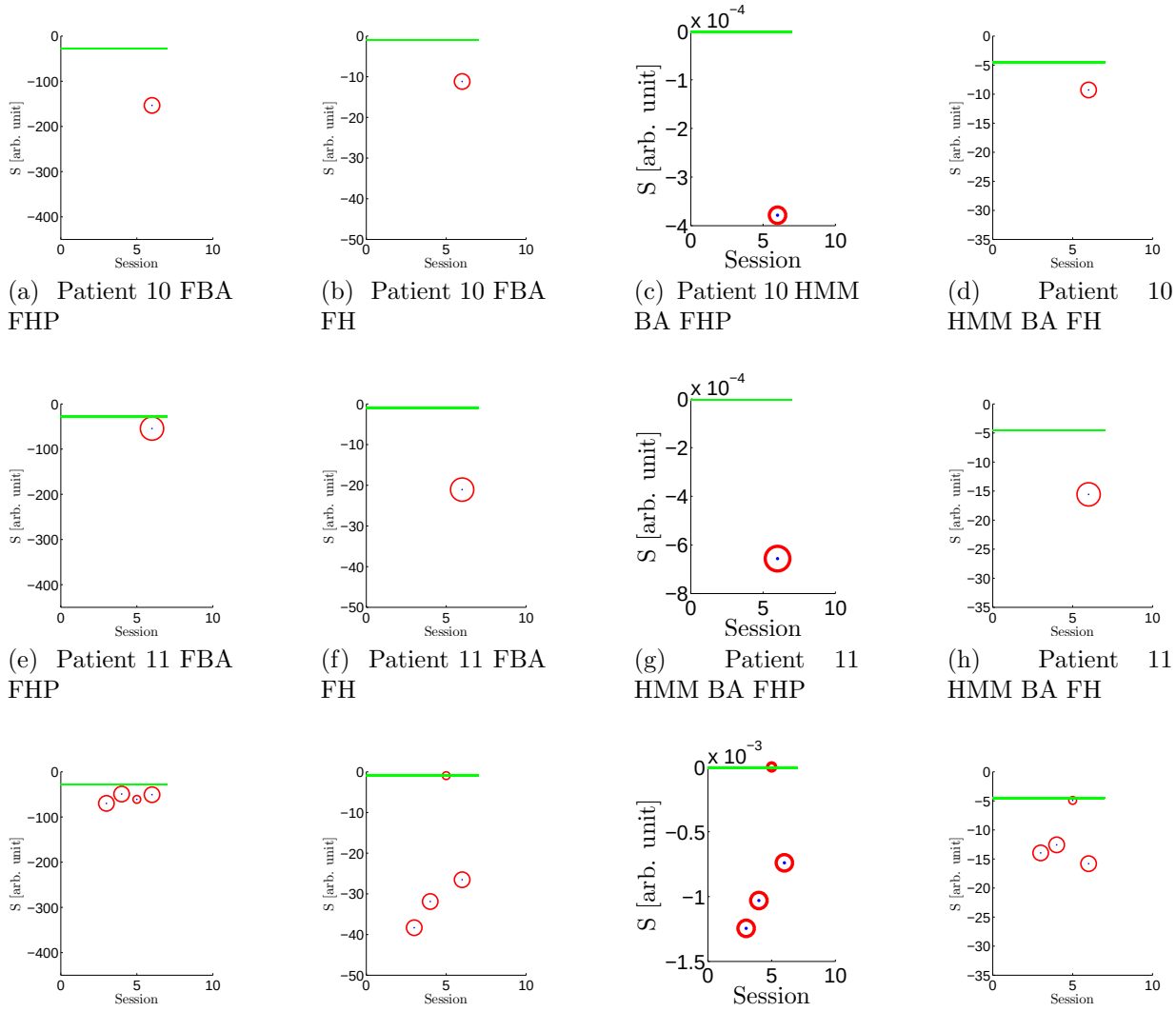
(a) Patient 7 FBA FHP (b) Patient 7 FBA FH (c) Patient 7 HMM BA FHP (d) Patient 7 HMM BA FH

(e) Patient 8 FBA FHP (f) Patient 8 FBA FH (g) Patient 8 HMM BA FHP (h) Patient 8 HMM BA FH

(i) Patient 9 FBA FHP (j) Patient 9 FBA FH (k) Patient 9 HMM BA FHP (l) Patient 9 HMM BA FH

Figure E.3: Figures depict the score measure S (○) for patients 7-9 The green line — shows the best score of the patients in their last physiotherapy session.

approach, and the DMP-based approach. Both approaches are capable of capturing the trend of progress for some patients,

