

A Deep Learning Framework for Assessing Physical Rehabilitation Exercises

Yalin Liao, Aleksandar Vakanski, *Member, IEEE*, and Min Xian¹

Abstract—Computer-aided assessment of physical rehabilitation entails evaluation of patient performance in completing prescribed rehabilitation exercises, by processing movement data captured with a sensory system. Despite the essential role of rehabilitation assessment toward improved rehabilitation outcomes and reduced healthcare costs, existing approaches for computer-aided monitoring and evaluation of patient performance lack versatility, robustness, and practical relevance. In this paper, we propose a deep learning-based framework for automated assessment of the quality of physical rehabilitation exercises. The main components of the framework are metrics for quantifying movement performance, scoring functions for mapping the performance metrics into numerical scores of movement quality, and deep neural network models for regressing quality scores of input movements via supervised learning. A performance metric based on the log-likelihood of a Gaussian mixture model used for encoding low-dimensional data representation obtained with a deep autoencoder network, is proposed in the paper. Multiple deep network architectures are repurposed for the task in hand and are validated by using a dataset of rehabilitation exercises. To the best of our knowledge, this is the first work that implements deep neural networks for the assessment of rehabilitation performance.

Index Terms—movement modeling, deep learning, performance metrics, physical rehabilitation

I. INTRODUCTION

PARTICIPATION in physical therapy and rehabilitation programs is necessary and critical in postoperative recovery by patients or for treatment of a wide array of musculoskeletal conditions. However, it is infeasible and economically unjustified to offer patients' access to a clinician for every single rehabilitation session [1]. Accordingly, current healthcare systems around the world are organized such that an initial portion of rehabilitation programs is performed in an inpatient facility under direct supervision by a clinician, followed by a second portion performed in an outpatient setting, where patients perform a set of prescribed exercises in their own residence. Reports in the literature indicate that more than 90% of all therapy sessions are performed in a home-based setting [2]. Under these circumstances, patients are asked to record their daily progress and to visit the clinic periodically for progress assessment. Still, numerous medical sources report low levels of patient motivation and adherence

to the recommended exercise regimens in home-based rehabilitation, leading to prolonged treatment times and increased healthcare cost [3], [4]. Although many different factors have been identified that contribute to the low compliance rates, the major impact factor is the absence of continuous feedback and timely oversight of patient exercises in a home environment by a healthcare professional [5].

Despite the development of a variety of new tools and devices in support of physical rehabilitation, such as robotic assistive systems [6], virtual reality and gaming interfaces [7], and Kinect-based assistants [2], [8], there is still a lack of versatile and robust systems for automatic monitoring and assessment of patient performance.

The article proposes a novel rehabilitation framework that encompasses formulation of metrics for quantifying movement performance, scoring functions for mapping the performance metrics into numerical scores of movement quality, and deep learning-based end-to-end models for encoding the relationship between movement data and quality scores.

The studied performance metrics are classified into model-less and model-based groups of metrics [9]. The model-less metrics employ distance functions, such as Euclidean and Mahalanobis distance, or dynamic time warping (DTW) [10] deviation between data sequences. The model-based metrics apply probabilistic approaches for modeling the movement data, and consequently, employ the log-likelihood for performance evaluation [11]. Next, the article investigates the effectiveness of deep autoencoder networks for dimensionality reduction of captured data. Further, we propose scoring functions for scaling the values of the studied performance metrics into the $[0, 1]$ range. The resulting movement quality scores are employed as the ground truth for training the proposed deep neural networks (NNs) for rehabilitation applications.

The proposed framework compares the performance of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hierarchical neural networks (HNNs) [12]. The framework is validated on the University of Idaho – Physical Rehabilitation Movement Dataset (UI-PRMD) [13]. To the best of our knowledge, this is the first framework that employs deep NNs for assessment of rehabilitation exercises.

The main contributions of the paper are: (1) A novel framework for assessment of rehabilitation exercises using deep NNs; (2) Comparison of NN architectures for movement assessment, and use of autoencoder NNs for dimensionality reduction of rehabilitation data; and (3) Comparison of performance metrics, and formulation of scoring functions for mapping the performance metrics into movement quality scores.

¹Manuscript submitted January 22, 2019. This work was supported by the Center for Modeling Complex Interactions (CMCI) at the University of Idaho Interactions through NIH Award #P20GM104420.

Yalin Liao, Aleksandar Vakanski, and Min Xian are with the Department of Computer Science, University of Idaho, 1776 Science Center Drive, Idaho Falls, ID, 83402, USA (e-mail: liao4728@vandals.uidaho.edu; vakanski@uidaho.edu; mxian@uidaho.edu).

The article is organized as follows. The next section provides an overview of related work. Section III describes the problem and introduces the mathematical notation. The ensuing section surveys common performance metrics for movement assessment. The proposed framework for rehabilitation assessment is presented in Section V, covering performance metric, dimensionality reduction, scoring functions, and deep NN architectures. The validation of the proposed framework on a dataset of exercises is reported in Section VI. The last two sections discuss the results and summarize the paper.

II. RELATED WORK

A. Human Movement Modeling

Conventional approaches for mathematical modeling and representation of human movements are broadly classified into two categories: top-down approaches that introduce latent states for describing the temporal dynamics of the movements, and bottom-up approaches that employ local features for representing the movements. Commonly used methods in the first category include Kalman filters [14], hidden Markov models [15], and Gaussian mixture models [16]. The main shortcomings of these methods originate from employing linear models for the transitions among the latent states (as in Kalman filters), or from adopting simple internal structures of the latent states (typical for hidden Markov models). The approaches based on extracting local features employ predefined criteria for identifying key points [17] or a collection of statistics of the movements (e.g., mean, standard deviation, mode, median) [18]. Such local features are typically motion-specific, which limits the ability to efficiently handle arbitrary spatio-temporal variations within movement data.

Recent developments in artificial NNs stirred significant interest in their application for modeling and analysis of human motions. Numerous works employed NNs for motion classification and applied the trained models for activity recognition, gait identification, gesture recognition, action localization, and related applications. NN-based motion classifiers utilizing different computational units have been proposed, including convolutional units [19], [20], long short-term memory (LSTM) recurrent units [21], [22], gated recurrent units [23], and combinations [24] or modifications of these computational units [25]. Also, NNs with different layer structures have been implemented, such as encoder-decoder networks [22], spatio-temporal graphs [26], and attention mechanism models [27], [28]. Besides the task of classification, a body of work in the literature focused on modeling and representation of human movements for prediction of future motion patterns [29], synthesis of movement sequences [22], and density estimation [11]. Conversely, little research has been conducted on the application of NNs for evaluation of movement quality, which can otherwise find use in various applications (physical rehabilitation being one of them).

B. Movement Assessment

Quantifying the level of correctness in completing prescribed exercises is important for the development of tools and devices in support of home- and clinic-based rehabilitation.

The movement assessment in existing studies is typically accomplished by comparing a patients performance of an exercise to the desired performance by healthy participants.

