Using Inertial Sensors to Estimate the Normal Force on the Castor Wheels during Handrim Wheelchair Propulsion In-Field

by

Louise Isabel Heringa

in partial fulfilment of the requirements for the degree of

Master of Science
in Biomedical Engineering
Track: Neuromusculoskeletal Biomechanics

at the Delft University of Technology,
Faculty of Mechanical, Maritime and Materials Engineering (3mE)

Student number:	5419875	
Date:	January 2023	
Supervisors:	MSc. M. P. van Dijk	TU Delft
	Prof. dr. H. E. J. Veeger	TU Delft
	Dr. M. J. M. Hoozemans	VU Amsterdam

T̃UDelft

Contents

Abstract

Goal The first aim of this master thesis is to explore whether the normal force on the castor wheels can be estimated from IMU data using a machine learning approach. The second aim is to evaluate whether incorporating the changing load distribution due to trunk movement could improve the friction power estimation compared with neglecting changes in load distribution.

Method Twenty-five subjects performed forward handrim wheelchair propulsions with no trunk, moderate and fast trunk movement on a treadmill in a wheelchair with six conditions regarding wheelchair mass and tire pressure. Two IMUs were placed on the trunk and wheelchair and two load pins in each castor wheel measured the normal force. After feature, model and hyperparameter selection, a model was trained to estimate the normal force on the castor wheels in percentage of the total weight from the IMU data. Accordingly, the predicted instantaneous normal force is used to calculate the friction power including changing mass distribution.

Results When using the linear velocity and acceleration of the wheelchair and the, linear acceleration of forward movement of the trunk, adequate estimations (MAE of 3.69% total weight) of the normal force from an LSTM model can be obtained for unseen subjects. This model is robust for wheelchair settings regarding wheelchair mass and tire pressure and for propulsions with no, moderate and fast trunk movement. The instantaneous friction power prediction incorporating the changing load distribution is proven to more accurate during propulsions with moderate and especially fast trunk movement.

Conclusion Coaches, sport scientists, and athletes may find this model useful for analysing the effect of different propulsion techniques or wheelchair conditions on the friction power. As part of a larger context, this research will contribute to the process of filling the technological gap of in-field monitoring mechanical power. Future research must validate the robustness of the model during game situations.

Acknowledgements

First of all, I would like to express my gratitude to Marit van Dijk, DirkJan Veeger and Marco Hoozemans for their assistance in completing this master's thesis. With her intelligent and enthusiastic view, Marit van Dijk always provided me the most high-quality feedback possible. With his abundance of knowledge and experience in the field of wheelchair biomechanics, DirkJan was always able to point me in the right direction when I was floundering in my swamp of data. With his brilliant statistic point of view, Marco could at one glace advice me on how to best represent my results.

Furthermore, I would like to thank Vera, Max, Ben and Marja. I was lucky to have had the most brilliant fellow thesis buddy, Vera. Besides that she could always answer my questions, she also became a close friend outside the university. My boyfriend Max, whom I could call for technical consult and emotional support. Along him, also his parents, Ben and Marja, a power couple that provided me with grammatical and data technical advise.

I highly appreciate my family and friends who were subjects of the experiments, you strengthened this study: Lara, Malou, Sonja, Maartje, Susanne, Anna, Tau and Celine. Also a special thanks, to my mother Sonja and my roommates in Rotterdam and Amsterdam, for their unconditional love during my ups and downs during my thesis process. Lastly, I am also grateful for the Monty Python voice of my father speaking to me when I got stuck during my research: "What is your quest?".

1 Introduction

One third of the Paralympic sports around are performed in a manual wheelchair [1], [2]. Over the last decade, the level of professionalism in wheelchair sports has been rising. Consequently, monitoring performance in wheelchair sports gained increased interest to optimize training programs [3], study the efficiency of different propulsion techniques [4] or evaluate new wheelchair conditions [5]. An important performance metric in cyclical endurance sports (e.g., cycling, rowing, running) is the mechanical power output, which is the external load that an athlete must provide to maintain a particular velocity over a given distance [5]. Athletes of various manual wheelchair disciplines, such as tennis, basketball, hockey, employ handrim propulsion to power the wheelchair. This movement is classified as an upper body cyclic task [6]. Therefore, for wheelchair sports using handrim wheelchair propulsion, mechanical power output can be an objective measurement for evaluating the athlete's performance. Yet, there is no non-invasive and low cost method available to monitor the mechanical power transferred by the athlete to the wheelchair during daily training and competition.

Mechanical power in sports can be estimated by solving the power equation defined according to *Van der Kruk et al.* [7] based on five terms: propulsion power generated by the athlete, kinetic power, frictional power, gravitational power and environmental power from external forces and moments, with the relationship described in Equation 1. In other words, the athlete needs to generate a certain propulsion power ($P_{propulsion}$) to overcome power losses due to restive forces ($P_{friction}$, $P_{gravitational}$ and $P_{environmental}$) resulting in movement of the athlete ($P_{kinetic}$) [8].

$$P_{propulsion} = P_{kinetic} + P_{friction} - P_{gravitational} - P_{environmental} \qquad (1)$$

The mechanical power generated by an athlete can either be estimated by estimating the propulsion power (left-hand-side of Equation 1) or through the sum of the kinetic power and power loss due to resistive forces (right-hand-side of Equation 1) [8]. For the left-hand-side method, force-instrumented push rims are available for accurate estimation of the direct mechanical power. However, these are not applied in practice, because they increase the total weight of the wheelchair with ~50–90%, which influences the wheelchair dynamics [8]. A promising non-invasive and low cost method [9] to estimate the power is via the right-hand-side by predicting the power terms from the data obtained with inertial measurement units (IMU) [8]. IMUs are wireless wearable electronic devices that contain an accelerometer, gyroscope and magnetometer, which measure 3-dimensional (3D) linear acceleration, angular velocity and strength and direction of the local magnetic field (i.e., orientation of the device) over time, respectively [10], [11]. Currently, these sensors are used to monitor velocity, acceleration, covered distance and rotations of the wheelchair [12]. A study by *Uddin et al.* [13] showed that IMUs in combination with machine learning have also the potential to assess mechanical power output during country skiing.

This research will focus on estimating the friction power ($P_{friction}$) from IMU data. During the most popular Paralympic wheelchair sports, wheelchair basketball, tennis and rugby [3], this is power loss is mainly due to rolling (and internal wheelchair) resistance. Given that these sports are performed at low velocities, the athlete and wheelchair experience negligible air resistance. Equations 2 and 3 describe the power loss due to rolling resistance, where F_{normal} is the total vertical force the surface exerts on the rear (back) and castor (front) wheels, λ is the rolling resistance parameter, r is the wheel radius and v is the velocity of the wheelchair [14]. The coefficient of rolling resistance (μ = λ/r [no unit]) is affected by wheel conditions (e.g., tire type and pressure) and surface conditions (e.g., gym court, treadmill).

$$P_{friction} = F_{rolling} \cdot v_{wheelchair} \qquad (2)$$

$$F_{rolling} = (\lambda_{rear} / r_{rear}) \cdot F_{normal, rear} + (\lambda_{castor} / r_{castor}) \cdot F_{normal, castor} \qquad (3)$$

The rear and castor wheels of a court wheelchair have different rolling resistance coefficients (μ) due to its substantial difference in radius [14]. For example, in the study by *Sauret et al.* [14], the castor wheels had a rolling resistant coefficient that is nearly three times higher than the rear wheels. Moreover, during a handrim propulsion cycle, the athlete moves its trunk to guide the push action of the hands on the rim. For example, the same study showed that the measured load on the castor wheels varied from 30% to 60% during one propulsion cycle [14]. Consequently, when the trunk moves forward, the normal force on the castor wheels with high rolling coefficient will increase, leading to an increased total rolling resistance force

and resulting friction power [14]. Therefore, to accurately estimate the power loss due to rolling resistance, knowledge on the real-time normal force distribution on the rear and castor wheels is required.