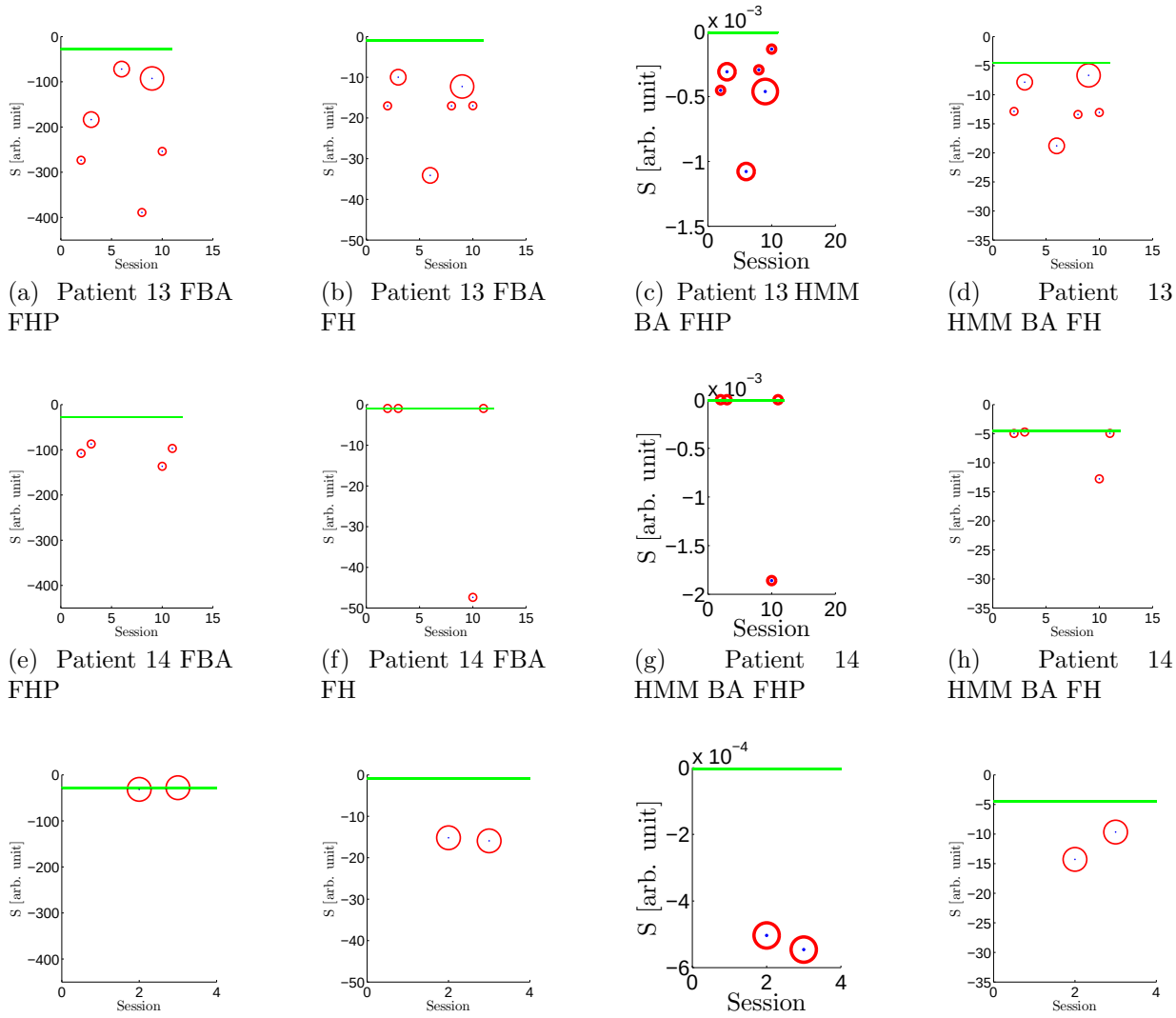


(a) Patient 10 FBA FHP (b) Patient 10 FBA FH (c) Patient 10 HMM BA FHP (d) Patient 10 HMM BA FH

(e) Patient 11 FBA FHP (f) Patient 11 FBA FH (g) Patient 11 HMM BA FHP (h) Patient 11 HMM BA FH

(i) Patient 12 FBA FHP (j) Patient 12 FBA FH (k) Patient 12 HMM BA FHP (l) Patient 12 HMM BA FH

Figure E.4: Figures depict the score measure S (○) for patients 10-12. The green line — shows the best score of the patients in their last physiotherapy session.