Several studies in the literature on exercise evaluation employed machine learning methods to classify the individual repetitions into *correct or incorrect classes* of movements. Methods used for this purpose include Adaboost classifier [30], k -nearest neighbors [31], Bayesian classifier [32], and an ensemble of multi-layer perceptron NNs [33]. The outputs in these approaches are discrete class values of 0 or 1 (i.e., incorrect or correct repetition). However, these methods do not provide the capacity to detect varying levels of movement quality or identify incremental changes in patient performance over the program duration.

The majority of related studies employed *distance functions* for deriving movement quality scores. Concretely, Houmanfar et al. [18] used a variant of the Mahalanobis distance to quantify the level of correctness of rehabilitation movements, based on a calculated distance between patient-performed repetitions and a set of repetitions performed by a group of healthy individuals. Similarly, a body of work utilized the dynamic time warping (DTW) algorithm [10] for calculating the distance between a patients performance and healthy subjects performance [34]–[36]. The advantage of the distance functions is that they are not exercise-specific, and thus can be applied for assessment of new types of exercises. However, the distance functions also have shortcomings, because they do not attempt to derive a model of the rehabilitation data, and the distances are calculated at the level of individual time-steps in the raw measurements.

Another body of research work utilized *probabilistic approaches* for modeling and evaluation of rehabilitation movements. Studies based on hidden Markov models [37], [38] and mixtures of Gaussian distributions [11] typically perform a quality assessment based on the likelihood that the individual sequences are being drawn from a trained model. Whereas the probabilistic models are advantageous in handling the variability due to the stochastic character of human movements, models with abilities for a hierarchical data representation can produce more reliable outcomes for movement quality assessment, and better generalize to new exercises.

III. PRELIMINARIES

A. Problem Description

In a physical rehabilitation setting, a clinician prescribes a collection of rehabilitation exercises to a patient, by either performing the movements in front of the patient or physically moving the patients body parts along the required paths. It is assumed that the clinician provides several demonstrations of the exercises in order to reinforce the perception of the required movements by the patient, in relation to the range and speed of movements, or other postural constraints in the performance of the movements. Then the patient is tasked to perform the prescribed set of exercises at home, according to the instructed rehabilitation regimen. It is assumed here that a daily rehabilitation *session* requires completing a series of

exercises, where the patient is instructed to complete a certain number of *repetitions* of each exercise during each session.

This article applies machine learning algorithms for mathematical modeling and assessment of rehabilitation movements captured in the form of skeletal data, consisting of time-ordered sequences of position or angular displacements of the joints in the human body. The main task is to evaluate the level of correctness in performing the exercises. Toward this goal, we employ performance metrics to quantify the level of performance, and define scoring functions for assigning quality scores to the repetitions of an exercise. NNs are trained in a supervised manner by using the obtained quality scores as labels for the individual repetitions of an exercise.

B. Notation

We assume that a sensory system is available to capture the skeletal data during the performance of rehabilitation exercises. It is further assumed that a control group of healthy subjects will perform multiple repetitions of each rehabilitation exercise. The healthy subjects will provide reference movement performance completed in a correct manner. The acquired data by the sensory system for the movements for one particular exercise performed by S subjects is denoted by \mathbb{X} , and hereafter they are referred to as *reference movements*. The symbol R_s is used for the number or repetitions of the exercise by the s -th subject. The combined data for all R_s repetitions of the exercise by the s -th subject is denoted Δ_s . Similar, R is used for the total number of all repetitions by the S subjects, i.e., $R = \sum_{s=1}^S R_s$. Using the notation $\mathbf{X}_{s,r}$ for the collected data of the r -th repetition by the s -th subject, we have $\mathbb{X} = \{\Delta_s\}$ for $s \in \underline{S}$, where $\Delta_s = \{\mathbf{X}_{s,r}\}$ for $r \in \underline{R_s}$. For convenience, throughout the text the underscore symbol denotes a set of indices, e.g., $\underline{S} = \{1, 2, \dots, S\}$ for any positive integer S . The data for each repetition $\mathbf{X}_{s,r}$ is a temporal sequence of T measurements, therefore $\mathbf{X}_{s,r} = (\mathbf{x}_{s,r}^{(1)}, \mathbf{x}_{s,r}^{(2)}, \dots, \mathbf{x}_{s,r}^{(T)})$, where the superscripts are used for indexing the temporal order of the displacement vectors within the repetition. Furthermore, the individual measurement $\mathbf{x}_{s,r}^{(t)}$ for $t \in \underline{T}$ is a D -dimensional vector, consisting of the values for all joint displacements in the human body, i.e. $\mathbf{x}_{s,r}^{(t)} = [x_{s,r}^{(t,1)}, x_{s,r}^{(t,2)}, \dots, x_{s,r}^{(t,D)}]$.

Similarly, it is assumed that movement data will be collected from a group of patients, who may not be able to perform the exercises in a correct manner due to a musculoskeletal condition. The collected data for the patients group are referred in the article as *patient movements*, and are denoted with the symbol \mathbb{Y} . By analogy to the introduced notation for the reference movements, $\mathbb{Y} = \{\mathbf{Y}_{s,r}\}$, where $\mathbf{Y}_{s,r}$ is the data of the r -th repetition by the s -th subject. Analogously, the repetition $\mathbf{Y}_{s,r} = (\mathbf{y}_{s,r}^{(1)}, \mathbf{y}_{s,r}^{(2)}, \dots, \mathbf{y}_{s,r}^{(T)})$ is comprised of a sequence of multidimensional vectors $\mathbf{y}_{s,r}^{(t)} = [y_{s,r}^{(t,1)}, y_{s,r}^{(t,2)}, \dots, y_{s,r}^{(t,D)}]$.

IV. SURVEY OF PERFORMANCE METRICS FOR MOVEMENT ASSESSMENT

Performance metrics evaluate the level of correctness of each repetition with respect to the set of reference movements.

This section surveys metrics that have been commonly used in prior studies on rehabilitation assessment [18], [34], [36], [37]. The metrics are classified into two categories: model-less and model-based [9]. The *model-less metrics* are calculated directly from low-level measurements of trajectories of body joints as acquired by the sensory system, without modeling the movement data. The *model-based metrics* evaluate the repetitions data with respect to a model of the exercise, e.g., using probabilistic modeling approaches and employing the log-likelihood for performance evaluation. In addition, considering that existing datasets of movements contain data collected by multiple subjects, the performance metrics can be calculated based on *between-subject within-subject* case.

A. Model-less Performance Metrics

Euclidean distance: the Euclidean distance between two movement data $\mathbf{Y}_{s,r}$ and $\mathbf{X}_{s',r'}$ has been commonly used for movement assessment, and it is defined by

$$\begin{aligned} d_E(\mathbf{Y}_{s,r}, \mathbf{X}_{s',r'}) &= \sum_{t=1}^T \left\| \mathbf{y}_{s,r}^{(t)} - \mathbf{x}_{s',r'}^{(t)} \right\| \\ &= \sum_{t=1}^T \sqrt{\sum_{d=1}^D (y_{s,r}^{t,d} - x_{s',r'}^{t,d})^2} \quad (1) \end{aligned}$$

for $s, s' \in \underline{S}$ and $r, r' \in \underline{R_s}$.