As mentioned above, an non-invasive and low cost solution to estimate the distribution of the normal force in-field is by using IMUs. Therefore, the load distribution change during propulsion can be captured by measuring angular velocity and linear acceleration of the trunk with IMUs [15]. Subsequently, this kinematic data can be used to obtain the normal force distribution by either a biomechanical modelling or machine learning approach. Biomechanical modelling is a technique that makes predictions by being explicitly programmed. Machine learning is a technique that uses a certain algorithm to learn from data and make predictions about events in the real world without being explicitly programmed. An advantage of machine learning over biomechanical modelling is that it can determine the relation between input and output within minutes using a simple programmable algorithm. A disadvantage is that it requires a relatively large dataset to determine the input-output relationship with reasonable accuracy. A recent comparative study on biomechanical modelling and machine learning with artificial neural networks yielded statistically comparable predictions of ground reaction forces during running and walking using IMU data [16]. Currently, there is no machine learning model in the scientific literature that examines the effect of the wheelchair user's actions on the normal force distribution. Therefore, this research will explore the possibility of estimating the normal force distribution on the castor wheels (output) from IMU data of the wheelchair and trunk (input) in-field using a machine learning approach.

To explore the validity of this method, the predicted normal force distribution from IMUs should be compared to a reference force measurement. During wheelchair propulsion, the total normal force calculation could be simplified to the total mass multiplied by the gravitational force equivalent ($F_{normal,total} = m \cdot g$), disregarding the vertical acceleration of the wheelchair-user system. When the total normal force is known, either the normal force on the rear or the castor wheels could be used to calculate the other. In this research, two load pins are placed in the axis of the castor wheel, because these have the highest rolling resistance coefficient and, therefore have the biggest impact on friction power.

The first aim of this master thesis is to explore whether the normal force on the castor wheels can be estimated from IMU data using a machine learning approach. Accordingly, with the obtained data, friction power can be calculated by filling in Equation 2. The second aim is to evaluate whether incorporating the changing load distribution due to trunk movement could improve the friction power estimation compared with neglecting changes in load distribution. Therefore, this research will contribute to the process of filling the technological gap of in-field monitoring mechanical power via the right-hand-side method during wheelchair sports.

2 Materials & Methods

2.1 Study design

2.1.1 Study population

Twenty-five healthy non-wheelchair bound subjects (19 females, age: 30 ± 11 years, body height: 170 ± 7 cm, body mass: 68 ± 11 kg) performed forward handrim wheelchair propulsions on a treadmill in an all-court sports wheelchair. Prior to the experiment, the participants were informed about the aim and procedure of the study and provided written informed consent. The experiment was approved by the ethical committee of the Technical University of Delft.

2.1.2 Experiment procedure

The subjects performed forward handrim wheelchair propulsions on a treadmill for 6 blocks of 3.5 minutes each. The wheelchair condition was changed for each block by either placing additional mass under the wheelchair seat or lowering the tire pressure of the rear wheels. This was to examine whether a machine learning model can predict the normal force on the castor wheels in different wheelchair conditions. Moreover, the addition of mass on the rear axle can also imitate a situation in which the user has a high ratio upper body mass to total mass, such as when missing a leg. Blocks 1 and 4 were performed in a wheelchair with original conditions. In blocks 2 and 3, mass was added (+5kg and +15 kg) to the initial wheelchair mass of 13.17 kg. In block 5 and 6, pressure was reduced (-1.75 and -3.5 bar) with respect to the initial tire pressure of 5.25 bar. The experiments of subjects 1-10 occurred in the sequence of first block 1-3 (i.e., increasing mass wheelchair) followed by block 4-6 (i.e., decreasing tire pressure). The experiments of subjects 11-25 occurred in opposite sequence to mitigate learning behaviour. Table 1 provides an overview of the wheelchair conditions per block.

During each block, the subjects successively performed three different propulsion types: moderate, no and fast trunk movement. This was to examine whether the machine learning model could predict the normal force on the castors when the subject uses a different intensity in trunk movement. No trunk movement can occur in circumstances where the user cannot use their trunk, for example due to a spinal cord injury. Moreover, even wheelchair athletes who are able to move their trunk, do not always use it. Moreover, the subjects were guided by attention signals to identify the start of a new propulsion type. Two measures were taken to allow the participants to control the forward propulsion in a linear trajectory. First, the duration of the first propulsion type was 90 seconds compared to 60 seconds for the second and third to familiarize the subjects with the propulsion. Secondly, the subjects were instructed to start the push action of the hands on the rim in the same rhythm of a metronome. During the first propulsion type, called moderate trunk movement, the subjects performed wheelchair propulsion with trunk inclination at a treadmill rolling 1.2 m/s and the metronome was set at 27 bpm. During the second propulsion type, called no trunk movement, the subjects performed the propulsion at the same speed without trunk inclination. The metronome was increased to 40 bpm, because the trunk remained in straight position and therefore could not guide the push action. The subjects had to perform more less-powerful push actions to maintain the same speed as during propulsion type 1. During the last propulsion type, called fast movement, the subjects performed the propulsion at a speed of 1.7 m/s with trunk inclination again. The metronome remained at 40 bpm, because, despite trunk inclination, the subjects had to perform the push action more frequent to maintain a higher speed. To allow participants to predict the next propulsion speed and trunk tilt, the sequencing of the propulsion types was not randomized between blocks. Table 2 provides an overview of the three wheelchair propulsion types performed during each block.

Prior to performing propulsion on the treadmill, the subjects were instructed to perform some tasks a force plate from a stand still for approximately 2 minutes to calibrate the load pins (discussed in Appendix A). The rear wheels were fixed to a dummy plate and the castor wheels were placed on a force plate. The subjects were instructed to sit in the wheelchair in straight position and subsequently, to perform simulated propulsion motions with no trunk, moderate and fast trunk movement. Furthermore, drag tests on the treadmill were performed after block 4,5 and 6 to determine the rolling resistance coefficients of the rear wheel for each tire pressure and the coefficient of the castor wheels. The treadmill maintained a constant velocity and the wheelchair was attached to a force sensor. One drag tests consisted of six sets of tests carried out in succession. The subjects were instructed to alternatively sit straight in the wheelchair and

bend forwards for approximately 20 seconds when no and an additional mass was placed in the subject's lap and on the footrest. These six sets simulate conditions in which the mass distribution on the wheels alters. The total drag force was measured by the force sensor, the force on the castor wheels by the load pins and the total normal force by mass times the gravitational force equivalent. Finally, rolling resistance coefficients were computed using the least square linear regression.

Before any data collection commenced, the participants without wheelchair experience were instructed to perform one test block to familiarize themselves with wheelchair propulsion on a treadmill. The experiment was performed in January to March 2022 in the human motion lab at the VU Amsterdam.

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
Wheelchair condition	original	+5 kg	+15 kg	original	-1.75 bar	-3.5 bar
Tire pressure (bar)	5.25	5.25	5.25	5.25	3.50	1.75
Wheelchair mass (kg)	13	18	28	13	13	13

Table 1. The wheelchair conditions per block.

	Propulsion type 1	Propulsion type 2	Propulsion type 3
Trunk movement	Moderate	No	Fast
Time	90 sec	60 sec	60 sec
Treadmill velocity	1.2 m/s	1.2 m/s	1.7 m/s
Trunk inclination	Yes	No	Yes
Propulsion frequency	27 bpm	40 bpm	40 bpm

Table 2. The propulsion task description per propulsion types in immediate succession during one block.