(a) Patient 13 FBA FHP

(b) Patient 13 FBA FH

(c) Patient 13 HMM BA FHP

(d) Patient 13 HMM BA FH

(e) Patient 14 FBA FHP

(f) Patient 14 FBA FH

(g) Patient 14 HMM BA FHP

(h) Patient 14 HMM BA FH

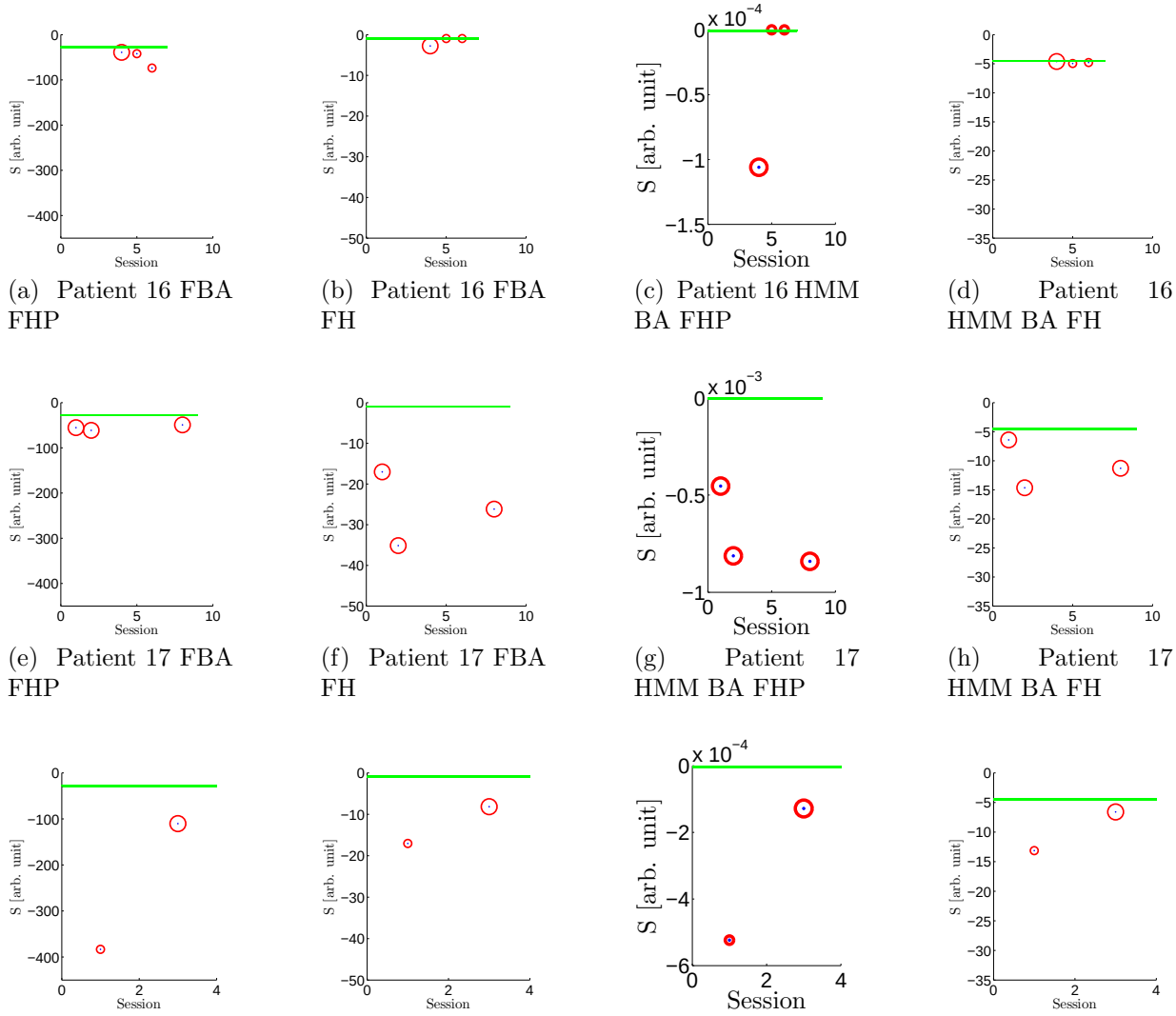
(i) Patient 15 FBA FHP

(j) Patient 15 FBA FH

(k) Patient 15 HMM BA FHP

(l) Patient 15 HMM BA FH

Figure E.5: Figures depict the score measure S (○) for patients 13-15. The green line — shows the best score of the patients in their last physiotherapy session.



(a) Patient 16 FBA FHP

(b) Patient 16 FBA FH

(c) Patient 16 HMM BA FHP

(d) Patient 16 HMM BA FH

(e) Patient 17 FBA FHP

(f) Patient 17 FBA FH

(g) Patient 17 HMM BA FHP

(h) Patient 17 HMM BA FH

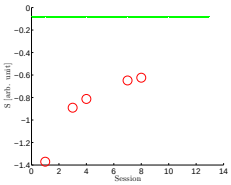
(i) Patient 18 FBA FHP

(j) Patient 18 FBA FH

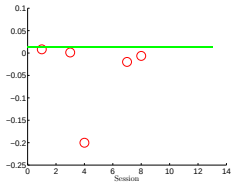
(k) Patient 18 HMM BA FHP

(l) Patient 18 HMM BA FH

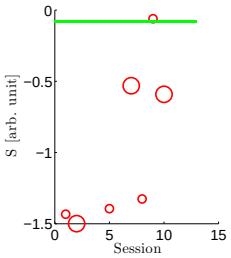
Figure E.6: Figures depict the score measure S (○) for patients 16-18. The green line — shows the best score of the patients in their last physiotherapy session.



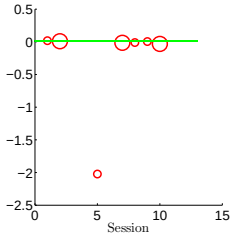
(a) Patient 1 K BA



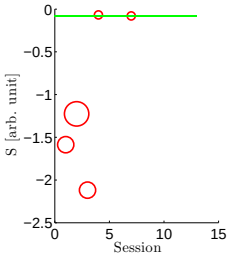
(b) Patient 1 DMP BA



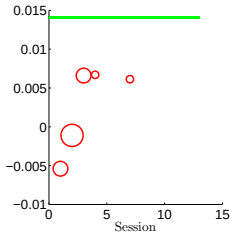
(c) Patient 2 K BA



(d) Patient 2 DMP BA

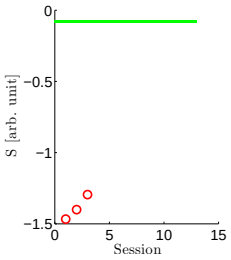


(e) Patient 3 K BA

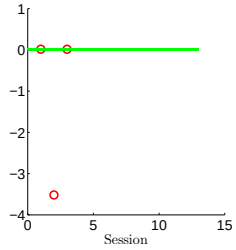


(f) Patient 3 DMP BA

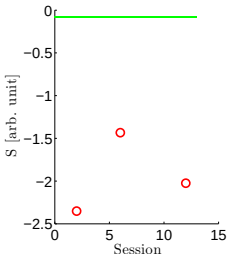
Figure E.7: Figures depict the score measure S (○) for patients 1-3. The green line — shows the best score of the patients in their last physiotherapy session.



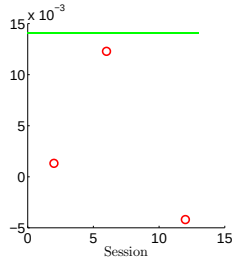
(a) Patient 4 K BA



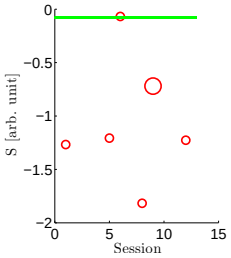
(b) Patient 4 DMP BA



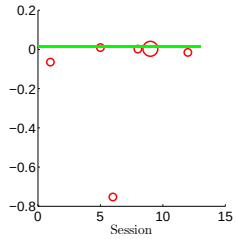
(c) Patient 5 K BA



(d) Patient 5 DMP BA

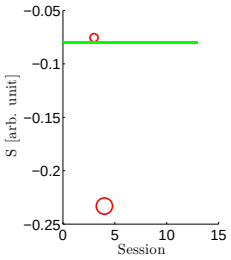


(e) Patient 6 K BA

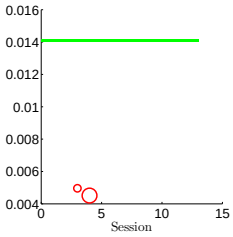


(f) Patient 6 DMP BA

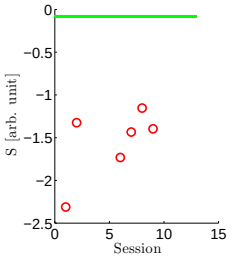
Figure E.8: Figures depict the score measure S (○) for patients 4-6. The green line — shows the best score of the patients in their last physiotherapy session.