Mahalanobis distance: The Mahalanobis distance between two sequences $\mathbf{Y}_{s,r}$ and $\mathbf{X}_{s',r'}$ is

$$\begin{aligned} d_M(\mathbf{Y}_{s,r}, \mathbf{X}_{s',r'}) &= \sum_{t=1}^T \sqrt{(\mathbf{y}_{s,r}^{(t)} - \mathbf{x}_{s',r'}^{(t)})^T \mathbf{V}^{-1} (\mathbf{y}_{s,r}^{(t)} - \mathbf{x}_{s',r'}^{(t)})} \quad (2) \end{aligned}$$

where \mathbf{V} is the covariance matrix of the set \mathbb{X} . In fact, the Euclidean distance represents a special case of the Mahalanobis distance when the covariance matrix is identity matrix.

DTW Distance: Dynamic time warping (DTW) algorithm is an algorithm for aligning time-series data via nonlinear warping of the temporal order of the data points in order to reduce a distance function between the time-series. The most commonly used distance function in DTW is the Euclidean distance. The optimal alignment path in DTW is calculated by minimizing the sum of the cumulative distances between the two time-series (e.g., $d_E(\mathbf{Y}_{s,r}, \mathbf{X}_{s',r'})$) and the minimum distances between the neighboring data points (here denoted $g(\mathbf{Y}_{s,r}, \mathbf{X}_{s',r'})$), i.e.,

$$\begin{aligned} d_{DTW}(\mathbf{Y}_{s,r}, \mathbf{X}_{s',r'}) &= d_E(\mathbf{Y}_{s,r}, \mathbf{X}_{s',r'}) \\ &\quad + g(\mathbf{Y}_{s,r}, \mathbf{X}_{s',r'}). \quad (3) \end{aligned}$$

For a more detailed description of DTW, refer to [10].

B. Model-based Performance Metrics

GMM Log-likelihood: Gaussian Mixture Model (GMM) is a parametric probabilistic model for representing data with a mixture of Gaussian probability density functions [39] GMM is frequently used for modeling human movements. For a

dataset consisting of multidimensional vectors $\mathbf{y}_{s,r}^{(t)}$, a GMM with C Gaussian components has the form

$$\mathcal{P}(\mathbf{y}_{s,r}^{(t)} | \lambda) = \sum_{c=1}^C \pi_c \mathcal{N}(\mathbf{y}_{s,r}^{(t)} | \mu_c, \Sigma_c), \quad (4)$$

where $\lambda = \{\pi_c, \mu_c, \Sigma_c\}$ are the mixing coefficient, mean, and covariance of the c -th Gaussian component, respectively. Subsequently, for a GMM model with parameters λ , the log-likelihood for the repetition $\mathbf{Y}_{s,r}$ is used as a performance metrics, which otherwise is calculated as

$$\mathcal{P}(\mathbf{Y}_{s,r} | \lambda) = - \sum_{t=1}^T \log \left\{ \sum_{c=1}^C \pi_c \mathcal{N}(\mathbf{y}_{s,r}^{(t)} | \mu_c, \Sigma_c) \right\}. \quad (5)$$

C. Between- and Within-Subject Metrics

Between-subject metrics calculate the deviation between a repetition $\mathbf{Y}_{s,r}$ of an exercise and the set of reference repetitions of the exercise performed by a group of healthy subjects \mathbb{X} . Hence, the deviation for the model-less metrics is obtained as

$$D_{BS}(\mathbf{Y}_{s,r}, \mathbb{X}) = \frac{1}{TR} \sum_{s'=1}^S \sum_{r'=1}^{R_{s'}} d(\mathbf{Y}_{s,r}, \mathbf{X}_{s',r'}), \quad (6)$$

where d is one of the metrics formulated in (1) – (3). For the model-based metric in (5), the deviation for the repetition $\mathbf{Y}_{s,r}$ and a GMM model of \mathbb{X} with parameters λ is

$$D_{BS}(\mathbf{Y}_{s,r}, \mathbb{X}) = \frac{1}{TR} \mathcal{P}(\mathbf{Y}_{s,r} | \lambda). \quad (7)$$

Within-subject metrics quantify the deviation between a repetition of an exercise $\mathbf{Y}_{s,r}$ and a set of repetitions of the exercise performed by the same subject Δ_s . The deviation between $\mathbf{Y}_{s,r}$ and Δ_s is given with

$$D_{WS}(\mathbf{Y}_{s,r}, \Delta_s) = \frac{1}{TR_s} \sum_{r'=1}^{R_s} d(\mathbf{Y}_{s,r}, \mathbf{X}_{s,r'}), \quad (8)$$

for $s \in \underline{S}$ and $r \in R_s$ for the model-less metrics. The deviation for the model-based metric is also obtained from (7), where a separate GMM model λ_s is used for encoding the repetitions data for each subject Δ_s , i.e., $D_{WS}(\mathbf{Y}_{s,r}, \Delta_s) = \frac{1}{TR_s} \mathcal{P}(\mathbf{Y}_{s,r} | \lambda_s)$.

V. PROPOSED METHOD

The proposed framework for assessing rehabilitation exercises encompasses dimensionality reduction, performance metrics, scoring functions, and NN models. A block-diagram of the envisioned framework is depicted in Fig. 1. The measured joint coordinates data by the sensory system are processed via dimensionality reduction, performance metric, and scoring function to obtain movement quality scores that are used for training a NN model. The trained NN model is afterward used to automatically generate movement quality scores for input movement data acquired by the sensory system.

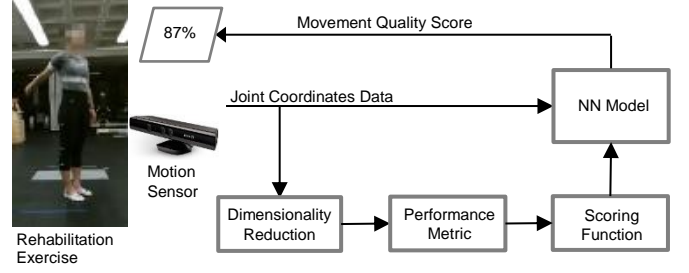


Fig. 1: Overview of the proposed framework for assessment of rehabilitation exercises.

A. Dimensionality Reduction

Outputs from the sensory systems for capturing human joint displacements are high-dimensional data, typically ranging between 40 and 120 dimensions. Dimensionality reduction of recorded data is often considered an essential step in processing human movements, in order to suppress unimportant, redundant, or highly correlated dimensions. The aim is to project the data $\mathbb{X} = \{\mathbf{X}_{s,r} = (\mathbf{x}_{s,r}^{(t)}) : \mathbf{x}_{s,r}^{(t)} \in \mathbb{R}^D\}$ into a lower-dimensional representation $\tilde{\mathbb{X}} = \{\tilde{\mathbf{X}}_{s,r} = (\tilde{\mathbf{x}}_{s,r}^{(t)}) : \tilde{\mathbf{x}}_{s,r}^{(t)} \in \mathbb{R}^M\}$, for $t \in \underline{T}$, $s \in \underline{S}$, $r \in R_s$, and where $M < D$. A common approach for dimensionality reduction of human movement data is maximum variance [40], which simply retains the first M dimensions with the largest variance and discards the remaining dimensions. PCA and its variants [41] are also widely used for reducing the dimensionality of movement data, where a matrix containing the leading M eigenvectors corresponding to the largest eigenvalues of the covariance matrix \mathbf{V} is used for projecting the data into a lower-dimensional space. Although PCA is one of the most common approaches for dimensionality reduction in general, it employs linear mapping of high-dimensional data into a lower-dimensional representation. Likewise, the shortcomings of maximum variance originate from its simplicity.