2.2 Data acquisition

The inertial data (3D acceleration and 3D angular velocity) of the wheelchair were measured by an IMU (NGIMU, X-IO Technologies, Colorado Springs, CO, United States) located at right axis of the rear wheel ($IMU_{wheelchair}$). The y-axis of the $IMU_{wheelchair}$ was aligned with the wheelchair axis and the wheel rotation was in the direction of the x-axis. The inertial data of the trunk were measured by an IMU mounted at the xiphoid process using a strap (IMU_{trunk}). The IMU_{trunk} is aligned with the trunk: sideways with the x-axis, up and down by y-axis and, for- and backwards by z-axis of the IMU_{trunk}. These two IMUs had a sample frequency of 100 Hz and a range of ±16 g and 2000 deg/sec. Simultaneously, the total vertical force (i.e., normal force) on both castor wheels were measured by two custom made load pins (MB1787, Batarow Sensorik GmbH, Karow, Germany). These load pins contain strain gauges with voltage as sensor output, which is proportional to the normal force. The voltage output was sent to a third IMU located at the frame of the wheelchair ($IMU_{loadpins}$). Appendix A.1 contains additional information about the operation regarding the load pin. The sample frequency of $IMU_{loadpins}$ was 50 Hz, because at a higher sample frequency the battery did not sustain a full experiment. A preliminary study showed that this sample frequency was able to capture the dynamics of the forces on the castor wheels. Hence, $IMU_{wheelchair}$ and IMU_{trunk} measured the inertia and the $IMU_{loadpins}$ captured the voltage of the load pins. Figure 1 displays the location of the three IMUs and two load pins. The data from the three IMUs were collected via Wi-Fi which enabled the sensors to be automatically time synchronized. Furthermore, a custom made 1x1 meter force plate was used for calibration of the load pins and determining the subject and wheelchair mass.
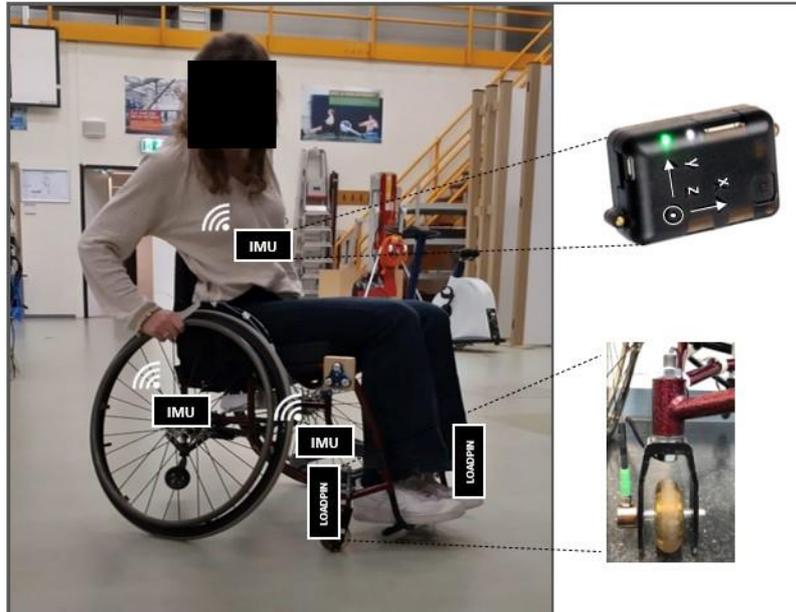
Figure 1. Overview of the sensors setup: IMU_{trunk} was located at the xiphoid of the subject, the $IMU_{wheelchair}$ at the right rear wheel, the $IMU_{loadpins}$ at the frame of the wheelchair underneath the seating area. The latter IMU was connected by wires to the two load pins, which were placed in the right and left castor wheel axis.

2.3 Data pre-processing

2.3.1 General data analysis

The data of the three IMUs was acquired in CVS format and imported into MATLAB (R2021b, The MathWorks Inc., Portola Valley, CA, United States). The sample frequency deviations from the IMUs were corrected using linear interpolation and resampled to 50Hz.

2.3.2 Predictor and target features

The trunk inclination during wheelchair propulsion leads to a change of normal force on the castor wheels. At the start of the propulsion cycle, the load distribution on the front wheels is at its lowest value, because the trunk is in upright position. When the trunk is moving forward during the push action, the load on the castor wheels increases. Moreover, due to the push action on the rear wheel, the wheelchair will accelerate in the forward direction. After the push action, the trunk is tilted backwards, leading to a decrease in the load on the castor wheels and at the same time the wheelchair will decelerate. From the data of the IMU_{trunk} and $IMU_{wheelchair}$, features were extracted that represent the change in mass distribution due to trunk inclination and wheelchair movement caused by a push action, respectively. These features are called the predictor features, because they can potentially predict the target feature, namely the normal force on the castor wheels. Figure 2 gives a schematic overview of the predictor and target features and were extracted from the IMU data as followed:

- *Predictor 1 – Linear velocity wheelchair (m/s).* This is calculated by Equation 4, where ω is the angular velocity from the y-axis of $IMU_{wheelchair}$ and $\varnothing_{rear}$ is the rear wheel diameter.

$$v_{wheelchair} = (\omega_{wheelchair} \cdot \varnothing_{rear} \cdot \pi) / 360 \tag{4}$$

- *Predictor 2 – Linear acceleration wheelchair (m/s²).* This is the derivative of predictor 1 and calculated as follows: the difference between the adjacent elements of the linear velocity of the wheelchair (predictor 1) is calculated and multiplied by the sample frequency. During this operation, a half sample shift forwards occurred and to correct this shift, spline interpolation was performed.
- *Predictor 3 – Angular velocity trunk inclination (deg/s).* This is the angular velocity from the x-axis of IMU_{trunk}.
- *Predictor 4 - Angular acceleration trunk inclination (deg/s²).* This is the derivative of angular velocity trunk inclination (predictor 3), using the same calculation as in predictor 2.
- *Predictor 5 - Linear acceleration for- and backward trunk (m/s²).* This is the linear acceleration from the z-axis of IMU_{trunk} times the gravitational force equivalent (g), because IMU gives the acceleration in g by default.

9

- *Predictor 6 - Linear acceleration up- and downward trunk (m/s²).* This is the linear acceleration from the y-axis of IMU_{trunk} times the gravitational force equivalent.
- *Predictor 7 - Resulting linear acceleration trunk (m/s²).* The resulting magnitude in acceleration is calculated by the three acceleration vectors using the Pythagoras theorem.
- *Target feature – Normal force on castor wheels (% total weight).* First, the voltage output of the left and right load pin from the $IMU_{loadpins}$ was added. Secondly, the voltage output was filled in the pre-defined polynomial (see Appendix A.1.2) to convert the force output to Newton. Subsequently, the normal force was divided by the subject and wheelchair weight (including the additional weight on the rear axle during block 2 and 3) to estimate the normal force of other subjects with different body or wheelchair mass. After the prediction of the force in percentage of the total weight, the force in Newton can be calculated by multiplying the value with the subject and wheelchair weight.
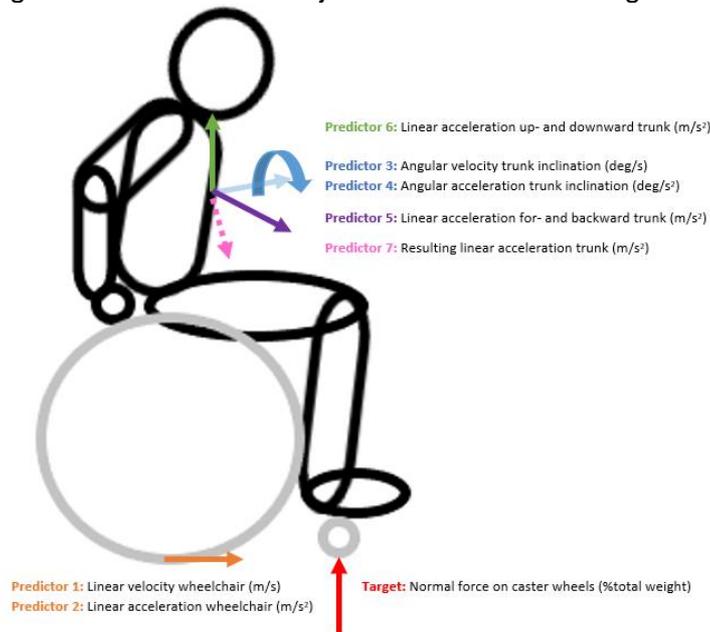

Figure 2. Overview of the predictor features and target feature with corresponding vector.

