

## Article

# Machine-Learning-Based Methodology for Estimation of Shoulder Load in Wheelchair-Related Activities Using Wearables

Sabrina Amrein <sup>1,2</sup>, Charlotte Werner <sup>1,3</sup> , Ursina Arnet <sup>2</sup>  and Wiebe H. K. de Vries <sup>2,\*</sup> 

<sup>1</sup> Rehabilitation Engineering Laboratory, Department of Health Science and Technology, ETH Zurich, 8049 Zurich, Switzerland

<sup>2</sup> Swiss Paraplegic Research, Guido A. Zächstrasse 4, 6207 Nottwil, Switzerland

<sup>3</sup> Spinal Cord Injury Center, University Hospital Balgrist, 8008 Zurich, Switzerland

\* Correspondence: wiebe.devries@paraplegie.ch

**Abstract:** There is a high prevalence of shoulder problems in manual wheelchair users (MWUs) with a spinal cord injury. How shoulder load relates to shoulder problems remains unclear. This study aimed to develop a machine-learning-based methodology to estimate the shoulder load in wheelchair-related activities of daily living using wearable sensors. Ten able-bodied participants equipped with five inertial measurement units (IMU) on their thorax, right arm, and wheelchair performed activities exemplary of daily life of MWUs. Electromyography (EMG) was recorded from the long head of the biceps and medial part of the deltoid. A neural network was trained to predict the shoulder load based on IMU and EMG data. Different cross-validation strategies, sensor setups, and model architectures were examined. The predicted shoulder load was compared to the shoulder load determined with musculoskeletal modeling. A subject-specific biLSTM model trained on a sparse sensor setup yielded the most promising results (mean correlation coefficient =  $0.74 \pm 0.14$ , relative root-mean-squared error =  $8.93\% \pm 2.49\%$ ). The shoulder-load profiles had a mean similarity of  $0.84 \pm 0.10$  over all activities. This study demonstrates the feasibility of using wearable sensors and neural networks to estimate the shoulder load in wheelchair-related activities of daily living.

**Keywords:** machine learning; wearable sensors; IMU; EMG; biomechanical modeling; spinal cord injury; joint force; shoulder



**Citation:** Amrein, S.; Werner, C.; Arnet, U.; de Vries, W.H.K.

Machine-Learning-Based Methodology for Estimation of Shoulder Load in Wheelchair-Related Activities Using Wearables. *Sensors* **2023**, *23*, 1577. <https://doi.org/10.3390/s23031577>

Academic Editor: Winson Lee

Received: 19 December 2022

Revised: 27 January 2023

Accepted: 30 January 2023

Published: 1 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Manual wheelchair users (MWUs) highly rely on their upper limbs for independent mobility during daily life. The shoulder complex becomes the main source of power for locomotion and activities of daily living (ADL). The high prevalence of shoulder problems reported in MWUs (30–70%), such as pain, pathologies, and a limited range of motion [1–5], drastically limits MWUs' participation in daily life, reduces their quality of life, and ultimately increases the health care cost for society [2].

Some ADL are known for an increased shoulder load, such as wheelchair (WC) propulsion, weight-relief lifts, and transfers [6,7]. These tasks are performed multiple times per day, which results in a high exposure of the shoulder [6,8]. Further, the risk factors for joint overload are known, namely the magnitude, frequency, and duration of joint load. If one or a combination of these three factors becomes too high, they can cause shoulder joint overload and consequently shoulder problems [9,10]. A methodology for shoulder-load estimation in MWUs in daily conditions is therefore of substantial importance from two viewpoints: (1) to monitor an MWU's daily shoulder load and (2) to study the longitudinal effects of shoulder load with respect to shoulder pain and pathologies. Knowledge of the daily shoulder load is imperative for the implementation of potential interventions to reduce the shoulder load and shoulder pain during ADL.

Until now, only laboratory-based measurements have examined the shoulder load during WC-related tasks [11–13]. Here, musculoskeletal modeling is used to estimate the

joint load, a method widely accepted as the gold standard. Musculoskeletal modeling requires kinematic data and kinetic data as input to calculate the joint load. Kinetic data alone, acquired, for example, using force plates or knowledge of handled weights, but especially the combination of kinematic and kinetic data can only be reliably measured within a laboratory setting [14,15]. Hence, musculoskeletal modeling for joint-load calculation is restricted to a laboratory setting. As with most laboratory measurements, the extensive equipment as well as the restricted area covered by the motion capture cameras used for kinematic data acquisition may hinder the participants' natural movements or execution of daily tasks. Furthermore, these measurements only represent a snippet of the daily shoulder loading in manual WC users. Additionally, musculoskeletal modeling based on laboratory data allows assessing only one of the three risk factors known from the literature for joint overload, namely the magnitude. To fully characterize shoulder-loading activities, not only the magnitude but also the frequency and duration of specific ADL need to be considered.

These drawbacks of musculoskeletal modeling underline the importance of a new, non-laboratory-based methodology for joint-load estimation in daily conditions. The use of alternative methods, such as wearable sensors, has been excessively investigated in biomechanical applications to address some of the limitations biomechanical laboratories possess [16–19]. Wearable sensors such as inertial measurement units (IMU) and electromyography (EMG) sensors do not require a complex laboratory setup but can track movement or muscle activity in daily conditions. Wearable sensors can easily be attached with straps to the participant or the equipment. They are small, lightweight, and wireless, thus unobtrusive measurement tools allowing full movements. First promising results by Goodwin et al. demonstrated the usefulness of wearable sensors in characterizing humeral elevation, a possible risk factor for shoulder pathology in MWUs with a spinal cord injury [16].

To evaluate sensor data, especially in a sparse measurement setup, machine learning techniques are necessary [20,21]. In recent years, IMU and machine learning techniques, including artificial neural networks (ANN), have been successfully applied in both classification and prediction tasks in biomechanical settings [21–27]. Traditional machine learning algorithms such as Random Forest, Support Vector Machine, etc., require time-intensive feature engineering and manual feature extraction. This is a serious drawback when handling complex, dynamic data with time dependencies, such as human movement and force exertions [21,28]. The advantage of ANNs is their ability to realize an arbitrary mapping of one vector space onto another vector space. ANNs capture some previously unknown information hidden in given data (training data set), learn from it, and apply the learned input–output relationship to new data. This allows a prediction of the behavior of a system with respect to unseen data [23,29].

Recurrent neural networks, including long short-term memory (LSTM) networks, use time dependencies in data to make predictions. By updating an internal state, previous information is retained and will affect future predictions, while irrelevant information is disregarded [30]. The key difference with ANN is that bidirectional recurrent neural networks, including bidirectional LSTM (biLSTM), propagate the input in two layers. In a forward layer, the input passes from past to future, while in a backward layer, the input passes from future to past. In this way, not only previous information but also future information will affect the current prediction [31–33]. This unique characteristic of biLSTM appears to be a huge advantage in the prediction of joint load from wearable sensor data. LSTM networks have been applied for the estimation of lower-limb kinematics [34,35] and the estimation of ground reaction forces and knee-joint kinetics [23] from IMU data. However, LSTM networks are intensive to train: large data sets and considerable memory capacity are required, and the training is time-consuming [30]. These are the main disadvantages of this method. Linear neural networks, on the other hand, are easy to train and often more generalizable if only small data sets can be provided [36]. Stetter et al. utilized wearable sensors in combination with a linear neural network for knee joint force estimation during

sports movements [25]. Similar work has been conducted on upper extremities by de Vries et al. [37]. They published a proof of concept for shoulder joint force estimation, demonstrating the feasibility of utilizing IMU and EMG sensors together with a linear neural network for the estimation of shoulder joint force. This project builds upon the work of de Vries et al. [37] by further exploring the feasibility of utilizing machine learning techniques for shoulder-load estimation in WC-related ADL. So far, no non-laboratory-based methodology to assess shoulder load during ADL in MWUs is available.

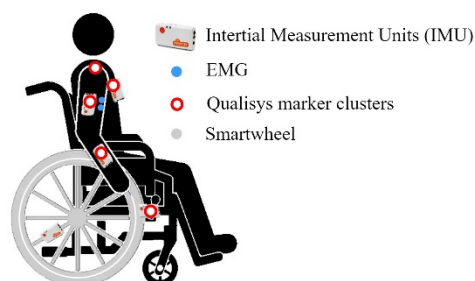
Therefore, this project aims to develop and evaluate a non-laboratory-based methodology for the estimation of shoulder joint load in daily conditions in WC-related ADL using wearable sensor data and ANNs.