In the proposed framework, we introduce **autoencoder NNs** [42] for dimensionality reduction. Autoencoder NNs is a non-linear technique for dimensionality reduction, which allows extracting richer data representations for dimensionality reduction in comparison to the linear techniques (such as PCA). Furthermore, deep autoencoder NNs created by stacking multiple consecutive layers of hidden neurons, can additionally increase the representational capacity of the network.

Autoencoders are an unsupervised form of NNs designed to learn an alternative representation of input data, through a process of data compression and reconstruction. The data processing involves an encoding step of compressing input data through one or multiple hidden layers, followed by a decoding step of reconstructing the output from the encoded representation through one or multiple hidden layers. If \mathcal{A} denotes a class of mapping functions from \mathbb{R}^M to \mathbb{R}^D , and \mathcal{B} is a class of mapping functions from \mathbb{R}^D to \mathbb{R}^M , then for any function $A \in \mathcal{A}$ and $B \in \mathcal{B}$, the encoder portion projects an input $\mathbf{x}_{s,r}^{(t)} \in \mathbb{R}^D$ into a lower-dimensional representation $\tilde{\mathbf{x}}_{s,r}^{(t)} \in \mathbb{R}^M$ (referred to as a code), and the decoder portion

converts the code into an output $A(B(x_{s,r}^{(t)})) \in \mathbb{R}^D$. Autoencoders are trained to find functions $A \in \mathcal{A}$ and $B \in \mathcal{B}$ which minimize the mean squared deviation between the input data and output data, i.e.,

$$\arg \min_D \|A(B(x_{s,r}^{(t)})) - x_{s,r}^{(t)}\|. \quad (9)$$

A graphical representation of the adopted architecture for the autoencoder network is presented in Fig. 2. The encoder portion consists of three intermediate layers of LSTM recurrent units with 30, 10, and 4 computational units, and the corresponding decoder portion has three intermediate layers of LSTM units with 10, 30, and 117 computational units, respectively. The input time-series data are 117-dimensional vectors of joint coordinates. The code representation of the proposed network is a temporal sequence of 4-dimensional vectors.

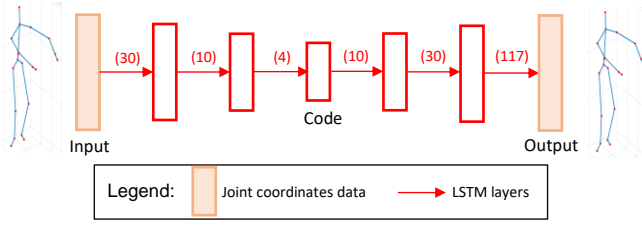


Fig. 2: The proposed autoencoder architecture projects an input movement data into a code representation, and re-projects the code into the movement data.

B. Performance Metric

We adopt the metric based on GMM log-likelihood given in (5) and (7). This choice stems from the demonstrated capacity of statistical methods to encode stochastic variability in human movements, which results in improved ability by the model-based metrics to handle spatio-temporal variations in rehabilitation data.

C. Scoring Functions

Scoring functions map the value of the performance metrics into a movement quality score in the range between 0 and 1. The resulting movement quality scores play a dual role in the proposed framework. First, in a real-world exercise assessment setting, the quality scores allow for intuitive understanding of the calculated values of the used performance metric. For instance, a movement quality score of 88% presented to a patient is easy to understand, and it can also enable the patient to self-monitor his/her progress toward functional recovery based on received quality scores over a period of time. Second, the quality scores are used here for supervised training of the studied NN models.

For a sequence of metric values of the reference movements $\mathbf{x} = (x_1, x_2, \dots, x_L)$ and a sequence $\mathbf{y} = (y_1, y_2, \dots, y_L)$ related to the patient movements, we propose the following scoring functions:

$$\begin{aligned} \bar{x}_k &= \left(1 + e^{\frac{x_k}{\mu+3\delta} - \alpha_1}\right)^{-1}, \\ \bar{y}_k &= \left(1 + e^{\frac{x_k}{\mu+3\delta} - \alpha_1} + \frac{y_k - x_k}{\alpha_1(\mu+3\delta)}\right)^{-1}; \end{aligned} \quad (10)$$

$$\begin{aligned} \bar{x}_k &= \left(1 + e^{\frac{\alpha_3 x_k}{M} - \alpha_4}\right)^{-1}, \\ \bar{y}_k &= \left(1 + e^{\frac{\alpha_3 y_k}{M} - \alpha_4}\right)^{-1}; \end{aligned} \quad (11)$$

$$\begin{aligned} \bar{x}_k &= 1 - |x_k - \mu|, \\ \bar{y}_k &= \begin{cases} 1 - |y_k - \mu| & \text{if } 1 - |y_k - \mu| > 0 \\ \alpha_5 & \text{if } 1 - |y_k - \mu| \leq 0 \end{cases}. \end{aligned} \quad (12)$$

In the above equations, $k \in \underline{L}$, $\mu = \frac{1}{L} \sum_{k=1}^L |x_k|$, $\delta = \sqrt{\frac{1}{L} \sum_{k=1}^L (|x_k| - \mu)^2}$, $M = \max_{i,j \in \underline{L}} \{x_i, y_j\}$, and α_1 to α_5 are data specific parameters. The proposed scoring functions in (10) and (11) are monotonically decreasing, and are designed to preserve the distribution of the values of performance metrics. The values for the reference movements x_k are scaled by $\mu+3\delta$ in (10) or by the maximum value M in (11) to ensure scores \bar{x}_k close to 1 for inputs x_k approximately in the range $(\mu-3\delta, \mu+3\delta)$ or no larger than M , respectively. Similarly, for the movements y_k the scoring functions in (10) and (11) are designed to preserve their distribution in mapping the metrics values to quality scores. The scoring function in (12) scales the metric values based on the absolute distance from the mean of the reference movements. Experimental evaluation of the scoring functions is provided in Section VI.

D. Neural Networks

Three different deep NN architectures are developed, implemented and evaluated in this work. These include CNNs, RNNs, and HNNs. For the networks we performed a grid search with various combinations of layers, numbers of layers, computational units per layer, size of convolutional filters, batch size, and other related hyperparameters. For all models, mean-squared-error was selected as a cost function, and Adam optimizer was employed. A batch size of 5 was applied, with early stopping regularization. Inputs are 117-dimensional sequences of joint displacements corresponding to single repetitions of an exercise. The output layer has linear activations, and outputs a numerical movement quality score for an input repetition.

The network architectures are presented in Table I, and are also illustrated in Fig. 3. The adopted CNNs contain three convolutional layers, two fully connected hidden layers, and an output layer. They utilize strided one-dimensional convolutional filters, leaky ReLU activations, and dropout of 0.2.

The RNN models with recurrent architecture consist of two bidirectional layers of LSTM units, one intermediate full connected layer, and an output layer. The recurrent layers use a recurrent dropout of 0.5, and are as well followed by a dropout layer with the rate of 0.25.