2.3.3 Filtering and segmentation
All the predictor features and target features were filtered with a second order low-pass Butterworth frequency filter with a cut-off frequency of 3 Hz. This cut-off frequency filters out the noise of the irregularities of the treadmill surface and while preserving the curve of the signal. This value was determined from the spectral density analysis of the normal force on the castor wheels from five randomly selected subjects during propulsions in a wheelchair without added mass or lower tire pressure. Figure 3a shows the power spectrum of a representative subject in a wheelchair condition without added mass or lowered tire pressure. Subject 18 was selected to be a representative, because the normal force on the castor wheels in sitting straight position was similar to the average force across all subjects. The power peaks at 0.44 and 0.66 Hz indicate the propulsion frequency during moderate, no and fast trunk movement, respectively. Figure 3b shows an example of the normal force on the castor wheels with and without filter during one propulsion cycle of representative subject with moderate trunk movement. This example shows that the filter can remove the noise while preserving the curve of the signal. During one propulsion cycle, the wheelchair–athlete combination accelerates by exerting force on the handrim (push phase), followed by a deceleration phase where the arms are brought back to a position so a new push phase can begin (recovery phase). The push phase was defined as local minimum of the wheelchair acceleration until the local maximum. Subsequently, the recovery phase was defined as the local maximum to the next local minimum. Therefore, one propulsion cycle was defined as a local minimum to the next local minimum in wheelchair acceleration.

Thereafter, the data of each block was manually segmented in such a manner to contain 60 seconds of each propulsion type, including the few seconds needed to transfer from type. As a result, the data of each block had a total sample duration of 180 seconds. Finally, the segmented data was concatenated into a single 10-column CVS database file including seven predictor features, the target feature, subject number and block number.
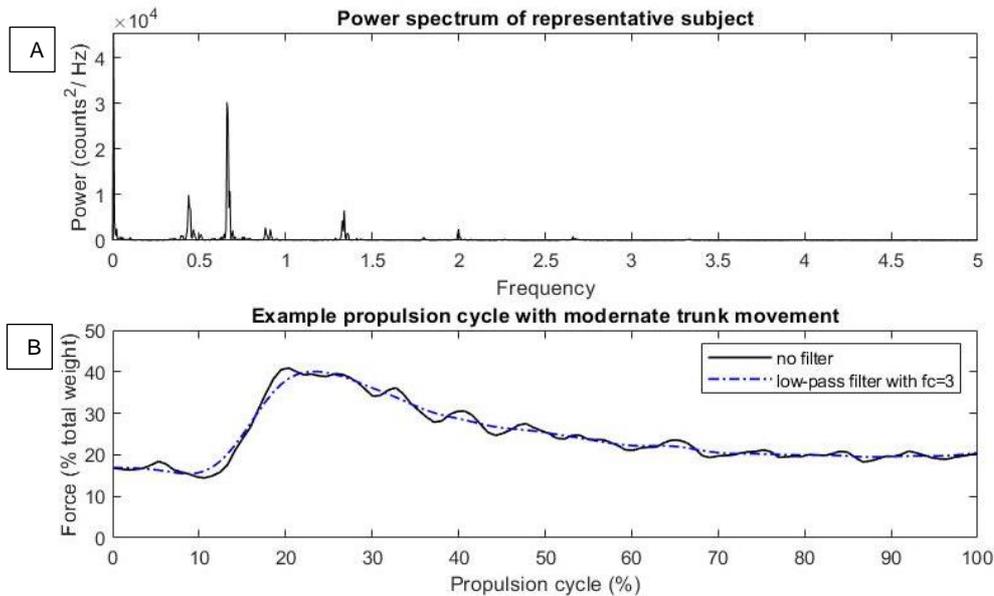
Figure 3. The power spectral density graph of a representative subject with normal wheelchair conditions shows the frequencies variations of the normal force on the castor wheels signal (A) and based on this graph the a filter with cut-off frequency of 3 Hz was chosen. An example of one propulsion cycle shows the normal force with and without filter applied (B).

2.4 Machine learning

After pre-processing, the database was imported into Python (version 3.9, Python Software Foundation, Wilmington, DE, United States). The data was used to build and evaluate the machine learning model with the following approach. First the data set was split into a training, validation and test set. Secondly, a feature selection method was applied to select the best combination set of the seven predictors features. Thirdly, the optimal machine learning model was selected. Followed by choosing the best hyperparameters of this model. Lastly, the final model with the optimal machine learning algorithm and hyperparameters was trained based on the training set with the optimal feature set and was evaluated on an unseen test set. Figure 4 provides an overview of each step of this machine learning approach and each will be discussed in the following sections.
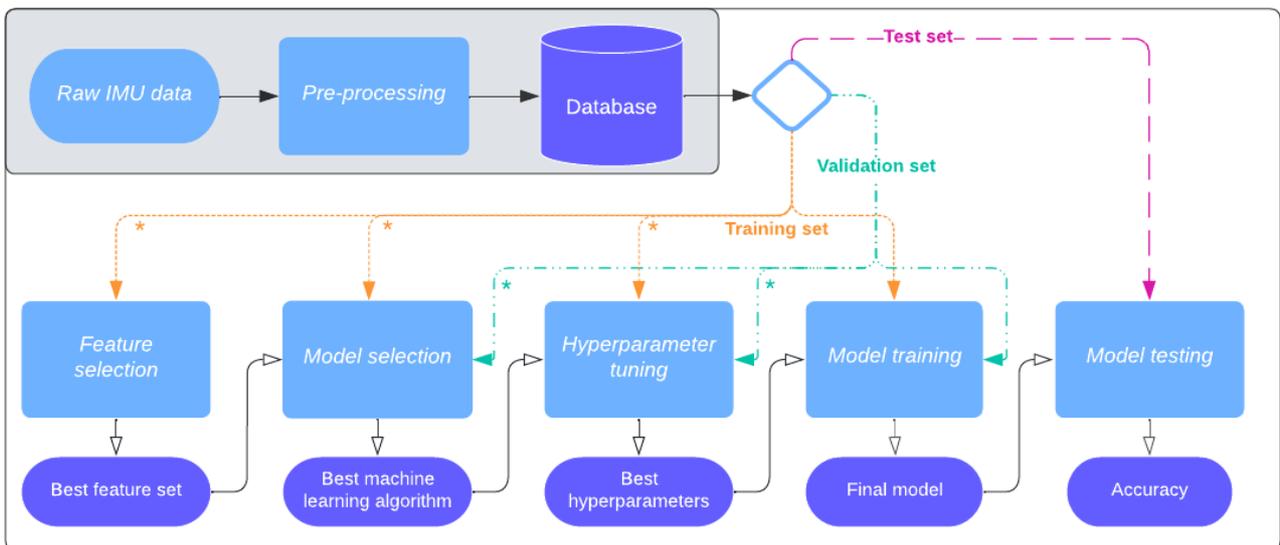

Figure 4. Overview of the machine learning approach to predict the normal force on the castor wheels. The asterisk indicates that only data from experiments in original wheelchair condition are used. The grey box corresponds to Section 2.3 Data pre-processing performed in MATLAB. The white box corresponds to Section 2.4 Machine learning performed in Python.