## 2. Materials and Methods

### 2.1. Data Collection

Ten able-bodied participants (7 female; age  $39 \pm 9.4$  years; height  $169 \pm 9.1$  cm; weight  $66 \pm 12$  kg) with no pain in the upper extremities and experienced in WC-related activities participated in this study. The study was approved by the Ethikkommission Nordwest- und Zentralschweiz (EKNZ, Project-ID: 2020-01961). The study follows the ICH Good Clinical Practice Guidelines and the Swiss regulation on research involving human beings. All participants were informed of the experimental procedures and gave informed written consent prior to the measurements. Before the actual measurements, participants were familiarized with the WC-related activities until they felt comfortable executing the activities.

The participants were equipped with IMU sensors, EMG sensors, motion capture marker clusters, and a Smartwheel as shown in Figure 1. Five IMU sensors (Shimmer3 IMU Unit, Shimmer, Dublin, Ireland) (sampling frequency 128 Hz,  $\pm 8$  g accelerometer,  $\pm 2000^\circ$ /s gyroscope) were attached to the participants' lower right arm, upper right arm, and thorax, and on the WC frame and wheel. IMU sensors collect information about the acceleration and the angular velocity of the body segment the sensor is attached to.



**Figure 1.** Equipment of a participant.

EMG data (Shimmer3 EMG Unit, Shimmer, Dublin, Ireland) (1024 Hz, gain 12) were collected from the medial deltoid and the long head of the biceps muscle of the participants' right arm. An eight-camera marker-based motion capture system (Oqus, Qualisys AB, Gothenburg, Sweden) (100 Hz) was used to obtain upper body kinematics conforming to Wu et al. [38]. A SmartWheel (Three Rivers Holdings LLC, Mesa, Arizona, USA) (240 Hz) for collecting propulsion kinetics replaced the original right wheel of a standard active wheelchair (Küschall Compact 2017, Küschall AG, Witterswil, Switzerland). The data-collecting systems were synchronized during post-processing by cross-correlation. The participants executed a well-recognizable motion each time data acquisition was initiated.

The participants performed six different activities exemplary of the daily living of MWUs. All participants performed the activities in the same order, namely a first weight-relief lift of 10 s duration; a specified sequence of WC propulsion maneuvers in a restricted space; WC propulsion at 0.56 m/s and 1.1 m/s at 0% inclination, and 0.56 m/s at 6% inclination for 30 s each; a second weight-relief lift of 10 s duration; ascending and descending a short ramp of 12% inclination; manual material handling, specifically placing a weight of

2 kg on three different levels of a shelf, followed by putting the weight into the back pocket of the WC; and desk work for 30 s, such as typing, working with the computer mouse, and making a phone call. Each of these activities was recorded once for each participant.

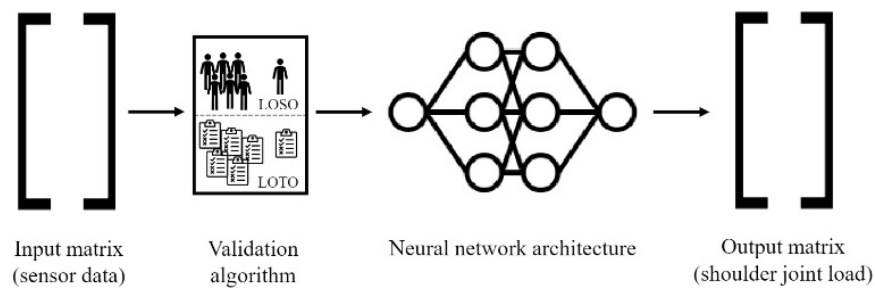
## 2.2. Data Processing and Biomechanical Modeling

The kinematic data from the motion capture system were filtered (Butterworth low-pass filter, fourth order, cut-off frequency 6 Hz). The SmartWheel data were offset corrected, filtered (Butterworth low-pass filter, fourth order, cut-off frequency 20 Hz), and downsampled to 100 Hz. Through the recorded upper-body kinematics and external forces (SmartWheel data for WC propulsion, weight-relief lifts, and ascending/descending a ramp; known weight of 2 kg for manual material handling), 3D shoulder-joint reaction forces (SJRF) were estimated using musculoskeletal modeling within OpenSim [39]. Each participant was individually modeled with a validated upper-extremity model scaled to the participant's height and weight [40]. During static optimization within the OpenSim processing pipeline, the data were downsampled to 25 Hz to reduce the computation time. The equation solved during static optimization considers constraints on the glenohumeral joint force direction which ensures that the calculated muscle forces produce sufficient stabilizing glenohumeral joint compression. The resultant SJRF was filtered (Butterworth low-pass filter, fourth order, cut-off frequency 4 Hz).

The IMU signals (acceleration and angular velocity) were filtered (Butterworth low-pass filter, fourth order, cut-off frequency 10 Hz) and downsampled to 25 Hz to correspond to the specific SJRF signals. EMG data were high-pass filtered (Butterworth, fourth order, cut-off frequency 20 Hz), corrected for offset, rectified, low-pass filtered (Butterworth, fourth order, cut-off frequency 3 Hz), normalized using submaximal isometric contraction, and downsampled to 25 Hz. The participants executed two static postures for normalization of the EMG signal by submaximal isometric contraction. For normalization of the medial deltoid, the participants were holding a weight of 2 kg in 90° abduction with the elbow extended and the thumb pointing frontally. For normalization of the long head of the biceps, the participants were holding a weight of 2 kg in elbow flexion with the forearm pointing frontally and the thumb pointing upwards.

## 2.3. Neural Network Modeling

The ANN developed for this study maps the signals from the 5 IMUs (tri-axial acceleration, tri-axial gyroscope) and 2 EMG sensors to the SJRF time series from the musculoskeletal modeling. The sensor signal matrix ( $N \times 32$ , where  $N$  depicts the trial length) served as input and the SJRF matrix ( $N \times 3$ ) served as target. Input and target were standardized by removing the mean and scaling to unit variance. Standardization happened independently on each input signal by computing the relevant statistics on the training data. A biLSTM was chosen as the preferred model due to its before-mentioned characteristics. The ANN was set up with the PyTorch library in Python (version 3.9.7). The biLSTM model consisted of three biLSTM layers [31,41]. Each biLSTM layer was followed by a dropout layer with a dropout probability of 0.37 to reduce overfitting. The three biLSTM layers each contained 128 neurons and were followed by a ReLU (Rectified Linear Unit) activation function. The model was trained for a maximum of 200 epochs. Training stopped if the validation loss did not decrease for six consecutive epochs. The model parameters resulting in the lowest validation loss during training were saved and reloaded for evaluation on the test data. During the initialization of an ANN, random weights are assigned to all internal connections, followed by the training process. For each initialization, this random weight assignment might result in a different outcome of the training process and hence in a different performance of the trained ANN [42]. Therefore, the model was initialized, trained, and evaluated for ten iterations. The model was tested on unseen data using a leave-one-out validation procedure. Figure 2 shows an overview of the neural network modeling pipeline.



**Figure 2.** Overview of the neural network modeling pipeline.

Different cross-validation strategies, different sensor setups, and different model architectures were examined. While it would have been interesting to train every cross-validation strategy with each sensor setup and model architecture, a stepwise approach was followed to systematize the data for evaluation and to reduce the processing time. This stepwise approach led to the potentially best combination of cross-validation strategy, sensor setup, and model architecture.

- **Cross-Validation Strategy:**

In step one, two different cross-validation strategies were compared. First, the biLSTM model was evaluated using a generalizable leave-one-subject-out (LOSO) cross-validation strategy. Specifically, the neural network was trained on all data from all but one participant (training set) and then tested on the data of the remaining participant (test set). This generalizable cross-validation strategy was compared to a subject-specific strategy, where the biLSTM was evaluated using the leave-one-trial-out (LOTO) strategy. Here, the training set consisted of all but one subtrials of one participant and the test set consisted of the remaining subtrial of the same participant.