The HNNs [12] are based on a hierarchical model that employs five recurrent sub-networks that take as inputs joint displacements of the left arm, right arm, left leg, right leg, and torso, respectively. The outputs from the five sub-networks are

TABLE I: NETWORK ARCHITECTURES

Network	Layers ^a
CNNs	Conv1D (60, 5, LR, D:0.2, St:2) \times Conv1D (30, 3, LR, D:0.2, St:2) \times Conv1D (10, 3, LR, D:0.2) \times FC (200, LR, D:0.2) FC (100, LR, D:0.2) \times FC (1, L)
RNNs	BiLSTM (20, RD:0.5, D:0.25) \times FC (30, LR, TH, D:0.5) \times BiLSTM (10, RD:0.5, D:0.25) \times FC (1, L)
HNNs	BiRNN (10, TH, RD:0.5) * 5 \rightarrow \times BiRNN (20, TH, RD:0.5) * 4 \rightarrow \times BiRNN (20, TH, RD:0.5) * 2 \rightarrow \times BiLSTM (30, TH, RD:0.5) FC (1, L)

^aAcronyms: Conv1D(N_K, N_S, \dots) – Layer with one-dimensional convolutional units with N_K kernels of size N_K , FC(\dots) – Fully connected layer, BiLSTM(\dots) – Layer with bidirectional LSTM units, BiRNN – Layer with bidirectional simple recurrent units, LR – Leaky ReLU activation, D – Dropout, RD – Recurrent dropout, St – Stride, L – linear activation, TH – Tanh activation, \rightarrow Merged layers.

progressively merged through a series of layers into a unified representation. Such hierarchical organization of the layers allows low-level spatial information from joint coordinates to be exploited for obtaining a high-level representation of the body parts movements in accomplishing the movements. The network employs three bidirectional layers with simple recurrent units, one bidirectional layer with LSTM units, and an output layer. The structure of the network was selected via a grid search and is displayed in Table I.

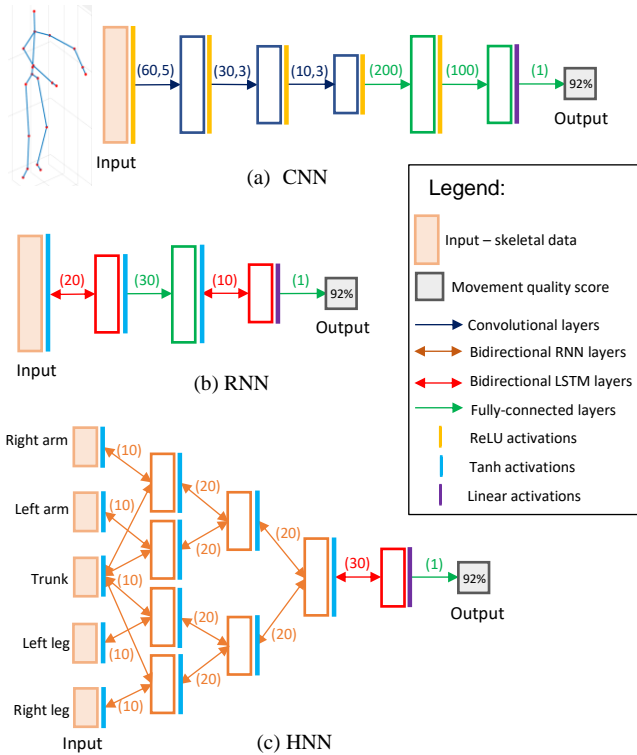


Fig. 3: NN architectures. (a) CNNs; (b) RNNs; (c) HNNs.

Data augmentation is crucial in image processing with NNs, where applying various transformations on input images leads to improved performance and model robustness. Consequently, we posit that data augmentation is important for processing movement data with NNs, particularly because existing datasets of human movement have relatively small size. For this purpose, data augmentation to the movement

TABLE II: EXERCISES IN THE UI-PRMD DATASET

Order	Exercise
E1	Deep squat
E2	Hurdle step
E3	Inline lunge
E4	Side lunge
E5	Sit to stand
E6	Standing active straight leg raise
E7	Standing shoulder abduction
E8	Standing shoulder extension
E9	Standing shoulder internalexternal rotation
E10	Standing shoulder scaption

sequences is performed by introducing additive noise

$$\hat{\mathbf{X}}_{s,r} = \mathbf{X}_{s,r} + k \cdot \gamma \quad (13)$$

for $s \in \underline{S}$ and $r \in \underline{R_s}$, where $\gamma \sim U(0, 1)$ represents a random number drawn from a uniform probability distribution U , and k is a constant parameter.

VI. EXPERIMENTAL RESULTS

A. Dataset

For validation of the presented framework, we created the UI-PRMD dataset [13]. The dataset comprises skeletal data collected from 10 healthy subjects. Each subject completed 10 repetitions of 10 rehabilitation exercises, listed in Table II. The data were acquired with a Vicon optical tracking system, and consist of 117-dimensional sequences of angular joint displacements. The subjects performed the exercises both in a correct manner, hereafter referred to as correct movements, and in an incorrect manner, i.e., simulating performance by patients with musculoskeletal constraints, hereafter referred to as incorrect movements. A detailed description of the UI-PRMD dataset is provided in [13].

B. Comparison of Performance Metrics

In this section, the performance metrics presented in Section IV—Euclidean, Mahalanobis, DTW distances, and GMM log-likelihood—are evaluated on the UI-PRMD dataset.

Data scaling: To compare the metrics on the same basis, their values are first linearly scaled to the same range. In this study the range $[1, 20]$ was selected based on an empirical understanding of the data. For obtained values of the performance metrics for repetitions of the correct movements denoted $\mathbf{x} = (x_1, x_2, \dots, x_L)$ and for the metrics of the incorrect movements $\mathbf{y} = (y_1, y_2, \dots, y_L)$ the following scaling functions were used

$$x'_i = \frac{19(y_i - m)}{M - m}; y'_i = \frac{19(y_i - m)}{M - m} \text{ for } i \in \underline{L} \quad (14)$$

where $x'_i, y'_i \in [1, 20]$ denote the scaled values of the performance metrics, $M = \max_{i,j \in \underline{L}} \{x_i, y_j\}$, and $m = \min_{i,j \in \underline{L}} \{x_i, y_j\}$.

The scaled values of the Euclidean distance for exercises E1 and E2 are shown in Fig. 4. Green circles markers are used for the repetitions of the correct movements, whereas the red squares symbolize the repetitions of the incorrect movements. Note that inconsistent data (associated with measurement errors or subjects performing the exercise with their left-arm/leg in a set of mostly right arm/leg exercises) were manually

removed from the original dataset, resulting in less than 100 repetitions per subject. E.g., there are 90 correct and incorrect movements for E1 in Fig. 4(a), and 55 correct and incorrect movements for E2 in Fig. 4(b).