2.4.1 Training, validation, test split

The dataset containing 150 subsets (25 subjects in 6 wheelchair condition) was divided into a training, a validation and a test set. First, the complete data set of three randomly selected subjects were set apart. This subject test set, was to examine whether a machine learning model can predict the normal force on the castor wheels for unseen users. Secondly, the data with wheelchair conditions +5kg and -1.75 bar were set apart. This condition test set, was to examine whether a machine learning model can predict the normal force for unseen wheelchair conditions. These two test sets were used for the final model evaluation and remained unseen during the feature selection, model selection, hyperparameter tuning and model training. From the remaining subsets 80% subsets were randomly designated to the training set and 20% to the validation set. For visualization, an example of the training, validation and test split is shown in Figure 5.
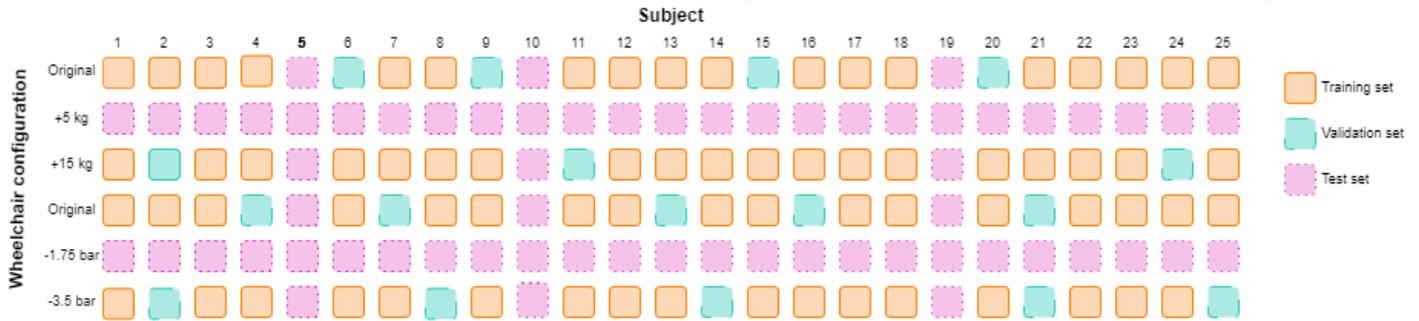

Figure 5. Example of a training, validation test split when all the experiments contain a complete dataset.

2.4.2 Feature selection

Feature selection was applied to select the best set of the seven predictors features to be used for training the machine learning model. By discarding irrelevant features, the computational time for model training will decrease and the model performance may increase by mitigating overfitting, as the model will not pick up on spurious correlations of redundant features. In this study, an exhaustive feature selection method was used to ensure the best combination of the seven features. Since this is a brute-force evaluation, to save computational time the training data of experiments in original wheelchair condition were used (the asterisk in Figure 4). It is assumed that the relationship between the predictor features and the normal force on the castor wheels does not change when the wheelchair has different settings in terms of mass and tire pressure.

The best set of features was selected by optimizing a performance metric given an arbitrary regressor. In this study, the performance metric used was the mean absolute error and the regressor used a random forest algorithm. The random forest regressor (RFR) was used because of its ability to be prone to overfitting [17]. A RFR operates by constructing several decision trees during training. Each tree is created from a different part of the training dataset and at each node a different feature is selected for splitting. Each of the trees makes its individual prediction and afterwards these predictions are averaged to a single result. This makes the model quite robust as the uncorrelated trees protect each other from their individual errors. In this study, a RFR algorithm [18] with 50 trees was trained on all the 126 possible feature combinations. The mean absolute error (MAE) was used as performance accuracy metric, because from interpretation standpoint is easy to understand and the size of the error influences the metric. Other metrics have been considered but discarded: root mean square error (RMSE), in which the errors are squared before they are averaged, gives a higher weight to large errors. This makes the value of the RMSE more difficult to interpret. Moreover, the coefficient of determination (R2) was not used as a metric, because it does not reflect the size of the error, solely its correlation between measured and predicted. Table 3 provides an overview of the formulas.

The optimal feature set was selected based on the elbow method. This method consists of plotting the lowest MAE of the possible feature combinations as a function of number of features and choosing the best combination of features at the (elbow) point were the MAE stagnates. Generally, increasing number of features will improve the prediction on the training set, but that at some point overfitting may occur. To prevent the model from overfitting, the combination of features is chosen where the model accuracy does not improve while adding more features, this is reflected by the elbow point.

12

Performance metric	Formula	Description
Mean error	$$ME = \frac{1}{N}\sum_{t=0}^{N} \hat{y}(t) - y(t)$$	Measures the *average* of the difference between the measured and predicted data.
Mean absolute error	$$MAE = \frac{1}{N}\sum_{t=0}^{N} |\hat{y}(t) - y(t)|$$	Measures the *average* of the absolute difference between the measured and predicted data.
Root mean squared error	$$RMSE = \sqrt{\frac{1}{N}\sum_{t=0}^{N} (\hat{y}(t) - y(t))^2}$$	Measures the *standard deviation* between the measured and predicted dataset.
Coefficient of determination	$$R2 = 1 - \frac{\sum_{t=0}^{N} (y(t) - \hat{y}(t))^2}{\sum_{t=0}^{N} (y(t) - \bar{y})^2}$$	Indicates the proportion of the variation of the predicted value that is predictable from the measured target value. It is a number assumed to be between 0 and 1. However, when the regression is worse than the mean, the R2 will be negative.

Table 3. Description of performance metrics used [19], where ŷ(t) represents the predicted and y(t) the measured normal force on the castor wheels and N the sample size.

2.4.3 Model selection

To determine which machine learning model was most suitable, five different regression algorithms were trained: a linear regression (LR), random forest regressor (RFR), multiple layer perceptron (MLP), long short-term Memory (LSTM) and gated recurrent unit (GRU). A LR was chosen to explore whether the relationship between the predictor and target features are linearly related. A RFR predicts the target feature by averaging the predictions produced by the trees. This algorithm was chosen, because it is considered to be robust as the uncorrelated trees protect each other from their individual errors. A MLP is a feedforward neural network that consist of one input layer, one or more hidden layers and one output layer with interconnected neurons. These connections called weights and are selected during training by optimizing the error between the input and output feature using a stochastic gradient descent during one iteration [18]. The learning rate determines the step size at each iteration while moving toward the minimum of a loss function. The batch size determines the number of samples for each iteration. This MLP algorithm was chosen, because in literature it has proven to predict the ground reaction forces during running and walking from IMU data [20], [21], [22]. An LSTM and GRU are recurrent neural networks, additional to the MLP model their neurons contain memory gates which remember a certain amount of previous time samples. These models predict the output based on pervious input and current input predict [23]. Therefore, they are mostly applied to temporal problems, such as language translation and speech recognition. The prediction of the normal force during propulsion can also be seen as a temporal problem as it follows a cycle. Moreover, in literature it has proven to predict the ground reaction forces during running and walking from IMU data using LSTM [10], [24] and GRU [25]. After the hidden layer, these studies added a drop out layer to improve generalization by randomly dropping a specified rate of neurons during model training [26]. Hence, these dropped neurons had no effect on the activation of the neuron in the output layer [24]. The five models consisted of a simple architecture with similar hyperparameters and are presented in Table 4. These were chosen based on default hyperparameters of the 'sklearn' package in Python [18] and literature estimating the ground reaction force from IMU data.