To prepare the data for the LOTO train-test split, the activities were divided into two, three, or six subtrials, depending on the activity's length and characteristics. Specifically, this resulted in six subtrials of WC propulsion on a treadmill (two for each condition), three subtrials of WC propulsion in restricted space, one subtrial each of ascending and descending a short ramp, two subtrials of weight-relief lift, three subtrials of manual material handling, and three subtrials of desk work. Seven unrelated subtrials from four different participants had to be excluded from further analysis due to a fault in the muscles' wrapping paths within the OpenSim processing pipeline, resulting in a total of 183 subtrials (69,971 samples) (10 participants  $\times$  19 subtrials—7 invalid subtrials).

The cross-validation strategy leading to better results was specified as the cross-validation strategy for the next steps in the analysis.

- **Sensor Setup:**

In step two, the input matrix was reduced towards a more pragmatic approach. For this pragmatic approach, the sensor setup was reduced to a sparse setup. Only the data from the IMUs attached to the participant's upper arm, the data from the two EMG sensors, and the data from both IMUs attached to the WC contributed to the input matrix ( $N \times 20$ ). The IMU attached to the participant's upper arm was included because it is the IMU placed on the segment distal to the shoulder, the joint of interest. Furthermore, upper-arm movement presumably provides more diverse information for the different ADL than the trunk movement. EMG was regarded as necessary information for SJRF estimation, as it delivers information on the musculoskeletal response of the activity and is related to the exerted force. The sensor setup leading to better results was specified as the sensor setup for the last step of the analysis.



- **Model Architecture:**

In step three, the model architecture was simplified to minimize the processing time. The complex biLSTM model was compared to the simpler linear model. The linear model had two hidden layers, one with 250 and one with 100 neurons. A ReLU activation function followed each hidden layer. Similar to the training procedure of the biLSTM model, the linear model was trained for a maximum of 200 epochs or until the validation loss did not decrease for six consecutive epochs. The model parameters resulting in the lowest validation loss during training were saved and reloaded for evaluation on the test data. The linear model was initialized, trained, and evaluated for ten iterations.

#### 2.4. Statistical Analysis

The total SJRF ( $F_{tot}$ ) was calculated as the Euclidean norm of the three individual components ( $F_x$ ,  $F_y$ ,  $F_z$ ). The similarity between the ground-truth  $F_{tot}$  from the musculoskeletal model and the ANN-predicted  $F_{tot}$  was analyzed using Pearson's correlation coefficient (PCC) and relative root-mean-squared error (rRMSE), which normalizes the RMSE by the range of the ground-truth  $F_{tot}$ . For evaluation of the neural networks' performances, the subtrials were concatenated to the original eight activities. The neural networks' performances were analyzed on an activity level and a participant level. For analysis on the activity level, each participant's mean performance across the ten iterations for the specific activity was calculated. As there were ten different participants, this led to a total of ten PCC and rRMSE values for each activity, one for each participant. For analysis on the participant level, all eight activities of one iteration were concatenated for each participant. Concatenation of all eight activities formed a complete iteration. Every complete iteration was analyzed, resulting in ten PCC and rRMSE values for each participant, one for each complete iteration.

The shoulder-load profiles were regarded as histograms and evaluated as such. To determine the similarity between the predicted shoulder-load profile ( $\hat{y}$ ) and the ground-truth shoulder-load profile ( $y$ ), the intersection ( $I$ ) of the two histograms was calculated according to Swain and Ballard's [43] conform Equation (1):

$$I(y, \hat{y}) = \frac{\sum_{i=1}^n \min(y_i, \hat{y}_i)}{\sum_{i=1}^n y_i} \quad (1)$$

### 3. Results

#### 3.1. Cross-Validation Strategy

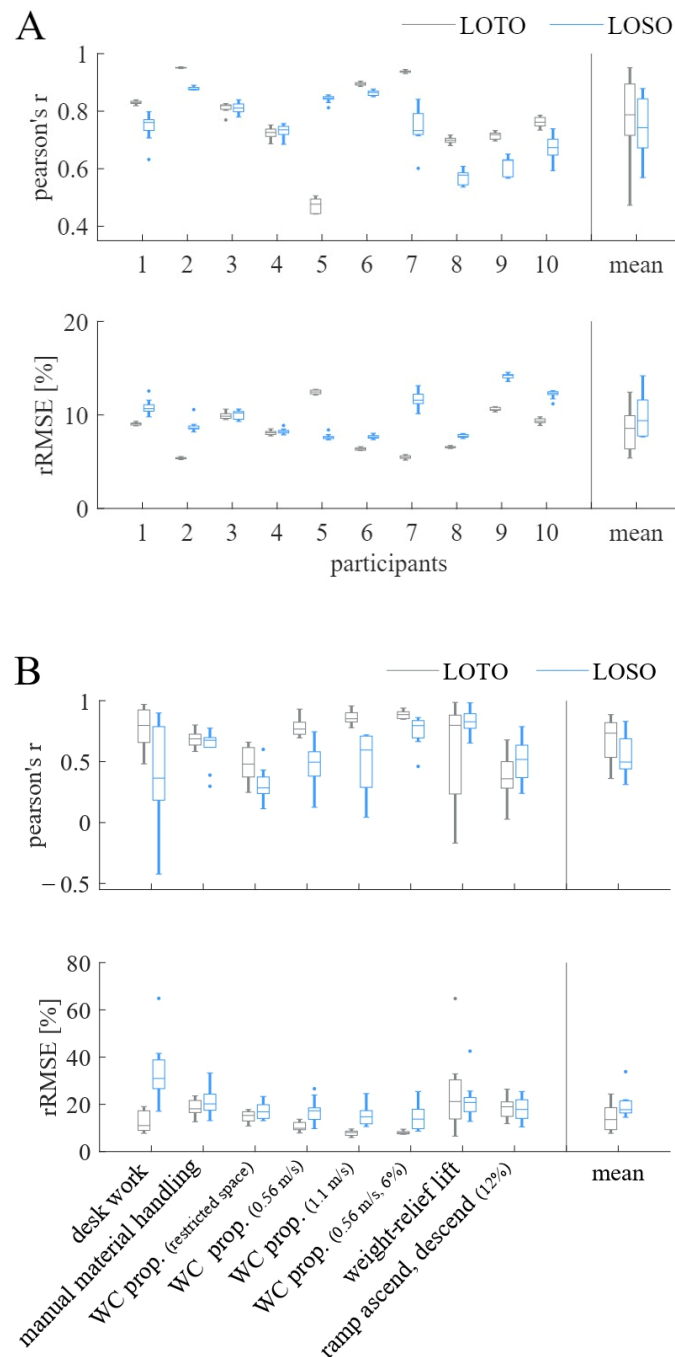
In step one, the two cross-validation strategies (LOTO vs. LOSO) were compared. Figure 3A,B shows the comparison between the subject-specific LOTO organization strategy and the generalizable LOSO organization strategy when using the biLSTM model. In Figure 3A, PCC and rRMSE are depicted for the ANN's performance on the participant level, while 3B shows the same metrics for the ANN's performance on the activity level. Using the subject-specific LOTO validation strategy results in a distinctly better prediction accuracy (higher PCC values, lower rRMSE values) for all participants but participants 3, 4, and 5. The mean prediction accuracy on the participant level for the LOTO validation strategy was PCC = 0.78, rRMSE = 8.32%, while the mean prediction accuracy for the LOSO validation strategy was PCC = 0.75, rRMSE = 9.89%. Hence, the LOTO was chosen as the preferred validation strategy.

Figure 3B shows a large difference in the prediction accuracy between participants for the weight-relief lift activity when using the LOTO organization strategy but not for the LOSO organization strategy.