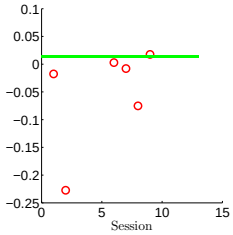
(a) Patient 7 K BA



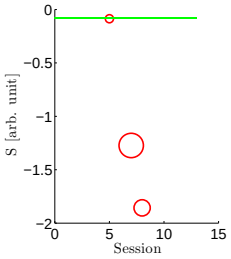
(b) Patient 7 DMP BA



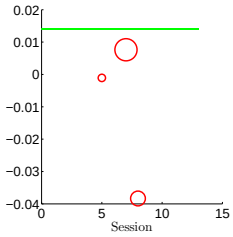
(c) Patient 8 K BA



(d) Patient 8 DMP BA

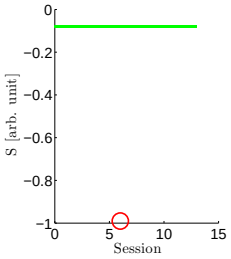


(e) Patient 9 K BA

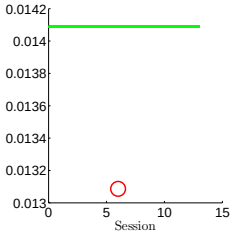


(f) Patient 9 DMP BA

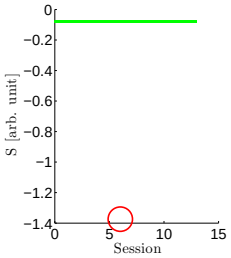
Figure E.9: Figures depict the score measure S (○) for patients 7-9. The green line — shows the best score of the patients in their last physiotherapy session.



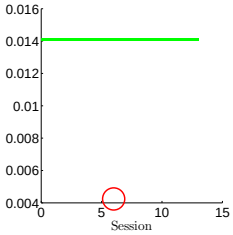
(a) Patient 10 K BA



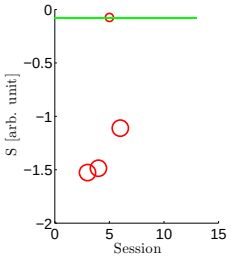
(b) Patient 10 DMP BA



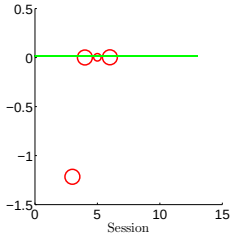
(c) Patient 11 K BA



(d) Patient 11 DMP BA

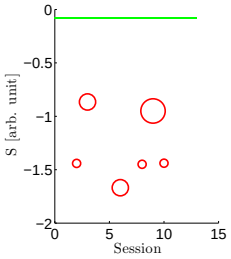


(e) Patient 12 K BA

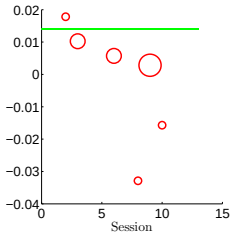


(f) Patient 12 DMP BA

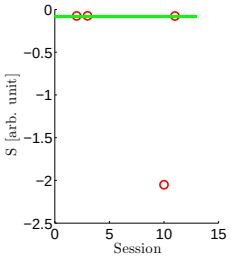
Figure E.10: Figures depict the score measure S (○) for patients 10-12. The green line — shows the best score of the patients in their last physiotherapy session.



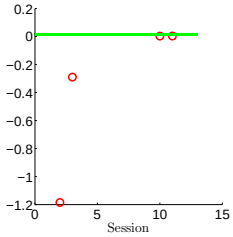
(a) Patient 13 K BA



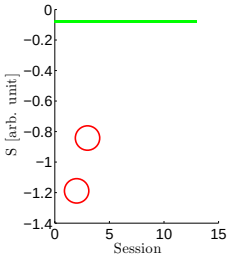
(b) Patient 13 DMP BA



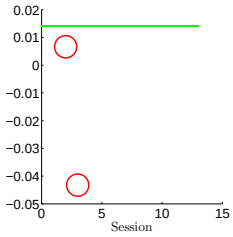
(c) Patient 14 K BA



(d) Patient 14 DMP BA

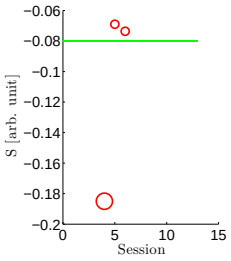


(e) Patient 15 K BA

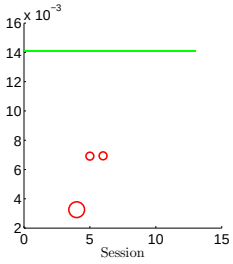


(f) Patient 15 DMP BA

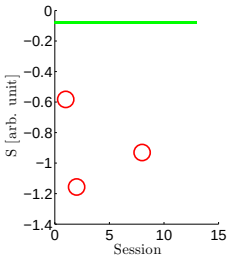
Figure E.11: Figures depict the score measure S (○) for patients 13-15. The green line — shows the best score of the patients in their last physiotherapy session.



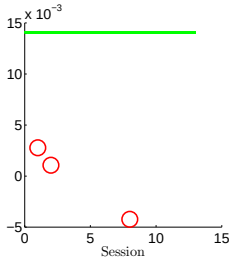
(a) Patient 16 K BA



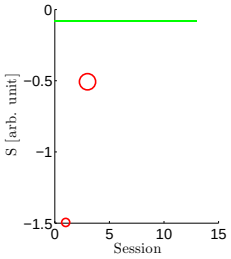
(b) Patient 16 DMP BA



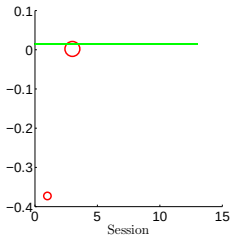
(c) Patient 17 K BA



(d) Patient 17 DMP BA



(e) Patient 18 K BA



(f) Patient 18 DMP BA

Figure E.12: Figures depict the score measure S (○) for patients 16-18. The green line — shows the best score of the patients in their last physiotherapy session.

References

[1] JK Aggarwal and M S Ryoo. Human activity analysis: A review. *ACM COMPUT SURV*, 43:16, 2011.

[2] R. Ali, L. Atallah, B. Lo, and G. Yang. Detection and analysis of transitional activity in manifold space. *IEEE T INF TECHNOL B*, 16(1):119–128, 2012.

[3] J Alon, S. Sclaroff, G. Kollios, and V. Pavlovic. Discovering clusters in motion time-series data. In *CVPR*, pages I–375–I–381, 2003.

[4] E Barshan, A Ghodsi, Z Azimifar, and M Zolghadri Jahromi. Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *PATTERN RECOGN*, 44(7):1357–1371, 2011.