The five models were trained on the training set with the best feature selection set and evaluated on the validation set containing again only with original wheelchair condition setting. The model with the average lowest MAE over all the validation sets was selected as best machine learning algorithm. As discussed in 2.4.2 Feature selection, this metric was chosen, because from an interpretation standpoint it easy to understand how the size of the error influences the metric. Moreover, the other accuracy performance metrics in Table 3 are also calculated to support the model selection and to prevent ambiguity between other (future) studies using different metrics. The mean error (ME) was used to determine whether the model is over- or underestimating the force.

LR	RFR	MLP	LSTM	GRU
No parameters	Number of trees: 50 Maximal number of splits: None Minimal samples per split: 2	Number hidden layers: 1 Number neurons per layer: 50 Learning rate: 0.01 Batch size: 256	Number hidden layers: 1 Number neurons per layer: 50 Learning rate: 0.01 Batch size: 256 Drop out rate: 0.2 Time steps: 20	Number hidden layers: 1 Number neurons per layer: 50 Learning rate: 0.01 Batch size: 256 Drop out rate: 0.2 Time steps: 20

Table 4. The default hyperparameters per machine learning model.

2.4.4 Hyperparameter tuning

After the best model was selected, candidate values for each hyperparameter in Table 5 were defined. Then the model was trained on each possible combination of hyperparameters and tested on the validation set with original wheelchair condition setting. The validation set was divided into subsets per subject per block. Thus the trained models with a unique hyperparameter combination were evaluated over these validation sets. When the grid search was performed, the best hyperparameter combination was again chosen based on the average lowest MAE.

2.4.5 Final model training and testing

For the last step, the final model with the optimal machine learning algorithm and hyperparameter combination was built based on the training set with the optimal feature set in multiple epochs. During one epoch, a neural network is trained on all the training data by one forward pass (calculating target feature from predictor features) followed by one backward pass (adjusting the weights to minimize the loss function). In this case, the data was divided into batches of numerous samples, so one epoch comprised of multiple iterations. The selected loss function was the mean squared error between the measured and predicted normal force on the castor wheels (note that this is not an evaluation metric as discussed in 2.3.3 Model selection). After each epoch, the minimum loss on the training and validation set was calculated to monitor overfitting of the model. Overfitting occurs if the model has learned the training dataset too well, including the statistical noise or random fluctuations in the training dataset. Therefore, the model becomes specialized to the training data and will be less able to generalize to new data, resulting in an increase in validation loss. A learning curve plot shows overfitting if the training loss continues to decrease per epoch and the validation loss decreases to a certain point after which it increases again. At this latter point of inflection, the training should be stopped to prevent model overfitting. Therefore, if the minimum loss on the validation set was improved compared to previous epoch, the model's current weights were saved. If the loss did not improve after five epochs, the model training was stopped. For the final model, the weights with the lowest loss in the validation set were used.

This model was tested on a test set containing three unseen subjects and on a test set two unseen wheelchair conditions. These two sets were to examine whether a machine learning model can predict the normal force on the castor wheels for unseen users or wheelchair conditions. Moreover, the performance accuracy per test subject per propulsion technique regarding no, moderate and fast trunk movement during original wheelchair condition was calculated. This was to examine whether the machine learning model could predict one sort of trunk movement better than another.

To examine whether the architecture of this model is prone to overfitting, a new model was built with a Gaussian noise layer added to original architecture. As it is a regularization layer, it is only active during training [26]. After training, the MAE per test subject during original wheelchair condition for this new model was calculated. If the MAE of this model is approximately the same as the original model, it means that the accuracy of this model architecture is not affected by white input noise. This makes the model robust as in the real world, input data from sensors will always contain some noise.

2.5 Friction power

This study aimed to explore whether IMUs could be utilized to estimate the normal force on the castor wheels during handrim wheelchair propulsion in-field. As part of a larger context, the normal force prediction could be a valuable tool for measuring the friction power. The total friction power during the experiments was calculated according to the Equations 3 and 4, described in 1. Introduction. The rolling resistance coefficients (μ) per wheel were determined by drag tests on the treadmill. The total normal force calculation was simplified to the total mass times the gravitational force equivalent, excluding the vertical acceleration of the centre of mass [14]. When subtracting the normal force of the castor wheels from the total normal force, the force the rear wheels was calculated. The total rolling resistance force was calculated with the obtained values according to Equation 3. Subsequently, the total friction power can be estimated by multiplying the rolling resistance force with the velocity of the wheelchair obtained from the $IMU_{wheelchair}$ data according to Equation 2.

In a simplified model, friction power can be estimated without the typical normal force variation resulting from trunk movement. The rolling resistance force was calculated according to Equation 5, where $\overline{X}_{castor}$ is normal force on the castor wheels in percentage of the total weight while the subject is sitting still in straight position.

$$F_{rolling} = (\mu_{rear} \cdot (1 - \overline{X}_{castor}) + \mu_{castor} \cdot \overline{X}_{castor}) \cdot F_{normal, total} \tag{5}$$

The friction power estimated from the measured normal force by the load pins was used as reference. An overview of the three different friction power calculations is given in Table 5. Note that the rolling resistance for the reference and the method including the changing mass distribution had a variable rolling resistance force and the excluding changing distribution method a constant force.

The instantaneous friction power during original wheelchair condition was estimated based on the including mass distribution method (incorporating the effect of the trunk inclination) and based on excluding the distribution (neglecting the effect of trunk inclination). Accordingly, the performance accuracy based on the reference friction was calculated per propulsion type to examine whether the prediction model will be valuable in predicting the friction power, especially when no trunk movement is applied.

	Reference	Incl. changing mass distribution	Excl. changing mass distribution
Normal forces	$F_{normal,castor}$ via load pin data $F_{normal,rear} = m \cdot g - F_{normal,castor}$	$F_{normal,castor}$ via IMU data $F_{normal,rear} = m \cdot g - F_{normal,castor}$	$\overline{X}_{castor}$ via force plate data $F_{normal,total} = m \cdot g$
Rolling resistance coefficients	μ_{rear} and μ_{castor} via drag tests		
Rolling resistance force	$F_{rolling} = \mu_{rear} \cdot F_{normal,rear} + \mu_{castor} \cdot F_{normal,castor}$		$F_{rolling} = (\mu_{rear} \cdot (1 - \overline{X}_{castor}) + \mu_{castor} \cdot \overline{X}_{castor}) \cdot F_{normal,total}$
Wheelchair velocity	$v_{wheelchair}$ via IMU data		
Friction power	$P_{friction} = F_{rolling} \cdot v_{wheelchair}$		

Table 5. Overview of the three different friction power calculations, where the rolling resistance for the reference and including changing mass distribution method has a variable rolling resistance force and the excluding changing mass distribution method a constant force.