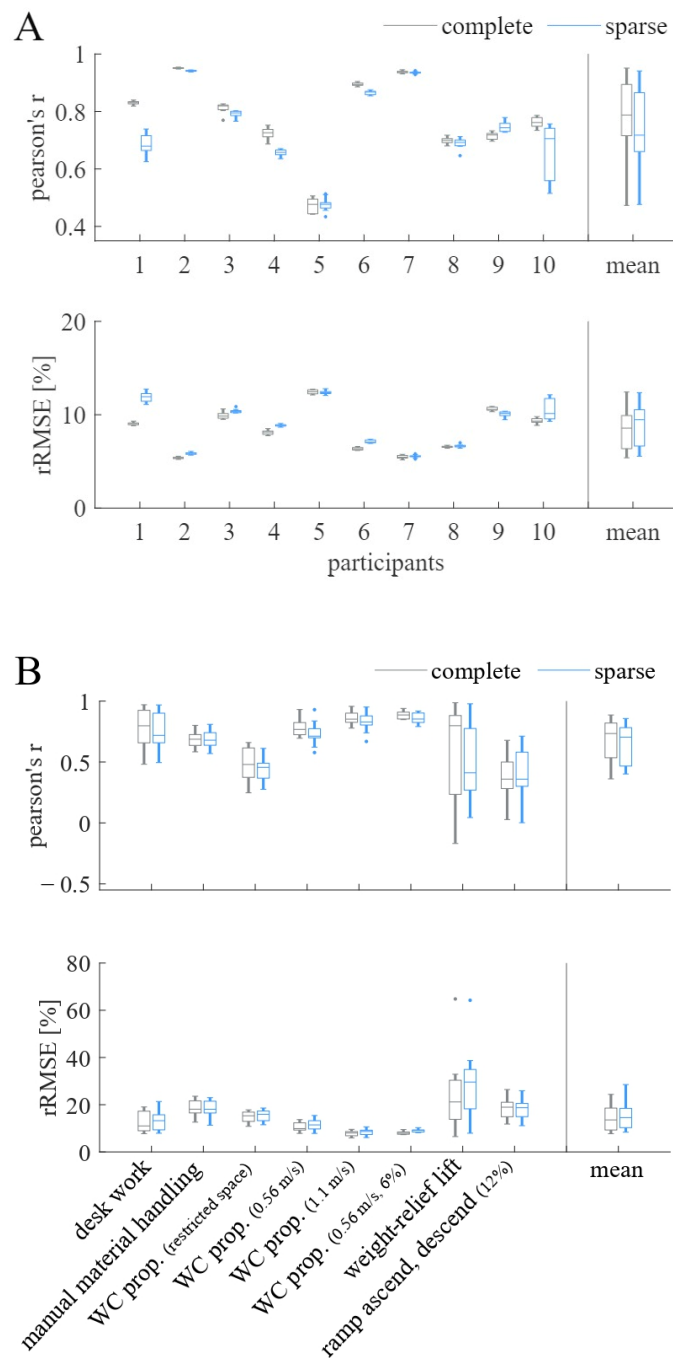
#### 3.2. Sensor Setup

In step two, the two different sensor setups (complete vs. sparse) were compared. Figure 4A,B visualizes the biLSTM model's performances on the participant level and the activity level, respectively, when combining the LOTO cross-validation strategy with

different sensor setups. Figure 4A shows that the model performs better having the complete sensor data available on the participant level (PCC = 0.78, rRMSE = 8.32%). With the sparse sensor setup, the mean prediction accuracy on the participant level was PCC = 0.74, rRMSE = 8.93%. However, on the activity level, these differences were negligible, as shown in Figure 4B. Due to the small differences between the two sensor setups the sparse setup was chosen.



**Figure 3.** Pearson's  $r$  and relative root-mean-squared error (rRMSE) for comparison of leave-one-trial-out (LOTO) and leave-one-subject-out (LOSO) cross-validation strategies when using the biLSTM model (A) on participants and (B) on activities. The mean performance of all participants and activities, respectively, are presented in the last column titled "mean".

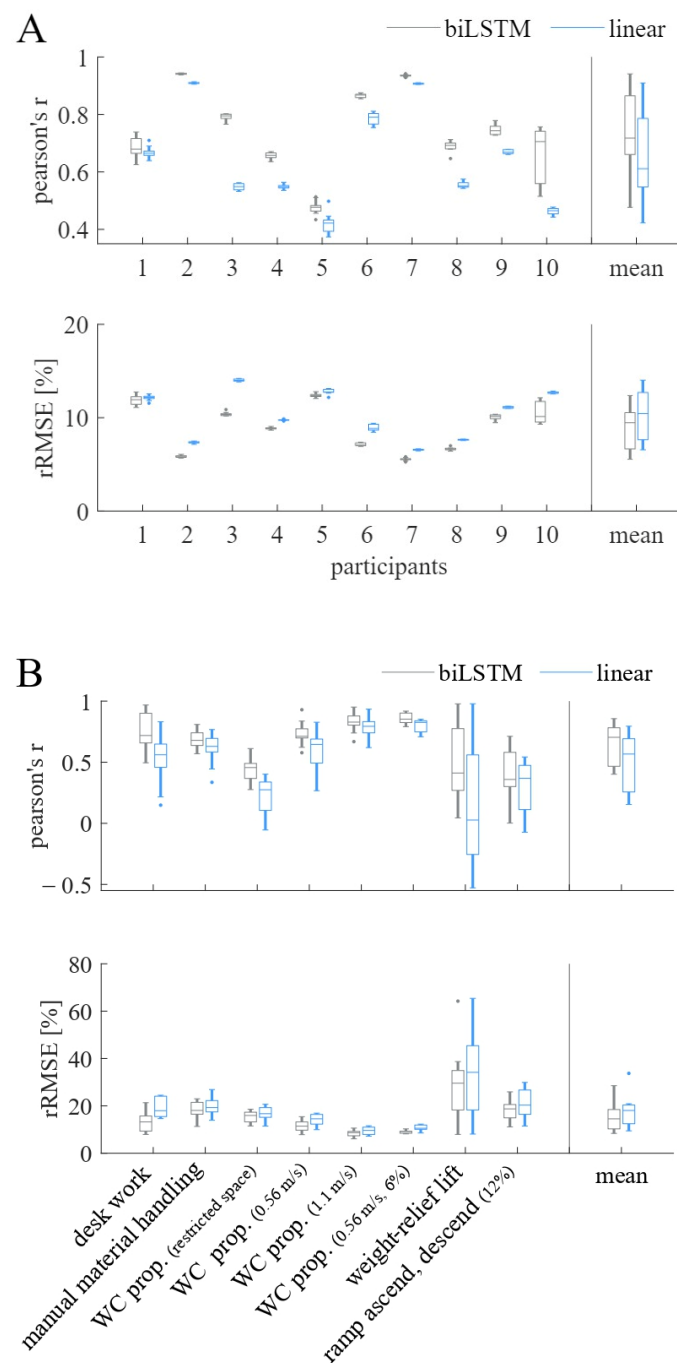


**Figure 4.** Pearson’s  $r$  and relative root-mean-squared error (rRMSE) for comparison of the complete sensor setup and the sparse sensor setup (upper-arm sensor and both WC sensors) when using the LOTO cross-validation strategy and biLSTM model (A) on participants and (B) on activities. The mean performance of all participants and activities, respectively, are presented in the last column, titled “mean”.

### 3.3. Model Architecture

In step three, the model architecture was simplified to minimize the processing time. Figure 5A,B compares the complex biLSTM model with the simpler linear model when using the LOTO cross-validation strategy and the sparse sensor setup. Using the linear model, the mean prediction accuracy on the participant level was  $PCC = 0.65$ ,  $rRMSE = 10.30\%$ . Overall, the biLSTM outperforms the linear model on both the activity and the participant level ( $PCC = 0.74$ ,  $rRMSE = 8.93\%$ ).

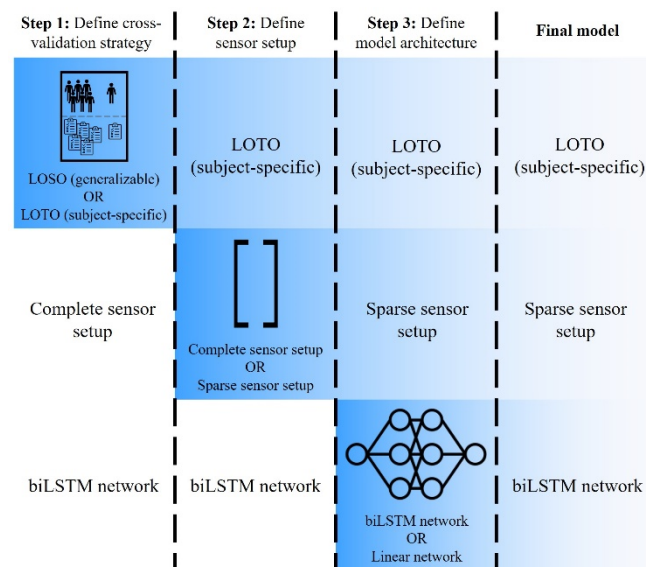




**Figure 5.** Pearson's  $r$  and relative root-mean-squared error (rRMSE) for comparison of the biLSTM model and the linear model when using the LOTO cross-validation strategy and the sparse sensor setup (A) on participants and (B) on activities. The mean performance of all participants and activities, respectively, area presented in the last column, titled "mean".