[5] G Baudat and F Anouar. Generalized discriminant analysis using a kernel approach. *NEURAL COMPUT*, 12(10):2385–2404, 2000.

[6] R Begg and J Kamruzzaman. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *J BIOMECH*, 38 (3):401–408, 2005.

[7] J Ben-Arie, Z Wang, P Pandit, and S Rajaram. Human activity recognition using multidimensional indexing. *PAMI*, 24(8):1091–1104, 2002.

[8] M Z Bendjaballah, A Shirazi-Adl, and DJ Zukor. Biomechanics of the human knee joint in compression: reconstruction, mesh generation and finite element analysis. *KNEE*, 2(2):69–79, 1995.

[9] L Bianchi, D Angelini, GP Orani, and F Lacquaniti. Kinematic coordination in human gait: relation to mechanical energy cost. *J NEUROPHYSIOL*, 79(4):2155–2170, 1998.

[10] EV Biryukova, A Roby-Brami, AA Frolov, and M Mokhtari. Kinematics of human arm reconstructed from spatial tracking system recordings. *J BIOMECH*, 33(8): 985–995, 2000.

[11] R Blake and M Shiffrar. Perception of human motion. *Annu. Rev. Psychol.*, 58: 47–73, 2007.

[12] C. Bregler. Learning and recognizing human dynamics in video sequences. In *CVPR*, pages 568–574, Jun 1997.

[13] H Buelow and A Birk. Gesture-recognition as basis for a human robot interface (hri) on a auv. In *OCEANS 2011*, pages 1–9. IEEE, 2011.

[14] E Burdet and TE Milner. Quantization of human motions and learning of accurate movements. *Biol Cybern*, 78(4):307–318, 1998.

[15] A Burns, B R Greene, M J McGrath, T J O'Shea, B Kuris, S M Ayer, F Stroiescu, and V Cionca. Shimmer–a wireless sensor platform for noninvasive biomedical research. *SEN*, 10(9):1527–1534, 2010.

[16] A Camurri, I Lagerlöf, and G Volpe. Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *INT J HUM-COMPUT ST*, 59(1):213–225, 2003.

[17] A Camurri, B Mazzarino, M Ricchetti, R Timmers, and G Volpe. Multimodal analysis of expressive gesture in music and dance performances. In *Gesture-based communication in human-computer interaction*, pages 20–39. Springer, 2004.

[18] A Cappozzo. Gait analysis methodology. *HUM MOVEMENT SCI*, 3(1):27–50, 1984.

[19] S Chan, C Wah Ngo, and K F Lai. Motion tracking of human mouth by generalized deformable models. *PATTERN RECOGN LETT*, 20(9):879–887, 1999.

[20] EY Chao, RK Laughman, E Schneider, and RN Stauffer. Normative data of knee joint motion and ground reaction forces in adult level walking. *J BIOMECH*, 16(3): 219–233, 1983.

[21] A Chaudhary, J L Raheja, and S Raheja. A vision based geometrical method to find fingers positions in real time hand gesture recognition. *Journal of Software*, 7(4): 861–869, 2012.

[22] J-JJ Chen and R Shiavi. Temporal feature extraction and clustering analysis of electromyographic linear envelopes in gait studies. *BME*, 37(3):295–302, 1990.

[23] L Chiari, U D Croce, A Leardini, and A Cappozzo. Human movement analysis using stereophotogrammetry: Part 2: Instrumental errors. *GAIT POSTURE*, 21(2): 197–211, 2005.

[24] S Corazza, L Mündermann, AM Chaudhari, T Demattio, C Cobelli, and TP Andriacchi. A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach. *ANN BIOMED ENG*, 34(6):1019–1029, 2006.

[25] C Cortes and V Vapnik. Support vector machine. *Machine learning*, 20(3):273–297, 1995.

[26] P. A. W. Cox, D.R. Lewis. *The statistical analysis of series of events*. Methuen (London), 1966.

[27] N Dalal, B Triggs, and C Schmid. Human detection using oriented histograms of flow and appearance. In *ECCV*, pages 428–441. Springer, 2006.

[28] TT Dao and MC Ho Ba Tho. Knowledge-based system for orthopedic pediatric disorders. In *IFMBE*, pages 125–128. Springer, 2012.

[29] B Dariush. Human motion analysis for biomechanics and biomedicine. *MACH VISION APPL*, 14(4):202–205, 2003.

[30] D DeCarlo and D Metaxas. Deformable model-based shape and motion analysis from images using motion residual error. In *ICCV*, pages 113–119. IEEE, 1998.

[31] F Dobson, M E Morris, R Baker, and H K Graham. Gait classification in children with cerebral palsy: a systematic review. *GAIT POSTURE*, 25(1):140–152, 2007.

[32] M Ekinci. Human identification using gait. *Turk J Elec Engin*, 14(2):267–291, 2006.

[33] N El-Sheimy, S Nassar, and A Noureldin. Wavelet de-noising for imu alignment. *AES*, 19(10):32–39, 2004.

[34] J F-S Lin and D Kulić. Segmenting human motion for automated rehabilitation exercise analysis. In *EMBC*, pages 2881–2884, 2012.

[35] T J Gabbett and M J Mulvey. Time-motion analysis of small-sided training games and competition in elite women soccer players. *JSCR*, 22(2):543–552, 2008.

[36] D Glowinski, N Dael, A Camurri, G Volpe, M Mortillaro, and K Scherer. Toward a minimal representation of affective gestures. *T-AFFC*, 2:106–118, 2011.

[37] M Gönen and E Alpaydın. Multiple kernel learning algorithms. *JMLR*, 999999: 2211–2268, 2011.

[38] R D Green and L Guan. Quantifying and recognizing human movement patterns from monocular video images-part i: a new framework for modeling human motion. *CSVT*, 14(2):179–190, 2004.

[39] A Gretton, K M Borgwardt, M J Rasch, B Schölkopf, and A Smola. A kernel two-sample test. *JMLR*, 13:723–773, 2012.

[40] Ju Han and Bir Bhanu. Individual recognition using gait energy image. *PAMI*, 28 (2):316–322, 2006.