3 Results

After visual inspection of the IMU data, it was recognised that 7 of the 150 experimental subsets contained incomplete data. This was due to empty batteries or due to the sensors not being activated. This resulted in a 66-17-60 subset training-validation-test split, or 46%-12%-42%. Figure 6a shows that there is no clear pattern in mean and distribution of the average normal force of each subject when adding mass or lowering tire pressure. This not as expected, as one would assume that normal force on the castor wheels in percentage total weight would decrease when mass is added on the rear wheel axis. Figure 6b shows that the propulsion technique with no trunk movement had a lower mean force compared to moderate and fast movement. This makes sense given the trunk remains in upright position and does not exert force on the castor wheels. Figure 6c shows that each of the three propulsion technique regarding trunk movement had a different pattern during one propulsion cycle.
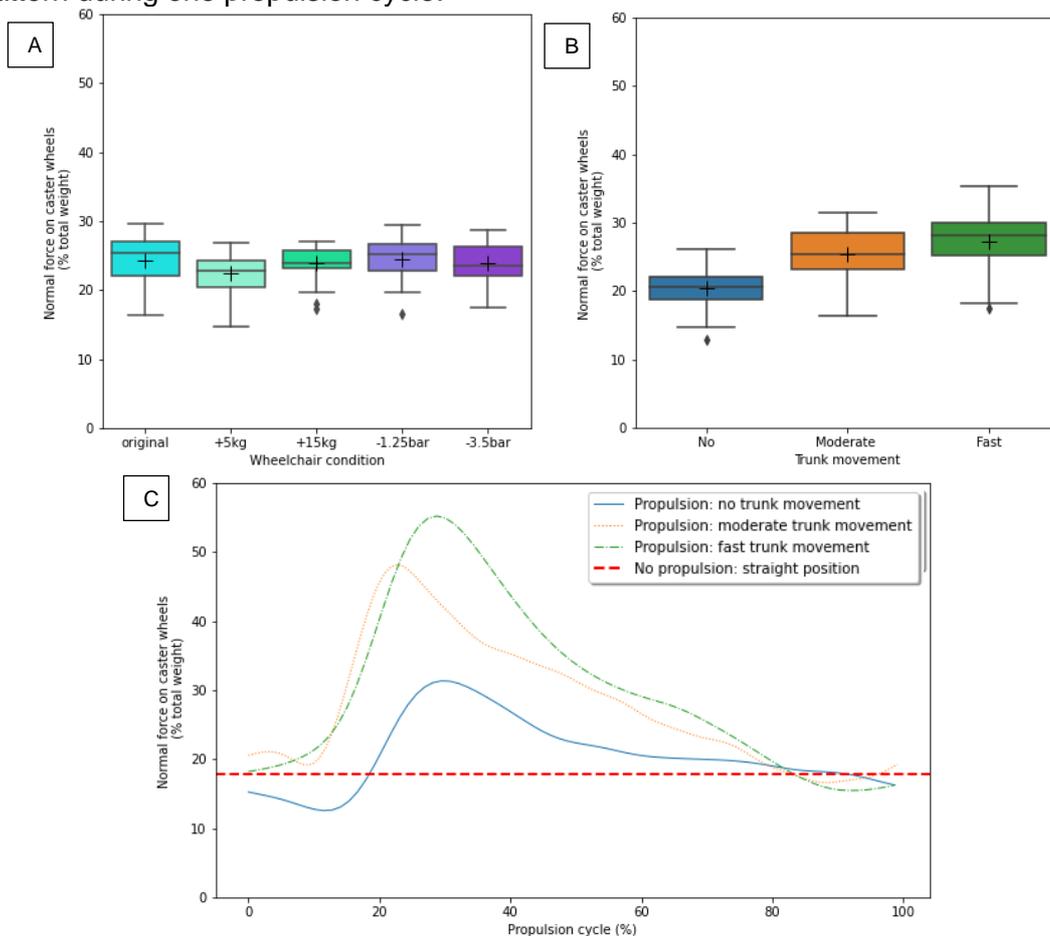

Figure 6. The box shows the interquartile range (IQR) of normal force on the castor wheels distribution and the whiskers show the rest of the distribution. The plus sign (+) indicates the mean normal force per subject and line in the box the median. The diamond (♦) indicates an outliers located outside the lower (Q1-1.5*IQR) and upper (Q3+1.5*IQR) limit. Figure A shows the distribution of the average normal force of each subject per wheelchair condition. For the original wheelchair condition, the data from both block 1 and 4 had been used. Figure B shows the distribution of the average normal force of each subject in original wheelchair condition setting per propulsion technique regarding trunk movement. Figure C shows an example of a normal force curve of the representative subject in original wheelchair condition during one propulsion cycle for each of the three trunk movement types. The red dotted line indicates the normal force on the castor wheels in original wheelchair conditions during no propulsion in straight position for the representative subject.

3.1 Feature selection

Figure 7 and Table 6 visualize the results of the exhaustive feature selection from the random forest regressor with a 7-fold cross validation (i.e., leave 5 of the 35 total training subsets out). The first seven feature combinations used only one feature and the last combination used all seven features. The MAE curve stagnates at the combination of three features: linear velocity and acceleration of the wheelchair and, linear acceleration of for- and backward of the trunk. According to the elbow-method, this combination was chosen as best combination of predictor features.
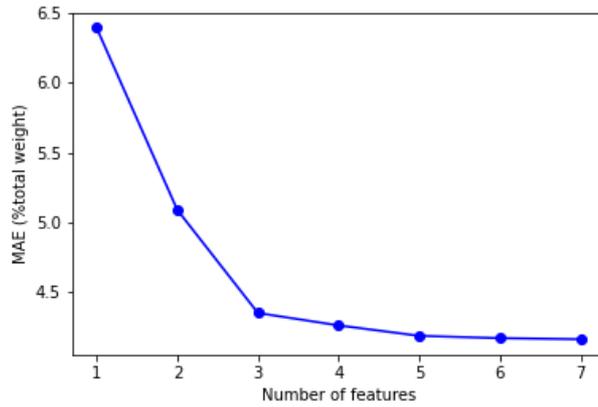
16

Figure 7. Results of the exhaustive feature selection, representing seven feature combinations with the lowest MAE between the predicted and measured normal force on the castor wheels from a random forest regression with 7-fold cross validation.

Number of features	1	2	3	4	5	6	7
Predictors	5	1,5	1,2,5	1,2,5,6	1,2,3,5,6	1,2,3,5,6,7	1,2,3,4,5,6,7
MAE (%total weight)	6.40 ±0.80	5.09 ±0.84	4.34 ±0.71	4.26 ±0.69	4.18 ±0.63	4.16 ±0.65	4.16 ±0.62

Table 6. Results of best feature combinations from the exhaustive feature selection.

3.2 Model selection

Figure 8 and Table 7 present the results of the five different machine learning algorithms. The LSTM had on average the lowest MAE over all the four models. Therefore, this model will be used as prediction model. To support this decision, the LSTM also had the lowest RMSE and highest R2 score.
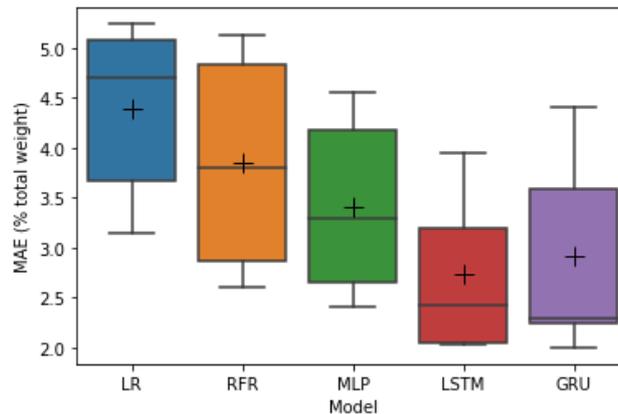

Figure 8. The box shows the interquartile range (IQR) and the whiskers show the rest of the normal force on the castor wheels distribution. The plus sign (+) indicates the mean normal force per subject and line in the box the median. The diamond (♦) indicates an outliers located outside the lower (Q1-1.5*IQR) and upper (Q3+1.5*IQR) limit, but are in this case not present.

	LR	RFR	MLP	LSTM	GRU
ME (%TW)	1.87	1.80	1.61	1.95	1.22
MAE (%TW) ±std	4.39 ±0.91	3.85 ±1.13	3.41 ±0.92	2.73 ±0.83	2.90 ±1.05
RSME (%TW)	5.93	5.09	4.60	3.51	3.62
R2	0.11	0.33	0.46	0.67	0.64

Table 7. The average performance accuracy between the measured and predicted normal force on the castor wheels for each machine learning model, where the ME, MAE and RMSE is in percentage of the total weight (%TW). A positive ME means that the predicted value is overestimated compared to the measured value.