### 3.4. Final Model

Figure 6 provides an overview of the stepwise approach. It shows the setups used for the individual steps, the results of the individual steps, and the setup of the final model. Specifically, the final model combines the subject-specific LOTO cross-validation strategy with the biLSTM network and the sparse sensor setup.

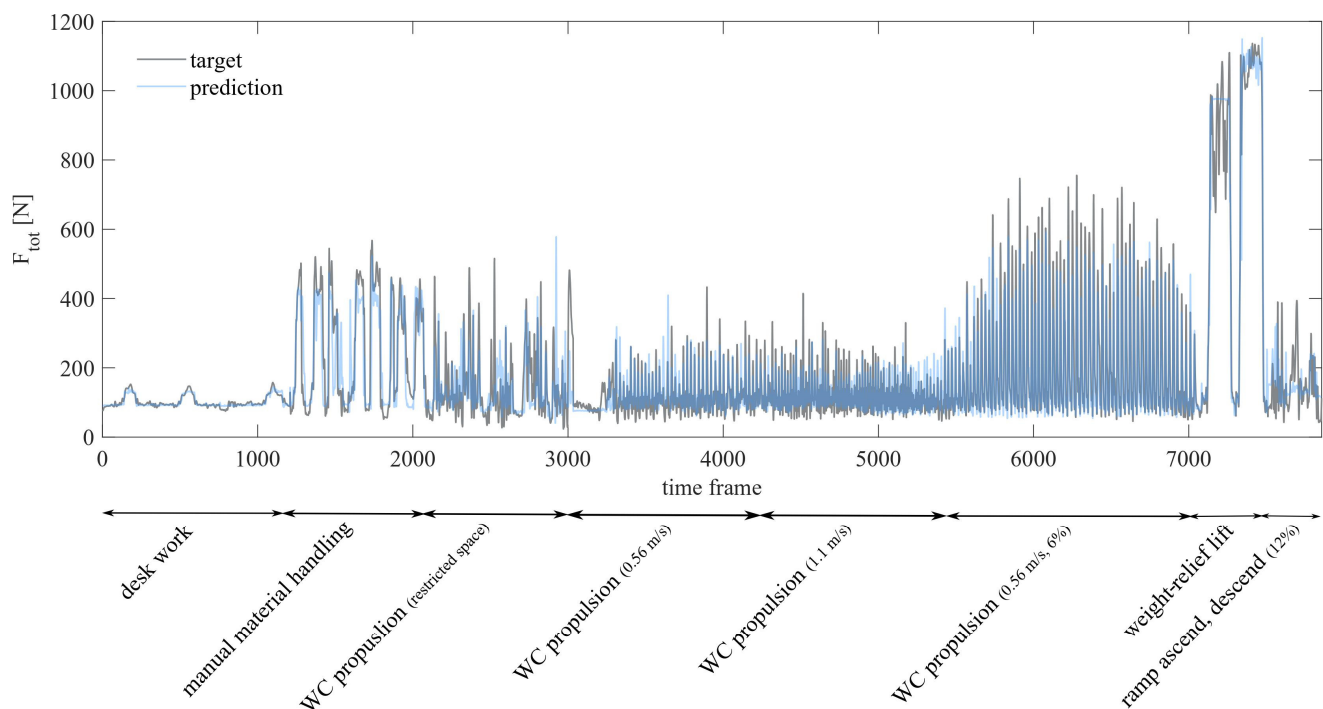


**Figure 6.** Stepwise approach for finding the best combination of cross-validation strategy (step 1), sensor setup (step 2), and model architecture (step 3) and the respective results.

The results of the final model show inter-participant differences (Figure 5A). Likewise, the prediction accuracy on the activity level follows a distinct pattern, independent of the model architecture (Figure 5B). Concretely, WC propulsion in a restricted space achieves a lower prediction accuracy than manual material handling, which in turn has a lower prediction accuracy than the three WC propulsion activities on the treadmill.

Figure 5B shows that the three WC propulsion activities on the treadmill reach the highest prediction accuracy with a small difference between participants.

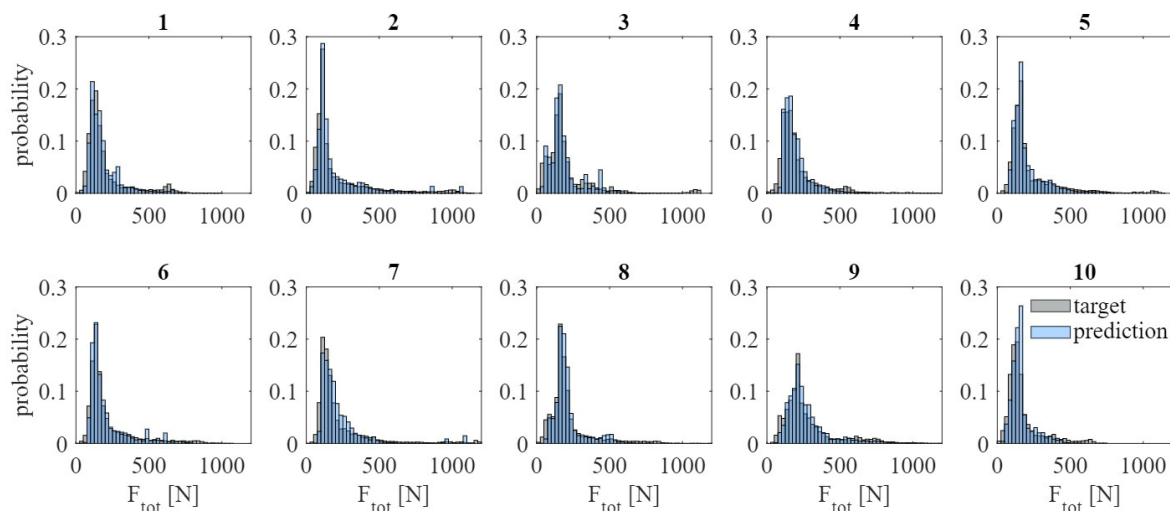
Figure 7 shows the predicted  $F_{tot}$  and the ground-truth  $F_{tot}$  of one complete iteration for one participant using the final model as described in Figure 6.



**Figure 7.** Ground-truth  $F_{tot}$  and predicted  $F_{tot}$  of one complete iteration for participant 2 using the final model.

### 3.5. Shoulder-Load Profiles

Figure 8 shows the shoulder-load profiles for all ten participants and a random iteration when using the final model as described in Figure 6. The mean intersection values are high for all participants with a consistently low standard deviation ( $I \geq 0.81 \pm 0.01$ ), as listed in Table 1.