[41] H Hatze. The complete optimization of a human motion. *Math Biosci*, 28(1):99–135, 1976.

[42] H Hatze. A comprehensive model for human motion simulation and its application to the take-off phase of the long jump. *J BIOMECH*, 14(3):135–142, 1981.

[43] H Hatze. Quantitative analysis, synthesis and optimization of human motion. *HUM MOVEMENT SCI*, 3(1):5–25, 1984.

[44] J B Hayfron-Acquah, M S Nixon, and J N Carter. Automatic gait recognition by symmetry analysis. *PATTERN RECOGN LETT*, 24(13):2175–2183, 2003.

[45] J. A. Howe, E. L. Inness, A. Venturini, J. I. Williams, and M. C. Verrier. The community balance and mobility scale-a balance measure for individuals with traumatic brain injury. *Clinical Rehabilitation*, 20:885–895, 2006.

[46] T Huynh and B Schiele. Analyzing features for activity recognition. In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies*, pages 159–163. ACM, 2005.

[47] A J Ijspeert, J Nakanishi, H Hoffmann, P Pastor, and S Schaal. Dynamical movement primitives: learning attractor models for motor behaviors. *Neural computation*, 25 (2):328–373, 2013.

[48] A Jaimes and N Sebe. Multimodal human–computer interaction: A survey. *COMPUT VIS IMAGE UND*, 108:116–134, 2007.

[49] K Jia and D-Y Yeung. Human action recognition using local spatio-temporal discriminant embedding. In *CVPR*, pages 1–8, 2008.

[50] Y Jiang, I Hayashi, and S Wang. Embodied knowledge extraction from human motion using singular value decomposition. In *FUZZ-IEEE*, pages 1–8. IEEE, 2012.

[51] E Jovanov, A Milenkovic, C Otto, and P C De Groen. A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *J NEUROENG REHABIL*, 2(1):6, 2005.

[52] S X Ju, M J Black, S Minneman, and D Kimber. Summarization of videotaped presentations: automatic analysis of motion and gesture. *CSVT*, 8(5):686–696, 1998.

[53] Jae-Yoon Jung, Janice I Glasgow, and Stephen H Scott. Feature selection and classification for assessment of chronic stroke impairment. In *BIBE*, pages 1–5, 2008.

[54] A Kale, N Cuntoor, B Yegnanarayana, AN Rajagopalan, and R Chellappa. Gait analysis for human identification. In *AVBPA*, pages 706–714. Springer, 2003.

[55] A Kanaujia, C Sminchisescu, and D Metaxas. Semi-supervised hierarchical models for 3d human pose reconstruction. In *CVPR*, pages 1–8. IEEE, 2007.

[56] M Kapadia, I-k Chiang, T Thomas, N I Badler, J T Kider Jr, et al. Efficient motion retrieval in large motion databases. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pages 19–28. ACM, 2013.

[57] M Karg, G Venture, J Hoey, and D Kulic. Human movement analysis as a measure for fatigue: A hidden markov-based approach. 2014.

[58] J J Kavanagh and H B Menz. Accelerometry: a technique for quantifying movement patterns during walking. *GAIT POSTURE*, 28(1):1–15, 2008.

[59] Y. Ke, R Sukthankar, and M Hebert. Efficient visual event detection using volumetric features. In *ICCV*, volume 1, pages 166–173. IEEE, 2005.

[60] N Kern, B Schiele, and A Schmidt. Multi-sensor activity context detection for wearable computing. In Emile Aarts, RenW. Collier, Evert Loenen, and Boris Ruyter, editors, *Ambient Intelligence*, volume 2875 of *Lecture Notes in Computer Science*, pages 220–232. Springer, 2003. ISBN 978-3-540-20418-3.

[61] C Kervrann, F Davoine, P Pérez, R Forchheimer, and C Labit. Generalized likelihood ratio-based face detection and extraction of mouth features. *PATTERN RECOGN LETT*, 18(9):899–912, 1997.

[62] R M Kiss. Verification of determining the spatial position of the lower extremity by ultrasound-based motion analyser. *Civil Engineering*, 51(1):39–43, 2007.

[63] A Kralj, R J Jaeger, and M Munih. Analysis of standing up and sitting down in humans: definitions and normative data presentation. *J BIOMECH*, 23(11):1123–1138, 1990.

[64] H I Krebs, M L Aisen, B T Volpe, and N Hogan. Quantization of continuous arm movements in humans with brain injury. *PNAS*, 96(8):4645–4649, 1999.

[65] W H Kruskal and W A Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[66] D Kulić, G Venture, and Y Nakamura. Detecting changes in motion characteristics during sports training. In *EMBC*, pages 4011–4014, 2009.

[67] D Kulić, C Ott, D Lee, J Ishikawa, and Y Nakamura. Incremental learning of full body motion primitives and their sequencing through human motion observation. *International Journal of Robotics Research*, 31:330 – 345, 2012.

[68] Aristidis L, Nikos V, and Jakob J. V. The global k-means clustering algorithm. *Pattern Recognition*, 36(2):451 – 461, 2003. Biometrics.

[69] MA Lafortune, PR Cavanagh, HJd Sommer, and A Kalenak. Three-dimensional kinematics of the human knee during walking. *J BIOMECH*, 25(4):347–357, 1992.

[70] T HW Lam, R ST Lee, and D Zhang. Human gait recognition by the fusion of motion and static spatio-temporal templates. *Pattern Recognition*, 40(9):2563–2573, 2007.

[71] H K Lee and J H Kim. An hmm-based threshold model approach for gesture recognition. *PAMI*, 21:961–973, 1999.

[72] M F Levin, J A Kleim, and S L Wolf. What do motor recovery and compensation mean in patients following stroke? *NEUROREHAB NEURAL RE*, 23(4):313–319, 2009.

[73] Q Li, M Young, V Naing, and JM Donelan. Walking speed and slope estimation using shank-mounted inertial measurement units. In *ICORR*, pages 839–844. IEEE, 2009.

[74] X Li, S Sam Ge, Y Pan, K-S Hong, Z Zhang, and X Hu. Feature extraction based on common spatial analysis for time domain parameters. In *URAI*, pages 377–382. IEEE, 2011.