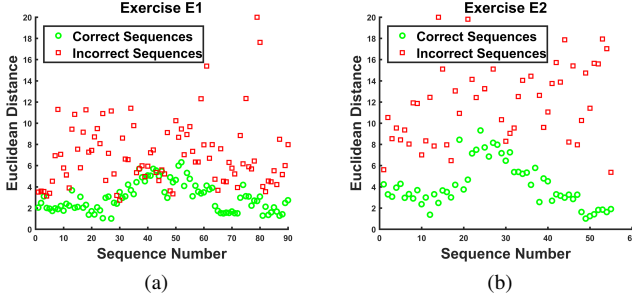


Fig. 4: Scaled values of the Euclidean distance for the between-subject case for: (a) first exercise E1 ($S_D = 0.384$); (b) second exercise E2 ($S_D = 0.497$).

Separation degree: For comparison of the scaled values of the performance metrics we propose the concept of separation degree. Specifically, for any positive real numbers x, y their separation degree is defined as $S_D(x, y) = \frac{x-y}{x+y} \in [-1, 1]$. The separation degree between two positive sequences $\mathbf{x} = (x_1, x_2, \dots, x_m)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is defined by

$$S_D(\mathbf{x}, \mathbf{y}) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n S_D(x_i, y_j). \quad (15)$$

Values of the separation degree close to 1 or -1 indicate good separation between the two sequences. Conversely, for values of the separation degree close to 0, the sequences don't separate well and they are almost mixed together.

When applied to the values of the distance metrics, the separation degree indicates greater ability of the used metric to differentiate between correct and incorrect repetitions of an exercise. For instance, in Fig. 4(b) one can observe a clearer differentiation between the correct and incorrect movements, in comparison to Fig. 4(a). This results in a larger value of the separation degree for the repetitions of exercise E2, which we calculated at 0.384 for E1, and 0.497 for E2, respectively.

The values for the separation degrees for the four studied performance metrics are presented in Table III. Each cell in the table corresponds to the average separate degree values S_D for the 10 exercises in the dataset. The shown values are the mean and in parenthesis is the standard deviation. For the comparison, scaled values of the metrics according to (14) are used. The table also compares the values of the metrics for the cases of raw 117-dimensional data, and low-dimensional data obtained with the presented methods of maximum-variance, PCA, and GMM log-likelihood. The largest values for the separation degree are indicated in each row with a bold font.

Conclusively, the GMM log-likelihood metric applied on a low-dimensional data with the autoencoder NN resulted in the largest separation between the correct and incorrect movements for both between- and within-subject cases. The within-subject case provides improved separation because the repetitions performed by the same subject are characterized

TABLE III: SEPARATION DEGREE FOR THE PERFORMANCE METRICS: MEAN (DEVIATION)

Metric	Euclidean distance	Mahalanobis distance	DTW distance	Log-likelihood GMM
Between-subject				
D ^a =117	0.445(0.087)	0.195(0.152)	0.487(0.063)	
D=3(MV)	0.309(0.101)	0.063(0.130)	0.310(0.100)	0.344(0.049)
D=3(PCA)	0.296(0.103)	0.108(0.169)	0.265(0.093)	0.360(0.060)
D=4(AE)	0.423(0.092)	0.229(0.102)	0.427(0.094)	0.515(0.106)
Within-subject				
D=117	0.568(0.058)	0.441(0.118)	0.570(0.059)	
D=3(MV)	0.472(0.048)	0.325(0.118)	0.455(0.053)	0.471(0.098)
D=3(PCA)	0.508(0.032)	0.322(0.169)	0.501(0.031)	0.518(0.057)
D=4(AE)	0.582(0.057)	0.474(0.133)	0.574(0.060)	0.603(0.073)

^aD: data dimensions; MV: maximum variance; PCA: principal component analysis; AE: autoencoder neural networks.

with a lower level of variability. The value of the GMM log-likelihood is not provided for the 117-dimensional data because GMM delivers better performance on low-dimensional data. Furthermore, the performance of the Euclidean and DTW distances in Table III is comparable, and better than the Mahalanobis distance. Also, the autoencoder NN lost less information in compressing the high-dimensional data sequences in comparison to maximum variance and PCA, because the separation degree values for all metrics using autoencoders are very close to the corresponding metric values of the 117-dimensions data without dimensionality reduction. In implementing GMM on the dataset, the number of Gaussian components C was set to 6.

C. Comparison of Neural Networks

Next, the performance of the presented NN architectures is evaluated. For training the networks, the movement quality scores based on the GMM log-likelihood calculated with autoencoder reduced data are employed. Only the case of between-subject is considered, because for the within-subject cases the number of repetitions per subject is too low for NNs training.

Scoring functions: The scoring function presented in (10) provided the best results for the used dataset, and therefore it is used to calculate the quality scores. The values of the parameters in (10) are empirically selected as $\alpha_1 = 3.2$ and $\alpha_2 = 10$. For example, Fig. 5 depicts the values of the log-likelihood and the corresponding performance scores for exercise E1 (i.e., deep squat). The scores for the correct movements are shown in Fig. 5(b) and have values close to 1, and the majority of the scores for the incorrect movements are in the range between 0.7 and 0.9.

NN movement assessment: Inputs to the NNs are pairs of repetitions data and quality scores. The networks are trained in a supervised regression manner, where the output is a predicted value of the movement quality for an input repetition (i.e., the quality scores can be considered equivalent to class labels in classification tasks). Also, note that inputs to the network are the raw measurement data with 117 dimensions.

Each network model is run five times, and we report the average absolute deviation between the input quality scores and the quality scores predicted by the network. The results for the 10 exercises in the UI-PRMD dataset are displayed in Table IV. Lower values of the deviation indicate low errors by

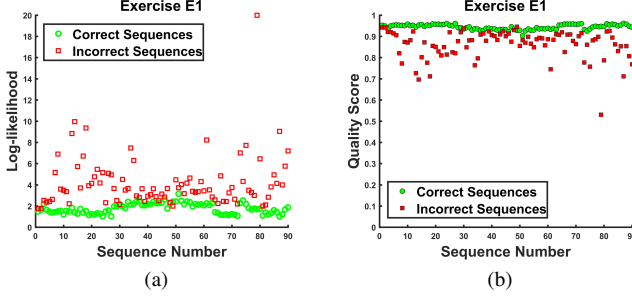


Fig. 5: (a) GMM log-likelihood values for exercise E1; (b) Corresponding quality scores.

TABLE IV: AVERAGE ABSOLUTE DEVIATION PER EXERCISE

Exercise	CNN	RNN	HNN
E1	0.01357	0.01670	0.03010
E2	0.02953	0.04934	0.07742
E3	0.04141	0.09382	0.13766
E4	0.01640	0.01609	0.03580
E5	0.01300	0.02536	0.06367
E6	0.02349	0.02166	0.04676
E7	0.03346	0.04090	0.19280
E8	0.02905	0.04590	0.07260
E9	0.02495	0.04419	0.06508
E10	0.03667	0.05198	0.16009

the network model in predicting the quality scores for input data. Accordingly, CNNs outperformed RNNs and HNNs on most of the exercise data. The results are further discussed in the subsequent section.

A performance example by the used CNN for exercise E1 is depicted in Fig. 6. The set of 90 correct and 90 incorrect repetitions was split 0.7/0.3 into a training set of 124 and a validation set of 56 repetitions. The input scores and predicted scores for the training and validation sets are shown in Figs. 6(a) and (b), respectively. In the two sub-figures the first half of the scores are for the correct sequences and have values close to one, and the second half of the scores pertain to the incorrect sequences and have lower quality scores. Overall, the network predictions are close to the assigned quality scores for almost all data instances.