**Figure 8.** Shoulder-load profiles for each participant over all activities using the final model (bin width = 25 N).

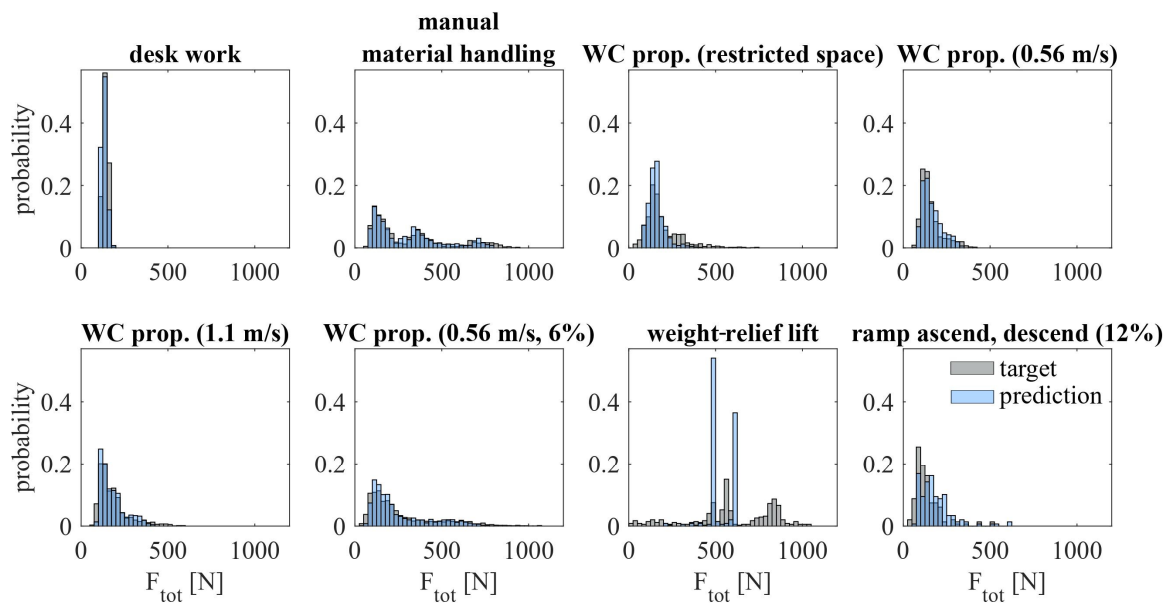
**Table 1.** Intersection values for participants, presented as mean (and standard deviation).

Participant	Intersection
1	0.83 (0.01)
2	0.89 (0.00)
3	0.84 (0.01)
4	0.85 (0.00)
5	0.81 (0.01)
6	0.87 (0.01)
7	0.89 (0.00)
8	0.85 (0.01)
9	0.86 (0.01)
10	0.83 (0.01)
Mean	0.85 (0.03)

Figure 9 shows the shoulder-load profiles for all activities of one participant exemplary of all participants. Weight-relief lift has the lowest mean intersection value and a distinctly higher standard deviation than all other activities, as listed in Table 2.

**Table 2.** Intersection values for activities, presented as mean (and standard deviation).

Activity	Intersection
Desk work	0.97 (0.01)
Manual material handling	0.81 (0.02)
WC propulsion (restricted space)	0.78 (0.03)
WC propulsion (0.56 m/s)	0.92 (0.02)
WC propulsion (1.1 m/s)	0.93 (0.01)
WC propulsion (0.56 m/s, 6%)	0.89 (0.01)
Weight-relief lift	0.66 (0.23)
Ramp ascend, descend	0.77 (0.10)
Mean	0.84 (0.10)



**Figure 9.** Exemplary shoulder-load profiles for all activities of one participant using the final model (bin width = 25 N).

#### 4. Discussion

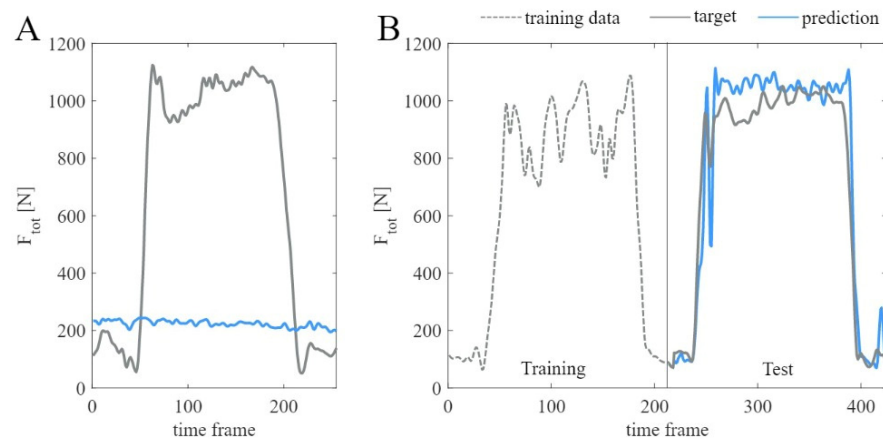
This study investigated the usage of ANNs for the continuous estimation of SJRF from wearable sensors in WC-related ADL. For this reason, different cross-validation strategies, different sensor setups, and different model architectures were examined.

##### 4.1. Cross-Validation Strategy

The results indicate that a subject-specific cross-validation strategy (LOTO) attains a higher prediction accuracy than a generalizable cross-validation strategy (LOSO). A generalizable strategy would have been preferable, as such a model needs to be trained only once on a broad spectrum of the activities of interest and can then reliably predict data from unseen participants. A model based on a subject-specific strategy, on the other hand, needs to be trained anew for each participant. Such a model is more resource-intensive, as it requires the collection of subject-specific data within a laboratory setting and time-consuming training of the model.

The large difference between participants observed for the weight-relief lift activity using the LOTO organization strategy can be explained by investigating the outlier, participant 5, more closely. For this participant, the LOSO cross-validation strategy performed clearly better than the LOTO strategy. Closer investigation of that participant's data revealed that one weight-relief lift had to be excluded from the data set due to an error within the OpenSim processing pipeline. Only one weight-relief lift activity remained in this participant's data set. When this weight-relief lift constituted the test set, no similar data were included in the training set. Accordingly, following the LOTO approach, the prediction of this specific activity was poor, as shown in Figure 10A. In contrast, when a participant's data set contained two weight-relief lift activities, one was always included in the training set if the other constituted the test set, as visualized in Figure 10B.

Using the generalizable LOSO organization strategy in this case massively increases the prediction accuracy. The model has seen weight-relief lift activities from other participants during training before predicting the data of participant 5. This observation is in line with the previously published literature, which all propose to focus the training data on the type of activity expected to be encountered in the test data [30,35,37].



**Figure 10.** Shoulder joint force for the weight-relief lift activity for participant 5 (A) and participant 2 (B) using the LOTO cross-validation strategy and biLSTM model with the sparse sensor setup. Activity-related training data for participant 5 were missing from the data set.

#### 4.2. Sensor Setup

For comparison of the two sensor setups, the LOTO cross-validation strategy and the biLSTM model were used. The ANN using the complete sensor setup performed only slightly better than the ANN using the sparse sensor setup. The IMU sensors attached to the participant's thorax and lower right arm, which are ignored for the sparse setup, possibly provide mostly redundant information already provided by the sensor attached to the participant's upper right arm. This finding is relevant insofar as it increases the applicability of the methodology. Using only one body-bound sensor is more convenient and less restrictive for the participant. Hence, the measured activities will be closer to the participant's natural activities. Additionally, a sparse sensor setup is less resource-intensive. Preparing the participant, processing the data, and training the ANN will demand less time. The benefits of the sparse sensor setup outweigh the small accuracy tradeoff and the sparse setup is therefore the preferred option to the complete setup. The usage of an even sparser sensor setup was not investigated. Reducing the setup further is not expected to improve the quality of the data. The WC-bound sensors do not hamper the participant's convenience or restrict the movements in any case.

#### 4.3. Model Architecture

For comparison of the two model architectures, the LOTO cross-validation strategy and the sparse sensor setup were used. The results suggest that the linear model cannot learn the intricate relation between sensor data and SJRF.

#### 4.4. Final Model

With each iteration, the ANN is initiated and random weights are assigned. The small difference between iterations implies that the ANN's predictive power is independent of the random weights' assignment. This suggests that for future application in research it is sufficient to initialize and train a single ANN.

The distinct pattern observed for participants and activities indicates that the prediction accuracy for an individual participant and activity is highly dependent on the training data. The absolute accuracy of the predicted SJRF changes with the type of ANN used. However, the prediction accuracy in relation to other participants or activities is consistent, independent of the ANN used.

The prediction accuracy of the ANN strongly varied between activities. Here, the accuracy seems to correlate with the duration of the initial activity. Long activities were split into several long subtrials. That way, more data focused on the activity of interest were included in the training set, which improved the performance on the test set. Ascending and descending a ramp is a short activity. When split into subtrials, only a little similar data

were included in the training set. This is reflected in the low prediction accuracy. These findings further underline the importance of including sufficient data in the training set. Sufficient data is difficult to define and a topic for further research. For this study, including a full activity and not only subtrials in the training set could potentially already improve the prediction accuracy.