[75] J F-S Lin and D Kulić. On-line segmentation of human motion for automated rehabilitation exercise analysis. *NSRE*, 22:168–180, 2013.

[76] J FS Lin and D Kulić. Human pose recovery using wireless inertial measurement units. *PHYSIOL MEAS*, 33(12):2099, 2012.

[77] J.F.S. Lin and D. Kulić. On-line segmentation of human motion for automated rehabilitation exercise analysis. *NSRE*, 22(99):168 – 180, 2013.

[78] J Little and J Boyd. Describing motion for recognition. In *ICCV*, pages 235–240. IEEE, 1995.

[79] W Lorenz, W Dietz, M Ennis, B Stinner, and A Doenicke. Histamine in anaesthesia and surgery: Causality analysis. In *Histamine and Histamine Antagonists*, pages 385–439. Springer, 1991.

[80] R A Malinzak, S M Colby, D T Kirkendall, B Yu, and W E Garrett. A comparison of knee joint motion patterns between men and women in selected athletic tasks. *CLIN BIOMECH*, 16(5):438–445, 2001.

[81] J Mantyjarvi, J Himberg, and T Seppanen. Recognizing human motion with multiple acceleration sensors. In *SMC*, volume 2, pages 747–752. IEEE, 2001.

[82] T McInerney and D Terzopoulos. Deformable models in medical image analysis: a survey. *To appear in Medical Image Analysis*, 1:2, 1996.

[83] D Mena, JM Mansour, and SR Simon. Analysis and synthesis of human swing leg motion during gait and its clinical applications. *J BIOMECH*, 14(12):823–832, 1981.

[84] A Nakazawa, S Nakaoka, K Ikeuchi, and K Yokoi. Imitating human dance motions through motion structure analysis. In *IROS*, pages 2539–2544, 2002.

[85] R C Nelson and R Polana. Qualitative recognition of motion using temporal texture. *CVGIP: Image understanding*, 56(1):78–89, 1992.

[86] M S Nixon and J N Carter. Automatic recognition by gait. *Proceedings of the IEEE*, 94(11):2013–2024, 2006.

[87] C. C. Norkin and D. J. White. *Measurement of Joint Motion: A Guide to Goniometry, 4th Edition*. F. A. Davis Company, 2009.

[88] Y Nubar and R Contini. A minimal principle in biomechanics. *B MATH BIOPHYS*, 23(4):377–391, 1961.

[89] W Park, D B Chaffin, and B J Martin. Toward memory-based human motion simulation: development and validation of a motion modification algorithm. *SMC*, 34(3): 376–386, 2004.

[90] M Parvis and F Ferraris. Procedure for effortless in-field calibration of three-axial rate gyro and accelerometers. *Sensors and Materials*, 7(5):311–330, 1995.

[91] V Pavlovic, J M Rehg, T J Cham, and K P Murphy. A dynamic bayesian network approach to figure tracking using learned dynamic models. In *ICCV*, volume 1, pages 94–101. IEEE, 1999.

[92] J Perš, M Bon, S Kovačič, Marko Šibila, and Branko Dežman. Observation and analysis of large-scale human motion. *Human Movement Science*, 21(2):295–311, 2002.

[93] P J Phillips, S Sarkar, I Robledo, P Grother, and K Bowyer. The gait identification challenge problem: data sets and baseline algorithm. In *ICPR*, pages 385–388, 2002.

[94] R Polana and R Nelson. Low level recognition of human motion (or how to get your man without finding his body parts). In *Motion of Non-Rigid and Articulated Objects, IEEE Workshop on*, pages 77–82. IEEE, 1994.

[95] R Poppe. Vision-based human motion analysis: An overview. *Comput. Vis. Image Underst.*, 108(1-2):4–18, 2007. ISSN 1077-3142.

[96] R Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010.

[97] D. Kulic R. Houmanfar, M. Karg. Movement analysis of rehabilitation exercises: Distance metrics for measuring patient progress. *ISJEB2*, page 12, 2014.

[98] L R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[99] HK Ramakrishnan and MP Kadaba. On the estimation of joint kinematics during gait. *J BIOMECH*, 24(10):969–977, 1991.

[100] L Ren, A Patrick, A A Efros, J K Hodgins, and James M Rehg. A data-driven approach to quantifying natural human motion. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 1090–1097. ACM, 2005.

[101] J G Richards. The measurement of human motion: A comparison of commercially available systems. *Human Movement Science*, 18(5):589–602, 1999.

[102] K Rohr. Towards model-based recognition of human movements in image sequences. *CVGIP-Image Understanding*, 59(1):94–115, 1994.

[103] B Rohrer, S Fasoli, H I Krebs, R Hughes, B Volpe, W R Frontera, J Stein, and N Hogan. Movement smoothness changes during stroke recovery. *J Neurosci*, 22(18): 8297–8304, 2002.

[104] S Sarkar, P J Phillips, Z Liu, I R Vega, P Grother, and K W Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *PAMI*, 27(2):162–177, 2005.

[105] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local svm approach. In *ICPR*, pages 32–36, 2004.

[106] D B Segala, J B Dingwell, and D Chelidze. Slow-time changes in human emg muscle fatigue states are fully represented in movement kinematics. *J BIOMECH ENG*, 131: 021004–1, 2009.

[107] S K. Semwal and J J. Hallauer. Biomechanical modeling: Implementing line-of-action algorithm for human muscles and bones using generalized cylinders. *Computers & Graphics*, 18(1):105 – 112, 1994.

[108] M L Sichitiu and C Veerarittiphan. Simple, accurate time synchronization for wireless sensor networks. In *WCNC*, volume 2, pages 1266–1273. IEEE, 2003.

[109] L Sigal and M J Black. Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. *Brown Univertsity TR*, 120, 2006.

[110] J R Smith and S-F Chang. Transform features for texture classification and discrimination in large image databases. In *ICIP*, volume 3, pages 407–411. IEEE, 1994.

[111] MV Srinivasan. Generalized gradient schemes for the measurement of two-dimensional image motion. *Biological Cybernetics*, 63(6):421–431, 1990.

[112] H Su and F-G Huang. Human gait recognition based on motion analysis. In *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, volume 7, pages 4464–4468. IEEE, 2005.