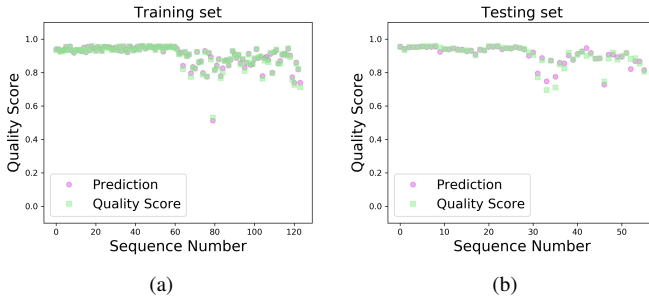


Fig. 6: (a) CNN predictions on the training set for exercise E1; (b) CNN predictions on the validation set for exercise E1.

Next, the effect of data augmentation on the employed NNs is investigated. Multiple values were adopted for the parameter k in (13), i.e., $k \in \{0.01, 0.03, 0.05, 0.07\}$, resulting in differing levels of added noise. By adding random noise to the 90 correct instances for exercise E1, 360 new instances were

TABLE V: AVERAGE ABSOLUTE DEVIATION FOR EXERCISE E1

Data type	CNN	RNN	HNN
Original data	0.01357	0.01670	0.03010
Augmented data	0.00656	0.00688	0.01404

synthetically created. After the data augmentation, movement quality scores were calculated for the generated sequences, and the NNs were trained on data including the original real data and synthetic data. The results are displayed in Table IV. The results indicate low errors in predicting the movement quality scores on the augmented dataset for all three network architectures.

VII. DISCUSSION

The article introduces a novel framework for the assessment of rehabilitation exercises via deep NNs. The framework includes performance metrics, scoring functions, and NN models.

Common metrics for quantifying the level of consistency in captured rehabilitation movements are surveyed and compared. The metrics include a model-less category that calculates distances between low-level measurement data points, and a model-less group that employs high-level latent states for estimating the data consistency. Studied metrics in the former group are Euclidean, Mahalanobis, and DTW distance, and in the latter group is GMM log-likelihood. The concept of separation degree is proposed for metric comparison. GMM log-likelihood outperformed the model-less metrics on the UI-PRMD dataset. Such results confirm our hypothesis that efficient movement assessment is strongly predicated on the provision of efficient models of human movements. Probabilistic approaches, such as the used GMM approach, have improved ability to handle the inherent variability and measurement uncertainty in human movement data, in comparison to the model-less approaches.

We compared the performance of PCA and maximum variance approaches for dimensionality reduction of human movements to a proposed approach that employs autoencoder NNs. Expectedly, the provision of nonlinear functions for neuron activations in autoencoders provided richer representational capacity of projected data into a lower dimensional space, in comparison to the linear technique of PCA and the simple concept of maximum variance.

The article introduces scoring functions for mapping the values of the performance metrics into quality scores in the range $[0, 1]$. The quality scores are afterward employed for training the NN models.

The performance of three deep NN architectures is investigated – CNNs, RNNs, and HNNs. The networks are trained via supervised regression, where for inputs comprising repetitions of rehabilitation exercises the inferred outputs are quality scores. The best performance was recorded by the CNN models. Although RNNs and HNNs employ recurrent layers that are specifically designed for processing sequential data, the results were not too surprising to our team for two key reasons: (1) the employed dataset is fairly small, consisting of less than 200 repetitions per exercises, and (2) a growing

body of work report of improved performance by CNNs on time-series and movement data [43], [44]. More specifically, recurrent networks utilize a larger number of parameters, thus they are more prone to overfitting on smaller datasets. Further, we observed improved network performance with additive noise for data augmentation, which may be of consideration in related biomedical studies, where data collection is expensive.

Our presented research has several limitations. First, the dataset used for validation of the approach comprises rehabilitation exercises collected with healthy subjects, rather than patients in rehabilitation programs. Second, we do not have a ground truth assessment of the movement quality by clinicians. For example, a movement quality score of 0.8 does not translate into a meaningful clinical score. Third, the approach is validated on measurements acquired with an expensive optical motion capturing system, whereas for practical applications in home-based rehabilitation we envision using an inexpensive color/depth camera (e.g., Kinect-like) for motion capture.

In future work, we will attempt to address the above-listed shortcomings of this study. I.e., we will focus on creating a dataset of rehabilitation exercises performed by patients and labeled by a group of clinicians who will assign quality scores. In addition, we will investigate the implementation of advanced NN architectures for hierarchical spatio-temporal modeling of rehabilitation data.

VIII. CONCLUSION

The article proposes a deep learning-based framework for assessment of rehabilitation exercises. To this end, an autoencoder NN is employed for reducing the dimensionality of skeleton data captured during the performance of repetitions of rehabilitation exercises. Further, the low-dimensional data representation is probabilistically modeled with a GMM and the log-likelihood of the movement repetitions is utilized as a metric for performance evaluation. A scoring function maps the values of the performance metric into movement quality scores. For each rehabilitation exercise a deep NN model is trained to learn the relationship between the movement data and quality scores, and generate quality scores for unseen repetitions of rehabilitation exercises. The experimental results indicate that the movement quality scores generated by the proposed deep learning-based framework closely follow the ground truth quality scores for the movements, and confirm the potential of deep learning models for assessing rehabilitation exercises.