A potential reason for the high prediction accuracy observed in the three WC propulsion activities on the treadmill is the repeatable characteristics of the activity. In contrast, activities with a higher rate of variation such as WC propulsion in a restricted space or manual material handling reach poorer accuracies. Stetter et al. made a similar observation for the prediction of knee joint forces, where they observed the highest predictive power for moderate running and only limited predictive power for activities with higher variation, such as sprint starts and full stops [25].

Stetter et al. [25] used an ANN for the prediction of the three individual components of knee joint force ( $F_x$ ,  $F_y$ ,  $F_z$ ). The ANN-predicted knee joint forces yielded PCC values ranging from 0.25 to 0.94 and rRMSE values ranging from 14.2% to 45.9%, depending on the component and the activity. The PCC value for the total knee joint force, although not reported, seems to be similar to our results; the rRMSE value is lower in our study. De Vries et al. [37] trained a linear model to predict SJRF during ADL in one healthy subject based on wearable sensors. They reported a good to excellent prediction accuracy (intraclass correlation coefficients ranging from 0.83 to 0.98). The use of different evaluation metrics makes a direct comparison of the results between our study and the study of de Vries et al. difficult. Additionally, the set of activities measured is more diverse and complex in our study.

#### 4.5. Shoulder-Load Profiles

There is a consistently high similarity between the ground-truth and the predicted shoulder-load profiles for all participants and most activities. One possible explanation for this high similarity between the shoulder-load profiles is the effect of binning. While a small deviation from the true SJRF has a possibly high effect on the prediction accuracy of SJRF, binning nullifies this effect. Another possible explanation is the equalization of deviations over time. The SJRF is predicted too high for some activities and too low for others. Both false predictions have a decisive effect on the predicted SJRF and hence on PCC. With shoulder-load profiles, however, these false predictions will equalize over time.

A potential reason for the noticeably low intersection value of the weight-relief lift has been extensively discussed in Section 4.1.

#### 4.6. Limitations and Future Research

Caution is required as the methodology was developed using data from able-bodied participants. A validation study with spinal-cord-injured participants is in progress. The results show that a subject-specific algorithm exceeds a generalizable algorithm in prediction accuracy. From a pragmatic point of view, a generalizable algorithm would be preferable to a subject-specific algorithm as it is less resource-intensive. Future research could focus on improving the prediction accuracy of the generalizable algorithm either by increasing the data set and hence the amount of training data or by performing a sensor-to-segment alignment and hence reducing the variability within the training data. The ground-truth SJRF for this study was based on musculoskeletal modeling. Musculoskeletal modeling in turn is based on several assumptions, such as intrinsic muscle parameters that cannot be measured in vivo, and therefore has its specific limitations. Providing the best means of reference data for the ANN modeling could help to predict the SJRF more precisely. A further limitation of the work might be seen in the absence of a wide comparison of available machine learning methods. This absence is justified by the careful selection of methods that have previously been successfully applied to similar problems in the estimation of joint load from wearable sensor data.



## 5. Conclusions

This work is a considerable step towards assessing shoulder load in daily life, which has not been achieved yet. The results of this study prove the feasibility of utilizing neural networks for quantifying the shoulder load in WC-related ADL based on IMU and EMG data. Specifically, the shoulder-load profiles for participants showed exceptional agreement between the ground-truth SJRF and the ANN-predicted SJRF. Knowledge of shoulder-load profiles combined with knowledge of the type of activity performed will introduce relevant targets for the reduction in shoulder-joint load, and hence reduction in shoulder pain and shoulder pathologies.

**Author Contributions:** Conceptualization, U.A. and W.H.K.d.V.; Data curation, S.A. and W.H.K.d.V.; Formal analysis, S.A., C.W. and W.H.K.d.V.; Investigation, S.A. and W.H.K.d.V.; Methodology, S.A., C.W., U.A. and W.H.K.d.V.; Project administration, S.A. and W.H.K.d.V.; Resources, U.A. and W.H.K.d.V.; Software, S.A.; Supervision, C.W. and W.H.K.d.V.; Validation, S.A.; Visualization, S.A.; Writing—original draft, S.A.; Writing—review and editing, S.A., C.W., U.A. and W.H.K.d.V.; Funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethikkommission Nordwest-und Zentralschweiz (EKNZ), project-ID 2020-01961.

**Informed Consent Statement:** Informed consent was obtained from all participants involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We would like to thank the participants for their cooperation in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jensen, M.P.; Hoffman, A.J.; Cardenas, D.D. Chronic pain in individuals with spinal cord injury: A survey and longitudinal study. *Spinal Cord* **2005**, *43*, 704–712. [[CrossRef](#)] [[PubMed](#)]
2. Turner, J.A.; Cardenas, D.D.; Warms, C.A.; McClellan, C.B. Chronic pain associated with spinal cord injuries: A community survey. *Arch. Phys. Med. Rehabil.* **2001**, *82*, 501–509. [[CrossRef](#)] [[PubMed](#)]
3. Arnet, U.; de Vries, W.H.; Eriks-Hoogland, I.; Wisianowsky, C.; van der Woude, L.H.V.; Veeger, D.H.E.J.; Berger, M. MRI evaluation of shoulder pathologies in wheelchair users with spinal cord injury and the relation to shoulder pain. *J. Spinal Cord. Med.* **2021**, *45*, 916–929. [[CrossRef](#)] [[PubMed](#)]
4. Eriks-Hoogland, I.E.; de Groot, S.; Post, M.W.; van der Woude, L.H. Correlation of shoulder range of motion limitations at discharge with limitations in activities and participation one year later in persons with spinal cord injury. *J. Rehabil. Med.* **2011**, *43*, 210–215. [[CrossRef](#)]
5. Bossuyt, F.M.; Arnet, U.; Brinkhof, M.W.G.; Eriks-Hoogland, I.; Lay, V.; Müller, R.; Sunnåker, M.; Hinrichs, T. Shoulder pain in the Swiss spinal cord injury community: Prevalence and associated factors. *Disabil. Rehabil.* **2018**, *40*, 798–805. [[CrossRef](#)] [[PubMed](#)]
6. Van Drongelen, S.; Van der Woude, L.H.; Janssen, T.W.; Angenot, E.L.; Chadwick, E.K.; Veeger, D.H. Mechanical load on the upper extremity during wheelchair activities. *Arch. Phys. Med. Rehabil.* **2005**, *86*, 1214–1220. [[CrossRef](#)] [[PubMed](#)]
7. Gutierrez, D.D.; Thompson, L.; Kemp, B.; Mulroy, S.J. The relationship of shoulder pain intensity to quality of life, physical activity, and community participation in persons with paraplegia. *J. Spinal Cord. Med.* **2007**, *30*, 251–255. [[CrossRef](#)] [[PubMed](#)]
8. Janssen, T.W.; van Oers, C.A.; van der Woude, L.H.; Hollander, A.P. Physical strain in daily life of wheelchair users with spinal cord injuries. *Med. Sci. Sports Exerc.* **1994**, *26*, 661–670. [[CrossRef](#)]
9. Walford, S.L.; Requejo, P.S.; Mulroy, S.J.; Neptune, R.R. Predictors of shoulder pain in manual wheelchair users. *Clin. Biomech.* **2019**, *65*, 1–12. [[CrossRef](#)]
10. Hoozemans, M.J.; van der Beek, A.J.; Frings-Dresen, M.H.; van Dijk, F.J.; van der Woude, L.H. Pushing and pulling in relation to musculoskeletal disorders: A review of risk factors. *Ergonomics* **1998**, *41*, 757–781. [[CrossRef](#)]
11. Arnet, U.; van Drongelen, S.; Scheel-Sailer, A.; van der Woude, L.H.; Veeger, D.H. Shoulder load during synchronous handcycling and handrim wheelchair propulsion in persons with paraplegia. *J. Rehabil. Med.* **2012**, *44*, 222–228. [[CrossRef](#)]
12. Veeger, H.E.; Rozendaal, L.A.; van der Helm, F.C. Load on the shoulder in low intensity wheelchair propulsion. *Clin. Biomech.* **2002**, *17*, 211–218. [[CrossRef](#)]