[113] B Sun, X Liu, J Shen, and Q Zhang. Joint angle measurements based on omni-directional lower limb rehabilitation platform. In *MHS*, pages 337–341, 2012.

[114] J AK Suykens and J Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[115] A Świtoński, M Stawarz, M Boczarska-Jedynak, A Sieroń, A Polański, and K Wojciechowski. The effectiveness of applied treatment in parkinson disease based on feature selection of motion activities. *PRZ ELEKTROTECHNICZN*, 88:103–106, 2012.

[116] B Taskar, P Abbeel, and D Koller. Discriminative probabilistic models for relational data. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 485–492. Morgan Kaufmann Publishers Inc., 2002.

[117] P E Taylor, G JM Almeida, T Kanade, and J K Hodgins. Classifying human motion quality for knee osteoarthritis using accelerometers. In *EMBC*, pages 339–343, 2010.

[118] R Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, pages 267–288, 1996.

[119] P Tormene, T Giorgino, S Quaglini, and M Stefanelli. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artificial Intelligence in Medicine*, 45(1):11–34, 2009.

[120] B Toro, C J Nester, and P C Farren. Cluster analysis for the extraction of sagittal gait patterns in children with cerebral palsy. *GAIT POSTURE*, 25(2):157–165, 2007.

[121] N F Troje, C Westhoff, and M Lavrov. Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics*, 67(4):667–675, 2005.

[122] A Tsanas, M A Little, P E McSharry, and L O Ramig. Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests. *BME*, 57(4):884–893, 2010.

[123] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *CSVT*, 18(11):1473–1488, 2008.

[124] Md Uddin, JJ Lee, T-S Kim, et al. Independent component feature-based human activity recognition via linear discriminant analysis and hidden markov model. In *EMBS*, pages 5168–5171. IEEE, 2008.

[125] L Van Gestel, T De Laet, E Di Lello, H Bruyninckx, Guy M, A Van Campenhout, E Aertbeliën, M Schwartz, H Wambacq, P De Cock, et al. Probabilistic gait classification in children with cerebral palsy: A bayesian approach. *RES DEV DISABIL*, 32(6):2542–2552, 2011.

[126] A M van Mourik, A Daffertshofer, and Peter J Beek. Deterministic and stochastic features of rhythmic human movement. *Biological cybernetics*, 94(3):233–244, 2006.

[127] C Vogler and D Metaxas. Asl recognition based on a coupling between hmms and 3d motion analysis. In *ICCV*, pages 363–369. IEEE, 1998.

[128] A Volmer, N T Krüger, and R Orglmeister. Posture and motion detection using acceleration data for context aware sensing in personal healthcare systems. In *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pages 71–74. Springer, 2009.

[129] L Wang, W Hu, and T Tan. Recent developments in human motion analysis. *Pattern recognition*, 36(3):585–601, 2003.

[130] L Wang, T Tan, W Hu, and H Ning. Automatic gait recognition based on statistical shape analysis. *IP*, 12(9):1120–1131, 2003.

[131] L Wang, T Tan, H Ning, and W Hu. Silhouette analysis-based gait recognition for human identification. *PAMI*, 25(12):1505–1518, 2003.

[132] A Webber, N Virji-Babul, R Edwards, and M Lesperance. Stiffness and postural stability in adults with down syndrome. *EXP BRAIN RES*, 155(4):450–458, 2004.

[133] D Weinland, E Boyer, and R Ronfard. Action recognition from arbitrary views using 3d exemplars. In *ICCV*, pages 1–7. IEEE, 2007.

[134] S Wolf, T Loose, M Schablowski, R Döderlein, Land Rupp, H J Gerner, G Bretthauer, and R Mikut. Automated feature assessment in instrumented gait analysis. *GAIT POSTURE*, 23(3):331–338, 2006.

[135] H J Woltring. On optimal smoothing and derivative estimation from noisy displacement data in biomechanics. *Human Movement Science*, 4(3):229–245, 1985.

[136] J Wu, J Wang, and L Liu. Feature extraction via kpca for classification of gait patterns. *Human movement science*, 26(3):393–411, 2007.

[137] Y Wu and T S. Huang. Human hand modeling, analysis and animation in the context of hci. In *IEEE Signal Processing Magazine, Special issue on Immersive Interactive Technology*, pages 6–10, 1999.

[138] Y Wu and T S Huang. Vision-based gesture recognition: A review. *Urbana*, 51: 61801, 1999.

[139] D Xu, S Yan, D Tao, S Lin, and H J Zhang. Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval. *IP*, 16(11):2811–2821, 2007.

[140] R Xu, D Wunsch, et al. Survey of clustering algorithms. *Neural Networks, IEEE Transactions on*, 16(3):645–678, 2005.

[141] T Yabe and K Tanaka. Similarity retrieval of human motion as multi-stream time series data. In *DANTE*, pages 279–286. IEEE, 1999.

[142] L. Yardley, N. Beyer, K. Hauer, G. Kempen, C. Piot-Ziegler, and C. Todd. Development and initial validation of the falls efficacy scale-international. *Age and Ageing*, 34(6):614–19, 2005.

[143] J-H Yoo, M S Nixon, and C J Harris. Extracting human gait signatures by body segment properties. In *Image Analysis and Interpretation, 2002. Proceedings. Fifth IEEE Southwest Symposium on*, pages 35–39. IEEE, 2002.

[144] B Zhang, T Kanno, W Chen, G Wu, and D Wei. Walking stability by age a feature analysis based on a fourteen-linkage model. In *CIT*, pages 145–150, 2007.

[145] R Zhang, C Vogler, and D Metaxas. Human gait recognition. In *CVPRW*, pages 18–18. IEEE, 2004.

[146] Z Zhang, Q Fang, L Wang, and P Barrett. Template matching based motion classification for unsupervised post-stroke rehabilitation. In *ISBB*, pages 199–202, 2011.

[147] H Zhou and H Hu. Human motion tracking for rehabilitationa survey. *BIOMED SIGNAL PROCES*, 3(1):1–18, 2008.

[148] G Zhu, Q Huang, C Xu, L Xing, W Gao, and H Yao. Human behavior analysis for highlight ranking in broadcast racket sports video. *MM*, 9:1167–1182, 2007.