REFERENCES

- [1] S. R. Machlin, J. Chevan, W. W. Yu, and M. W. Zodet, Determinants of utilization and expenditures for episodes of ambulatory physical therapy among adults, *Phys Ther*, vol. 91, no. 7, pp. 1018–1029, Jul. 2011.
- [2] R. Komatireddy, A. Chokshi, J. Basnett, M. Casale, D. Goble, and T. Shubert, Quality and Quantity of Rehabilitation Exercises Delivered By A 3-D Motion Controlled Camera: A Pilot Study, *Int J Phys Med Rehabil*, vol. 2, no. 4, Aug. 2014.
- [3] S. F. Bassett and H. Prapavessis, Home-based physical therapy intervention with adherence-enhancing strategies versus clinic-based management for patients with ankle sprains, *Phys Ther*, vol. 87, no. 9, pp. 1132–1143, Sep. 2007.
- [4] K. Jack, S. M. McLean, J. K. Moffett, and E. Gardiner, Barriers to treatment adherence in physiotherapy outpatient clinics: A systematic review, *Man Ther*, vol. 15, no. 32, pp. 220–228, Jun. 2010.
- [5] K. K. Miller, R. E. Porter, E. DeBaun-Sprague, M. Van Puymbroeck, and A. A. Schmid, Exercise after Stroke: Patient Adherence and Beliefs after Discharge from Rehabilitation, *Top Stroke Rehabil*, vol. 24, no. 2, pp. 142–148, 2017.
- [6] P. Maciejasz, J. Eschweiler, K. Gerlach-Hahn, A. Jansen-Troy, and S. Leonhardt, A survey on robotic devices for upper limb rehabilitation, *Journal of NeuroEngineering and Rehabilitation*, vol. 11, no. 1, p. 3, Jan. 2014.
- [7] L. V. Gauthier et al., Video Game Rehabilitation for Outpatient Stroke (VIGOROUS): protocol for a multi-center comparative effectiveness trial of in-home gamified constraint-induced movement therapy for rehabilitation of chronic upper extremity hemiparesis, *BMC Neurology*, vol. 17, no. 1, p. 109, Jun. 2017.
- [8] D. Antn, A. Goi, A. Illarramendi, J. J. Torres-Unda, and J. Seco, KiReS: A Kinect-based telerehabilitation system, in *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*, 2013, pp. 444–448.
- [9] A. Vakanski, J. M. Ferguson, and S. Lee, Metrics for Performance Evaluation of Patient Exercises during Physical Therapy, *International Journal of Physical Medicine & Rehabilitation*, vol. 05, no. 03, 2017.
- [10] H. Sakoe, Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, pp. 43–49, 1978.
- [11] A. Vakanski, J. M. Ferguson, and S. Lee, Mathematical Modeling and Evaluation of Human Motions in Physical Therapy Using Mixture Density Neural Networks, *J Physiother Phys Rehabil*, vol. 1, no. 4, Dec. 2016.
- [12] Y. Du, W. Wang, and L. Wang, Hierarchical recurrent neural network for skeleton-based action recognition, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1110–1118.
- [13] A. Vakanski, H. Jun, D. Paul, and R. Baker, A Data Set of Human Body Movements for Physical Rehabilitation Exercises, *Data*, vol. 3, no. 1, p. 2, Jan. 2018.
- [14] X. Yun and E. R. Bachmann, Design, Implementation, and Experimental Results of a Quaternion-Based Kalman Filter for Human Body Motion Tracking, *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1216–1227, Dec. 2006.
- [15] J. Yang, Y. Xu, and C. S. Chen, Human action learning via hidden Markov model, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 27, no. 1, pp. 34–44, Jan. 1997.
- [16] Y. Huang, K. B. Englehart, B. Hudgins, and A. D. C. Chan, A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses, *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 11, pp. 1801–1811, Nov. 2005.
- [17] A. Vakanski, I. Mantegh, A. Irish, and F. Janabi-Sharifi, Trajectory Learning for Robot Programming by Demonstration Using Hidden Markov Model and Dynamic Time Warping, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 4, pp. 1039–1052, Aug. 2012.
- [18] R. Houmanfar, M. Karg, and D. Kulic, Movement Analysis of Rehabilitation Exercises: Distance Metrics for Measuring Patient Progress, *IEEE Systems Journal*, vol. 10, no. 3, pp. 1014–1025, Sep. 2016.
- [19] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, Sequential Deep Learning for Human Action Recognition, in *Human Behavior Understanding*, 2011, pp. 29–39.
- [20] T. T. Um, V. Babakeshizadeh, and D. Kuli, Exercise motion classification from large-scale wearable sensor data using convolutional neural networks, in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 2385–2390.
- [21] G. Lefebvre, S. Berlemont, F. Mamalet, and C. Garcia, BLSTM-RNN Based 3D Gesture Classification, in *Artificial Neural Networks and Machine Learning ICANN 2013*, 2013, pp. 381–388.
- [22] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, Recurrent Network Models for Human Dynamics, in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Washington, DC, USA, 2015, pp. 4346–4354.
- [23] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing, in *Proceedings of the 26th International Conference on World Wide Web*, Republic and Canton of Geneva, Switzerland, 2017, pp. 351–360.

- [24] F. J. Ordez and D. Roggen, Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition, *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.
- [25] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis, *arXiv:1604.02808 [cs]*, Apr. 2016.
- [26] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, Structural-RNN: Deep Learning on Spatio-Temporal Graphs, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5308–5317.
- [27] Y. Wang, S. Wang, J. Tang, N. OHare, Y. Chang, and B. Li, Hierarchical Attention Network for Action Recognition in Videos, *arXiv:1607.06416 [cs]*, Jul. 2016.
- [28] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data, in *Association for the Advancement of Artificial Intelligence (AAAI)*, 2017, pp. 4263–4270.
- [29] J. Btepage, M. J. Black, D. Kragic, and H. Kjellstrm, Deep Representation Learning for Human Motion Prediction and Classification, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1591–1599.
- [30] P. E. Taylor, G. J. M. Almeida, T. Kanade, and J. K. Hodgins, Classifying human motion quality for knee osteoarthritis using accelerometers, *Conf Proc IEEE Eng Med Biol Soc*, vol. 2010, pp. 339–343, 2010.
- [31] Z. Zhang, Q. Fang, L. Wang, and P. Barrett, Template matching based motion classification for unsupervised post-stroke rehabilitation, in *International Symposium on Bioelectronics and Bioinformatics 2011*, 2011, pp. 199–202.
- [32] I. Ar and Y. S. Akgul, A computerized recognition system for the home-based physiotherapy exercises using an RGBD camera, *IEEE Trans Neural Syst Rehabil Eng*, vol. 22, no. 6, pp. 1160–1171, Nov. 2014.
- [33] J.-Y. Jung, J. I. Glasgow, and S. H. Scott, Feature selection and classification for assessment of chronic stroke impairment, in *2008 8th IEEE International Conference on Bioinformatics and BioEngineering*, 2008, pp. 1–5.
- [34] C.-J. Su, C.-Y. Chiang, and J.-Y. Huang, Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic, *Applied Soft Computing*, vol. 22, pp. 652–666, Sep. 2014.
- [35] Z. Zhang, Q. Fang, and X. Gu, Objective Assessment of Upper-Limb Mobility for Poststroke Rehabilitation, *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 859–868, Apr. 2016.
- [36] D. Antn, A. Goi, and A. Illarramendi, Exercise Recognition for Kinect-based Telerehabilitation, *Methods Inf Med*, vol. 54, no. 02, pp. 145–155, 2015.
- [37] M. Capecci et al., A Hidden Semi-Markov Model based approach for rehabilitation exercise assessment, *Journal of Biomedical Informatics*, vol. 78, pp. 1–11, Feb. 2018.
- [38] J. F. Lin, M. Karg, and D. Kuli, Movement Primitive Segmentation for Human Motion Modeling: A Framework for Analysis, *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, pp. 325–339, Jun. 2016.
- [39] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2011.
- [40] J. F. Lin and D. Kuli, Online Segmentation of Human Motion for Automated Rehabilitation Exercise Analysis, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 168–180, Jan. 2014.
- [41] F. Bashir, W. Qu, A. Khokhar, and D. Schonfeld, HMM-based motion recognition system using segmented PCA, in *IEEE International Conference on Image Processing 2005*, 2005, vol. 3, pp. III–1288.
- [42] H. Bourlard and Y. Kamp, Auto-association by multilayer perceptrons and singular value decomposition, *Biological cybernetics*, vol. 59, no. 4–5, pp. 291–294, 1988.
- [43] T. M. Le, N. Inoue, and K. Shinoda, A Fine-to-Coarse Convolutional Neural Network for 3D Human Action Recognition, *arXiv:1805.11790 [cs]*, May 2018.
- [44] M. Liu, C. Chen, and H. Liu, 3D action recognition using data visualization and convolutional neural networks, in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 925–930.