13. Mercer, J.L.; Boninger, M.; Koontz, A.; Ren, D.; Dyson-Hudson, T.; Cooper, R. Shoulder joint kinetics and pathology in manual wheelchair users. *Clin. Biomech.* **2006**, *21*, 781–789. [[CrossRef](#)] [[PubMed](#)]
14. Ancillao, A.; Tedesco, S.; Barton, J.; O'Flynn, B. Indirect Measurement of Ground Reaction Forces and Moments by Means of Wearable Inertial Sensors: A Systematic Review. *Sensors* **2018**, *18*, 2564. [[CrossRef](#)] [[PubMed](#)]
15. Camomilla, V.; Bergamini, E.; Fantozzi, S.; Vannozzi, G. Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review. *Sensors* **2018**, *18*, 873. [[CrossRef](#)] [[PubMed](#)]
16. Goodwin, B.M.; Cain, S.M.; Van Straaten, M.G.; Fortune, E.; Jahanian, O.; Morrow, M.M.B. Humeral elevation workspace during daily life of adults with spinal cord injury who use a manual wheelchair compared to age and sex matched able-bodied controls. *PLoS ONE* **2021**, *16*, e0248978. [[CrossRef](#)] [[PubMed](#)]
17. Robert-Lachaine, X.; Mecheri, H.; Larue, C.; Plamondon, A. Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis. *Med. Biol. Eng. Comput.* **2017**, *55*, 609–619. [[CrossRef](#)] [[PubMed](#)]
18. Teufl, W.; Miezal, M.; Taetz, B.; Fröhlich, M.; Bleser, G. Validity of inertial sensor based 3D joint kinematics of static and dynamic sport and physiotherapy specific movements. *PLoS ONE* **2019**, *14*, e0213064. [[CrossRef](#)] [[PubMed](#)]
19. Morrow, M.M.B.; Lowndes, B.; Fortune, E.; Kaufman, K.R.; Hallbeck, M.S. Validation of Inertial Measurement Units for Upper Body Kinematics. *J. Appl. Biomech.* **2017**, *33*, 227–232. [[CrossRef](#)]
20. Kaixuan, C.; Dalin, Z.; Lina, Y.; Bin, G.; Zhiwen, Y.; Yunhao, L. Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges and Opportunities. *ACM Comput. Surv.* **2021**, *54*, 1–40.
21. Gupta, S. Deep learning based human activity recognition (HAR) using wearable sensor data. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100046. [[CrossRef](#)]
22. Hendry, D.; Leadbetter, R.; McKee, K.; Hopper, L.; Wild, C.; O'Sullivan, P.; Straker, L.; Campbell, A. An Exploration of Machine-Learning Estimation of Ground Reaction Force from Wearable Sensor Data. *Sensors* **2020**, *20*, 740. [[CrossRef](#)] [[PubMed](#)]
23. Cerfoglio, S.; Galli, M.; Tarabini, M.; Bertozzi, F.; Sforza, C.; Zago, M. Machine Learning-Based Estimation of Ground Reaction Forces and Knee Joint Kinetics from Inertial Sensors While Performing a Vertical Drop Jump. *Sensors* **2021**, *21*, 7709. [[CrossRef](#)] [[PubMed](#)]
24. Jiang, X.; Gholami, M.; Khoshnam, M.; Eng, J.J.; Menon, C. Estimation of Ankle Joint Power during Walking Using Two Inertial Sensors. *Sensors* **2019**, *19*, 2796. [[CrossRef](#)]
25. Stetter, B.J.; Ringhof, S.; Krafft, F.C.; Sell, S.; Stein, T. Estimation of Knee Joint Forces in Sport Movements Using Wearable Sensors and Machine Learning. *Sensors* **2019**, *19*, 3690. [[CrossRef](#)]
26. de Vries, W.H.K.; Amrein, S.; Arnet, U.; Mayrhuber, L.; Ehrmann, C.; Veeger, H.E.J. Classification of Wheelchair Related Shoulder Loading Activities from Wearable Sensor Data: A Machine Learning Approach. *Sensors* **2022**, *22*, 7404. [[CrossRef](#)] [[PubMed](#)]
27. Fortune, E.; Cloud-Biebl, B.A.; Madansingh, S.I.; Ngufor, C.G.; Van Straaten, M.G.; Goodwin, B.M.; Murphree, D.H.; Zhao, K.D.; Morrow, M.M. Estimation of manual wheelchair-based activities in the free-living environment using a neural network model with inertial body-worn sensors. *J. Electromyogr. Kinesiol.* **2022**, *62*, 102337. [[CrossRef](#)]
28. Weiss, G.M.; Yoneda, K.; Hayajneh, T. Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living. *IEEE Access* **2019**, *7*, 133190–133202. [[CrossRef](#)]
29. Maggiora, G.M.; Elrod, D.W.; Trenary, R.G. Computational neural networks as model-free mapping devices. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 732–741. [[CrossRef](#)]
30. Mundt, M.; Johnson, W.R.; Potthast, W.; Markert, B.; Mian, A.; Alderson, J. A Comparison of Three Neural Network Approaches for Estimating Joint Angles and Moments from Inertial Measurement Units. *Sensors* **2021**, *21*, 4535. [[CrossRef](#)]
31. Rupapara, V.; Rustam, F.; Amaar, A.; Washington, P.B.; Lee, E.; Ashraf, I. Deepfake tweets classification using stacked Bi-LSTM and words embedding. *PeerJ Comput. Sci.* **2021**, *7*, e745. [[CrossRef](#)] [[PubMed](#)]
32. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [[CrossRef](#)]
33. Rahman, M.; Watanobe, Y.; Nakamura, K. A Bidirectional LSTM Language Model for Code Evaluation and Repair. *Symmetry* **2021**, *13*, 247. [[CrossRef](#)]
34. Mundt, M.; Thomsen, W.; Witter, T.; Koeppe, A.; David, S.; Bamer, F.; Potthast, W.; Markert, B. Prediction of lower limb joint angles and moments during gait using artificial neural networks. *Med. Biol. Eng. Comput.* **2020**, *58*, 211–225. [[CrossRef](#)] [[PubMed](#)]
35. Rapp, E.; Shin, S.; Thomsen, W.; Ferber, R.; Halilaj, E. Estimation of kinematics from inertial measurement units using a combined deep learning and optimization framework. *J. Biomech.* **2021**, *116*, 110229. [[CrossRef](#)]
36. Halilaj, E.; Rajagopal, A.; Fiterau, M.; Hicks, J.L.; Hastie, T.J.; Delp, S.L. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *J. Biomech.* **2018**, *81*, 1–11. [[CrossRef](#)] [[PubMed](#)]
37. de Vries, W.H.K.; Veeger, H.E.J.; Baten, C.T.M.; van der Helm, F.C.T. Can shoulder joint reaction forces be estimated by neural networks? *J. Biomech.* **2016**, *49*, 73–79. [[CrossRef](#)]
38. Wu, G.; van der Helm, F.C.; Veeger, H.E.; Makhous, M.; Van Roy, P.; Anglin, C.; Nagels, J.; Karduna, A.R.; McQuade, K.; Wang, X.; et al. ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: Shoulder, elbow, wrist and hand. *J. Biomech.* **2005**, *38*, 981–992. [[CrossRef](#)]
39. Delp, S.L.; Anderson, F.C.; Arnold, A.S.; Loan, P.; Habib, A.; John, C.T.; Guendelman, E.; Thelen, D.G. OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 1940–1950. [[CrossRef](#)]
40. Wu, W.; Lee, P.V.S.; Bryant, A.L.; Galea, M.; Ackland, D.C. Subject-specific musculoskeletal modeling in the evaluation of shoulder muscle and joint function. *J. Biomech.* **2016**, *49*, 3626–3634. [[CrossRef](#)]

41. Sharma, R.; Dasgupta, A.; Cheng, R.; Mishra, C.; Vikranth, H.N. Machine Learning for Musculoskeletal Modeling of Upper Extremity. *IEEE Sens. J.* **2022**, *22*, 18684–18697. [[CrossRef](#)]
42. de Vries, W.H.K. Estimation of the mechanical loading of the shoulder joint in daily conditions. Ph.D. Thesis., Technical University Delft, Delft, The Netherlands, 2015.
43. Swain, M.J.; Ballard, D.H. Indexing via Color Histograms. In *Active Perception and Robot Vision*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 261–273.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.