3.3 Hyperparameter selection

From the grid research, the hyperparameter combination with one hidden layer, consisting of 50 neurons, with 0.01 learning rate, 128 batch size, 0.1 drop out-rate and 20 time steps gave the lowest MAE value compared to the other 63 combinations. This combination, shown in Table 8, had a MAE of 2.59 ±0.58 %TW, therefore it had a 5.11% improvement compared to the LSTM model used in 3.2 Model selection. Figure 9 gives a schematic overview of these hyperparameters during model training.

Model hyperparameter	Options
Number of hidden layers (h)	1*, 2
Number of neurons per layer (nn)	50*, 70
Learning rate (α)	0.01*, 0.005
Batch size (ξ)	128*, 256
Drop out rate (δ)	0.1*, 0.2
Time steps (ts)	10, 20*

Table 8. Model hyperparameters for the LSTM model, where asterisk indicates the best option according to the grid search.
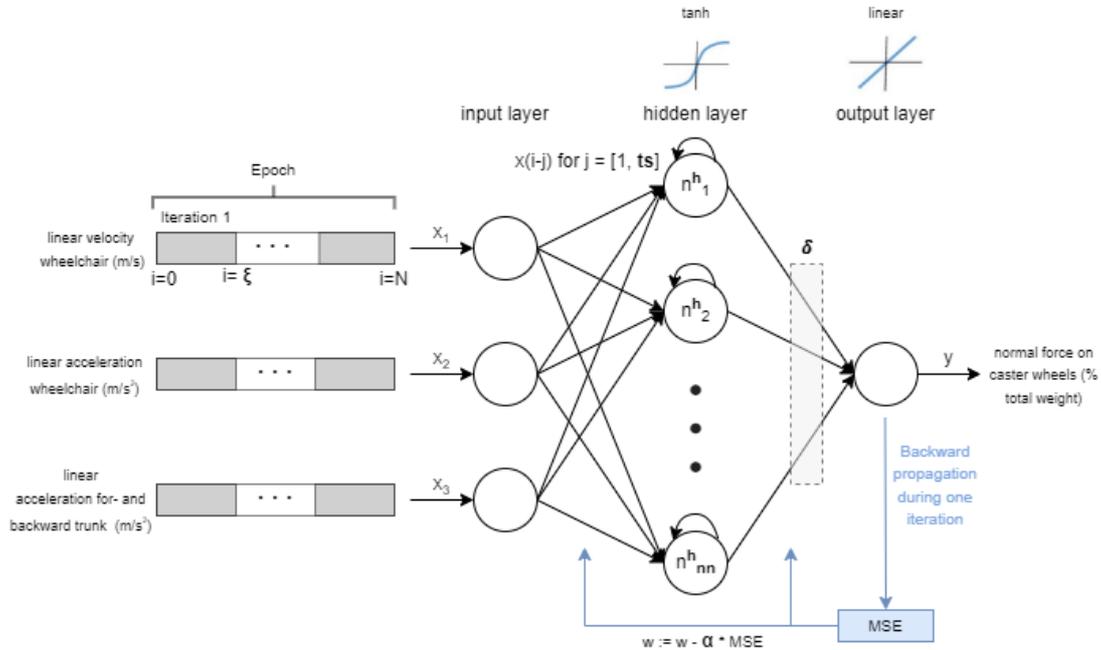

Figure 9. Schematic overview of the hyperparameters (h, nn, α, ξ, δ and ts) for model training, where N is the sample size and w the weights (each connection between nodes). During the first training iteration of one epoch, the first batch of 128 data samples of the three features are input to each of the 50 neuron in the hidden layer. Each neuron in the hidden layer is a LSTM node containing memory cells, therefore the input of a neuron are the current and 19 previous input features. The hyperbolic tangent (tanh) activation function transform the weighted sum of the inputs into an output value. Subsequently, five of the neurons are randomly dropped and had therefore no effect on the linear activation of the neuron in the output layer. Afterwards, the weights are updated by the mean gradient of the batch. These steps are iterated for each batch until the network has seen all the samples.

3.4 Final model training and testing

The execution time to train the model was 12 minutes. The training learning curve in Figure 10 demonstrates that the model is trained on a representative trainings data set as it does not show any noisy movements. The validation learning curve does show some noisy movements, indicating that the model is tested on a less representative data set.

Table 9 indicates that the average MAE for no, moderate and fast trunk movement is 4,4%, 4.4%, and 4.2%TW respectively. This demonstrates that there is no variation in performance accuracy between over the three propulsion techniques regarding trunk movement across the subject test sets in original wheelchair condition.

Table 10 shows that the MAE values of the three test subjects are 5.51%, 3.51% and 2.04%TW. When multiplying with the individual body and wheelchair weight, the values correspond to 35.8, 29.0 and 15.9 N respectively. The small difference between the ME and MAE values indicates that the predicted normal force on the castor wheels for subject test set 1 is overestimated (5.37%TW) and the subject test set 2 it is underestimated (-2.97%TW). Although these subject test set 1 and 2 have certain offset, the course of the normal force can still adequately be predicted, as shown in Figure 11. Moreover, it should be noted that some predictions in subject test set 1 have negative R2 values. This indicates that the mean measured normal force is a better predictor than the model. As the model is still able to predict the course of normal force curve, part of the low R2 values in subject test set 1 can possibly be explained by this offset. Moreover, the difference between the RSME and MAE is relatively small. This indicates is able to make predictions without large absolute errors, as RMSE gives a higher weight than MAE to large errors. To

summarize, the results of the subject test sets demonstrate that the model is able to predict the course of the normal force curve for an unseen subject, except some the predictions may contain an offset.

The last column of Table 10 shows that the average MAE values of the wheelchair condition with +5kg was 3.04 ±1.40%TW and the MAE of the condition with -1.25 bar was 2.83 ±1.01%TW over the training subjects. This demonstrates that the model can predict the normal force on the castor wheels in different wheelchair conditions.

To evaluate whether the architecture of this model is prone to overfitting, a new model was built with an Gaussian noise layer added to original architecture. For the model with the added noise layer, the MAE per test subject during original wheelchair condition was 5.61%, 5.37% and 2.26%TW. Compared to the MAE values of the original model, the new model increased in MAE per test subject with 0.10%, 0.14% and 0.22 %TW. Therefore, the models have similar results, which means that the accuracy of this model is robust to white input noise.
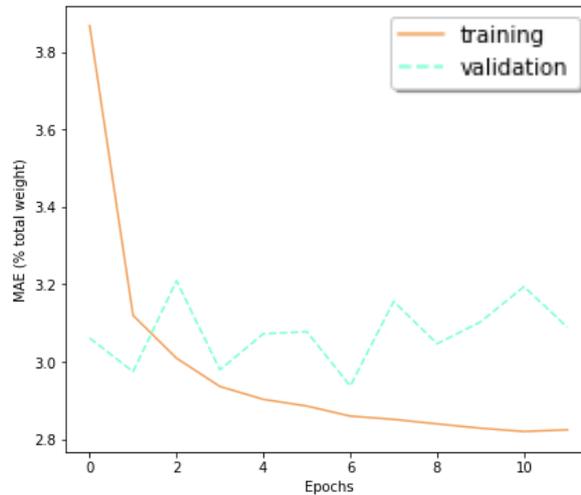

Figure 10. The model performance during training in mean absolute error on the training and validation set.

		Subject test set 1			Subject test set 2			Subject test set 3		
		Trunk movement			Trunk movement			Trunk movement		
		No	Moderate	Fast	No	Moderate	Fast	No	Moderate	Fast
Wheelchair condition: original	ME (%TW)	5.42	5.08	6.47	-6.02	-6.13	-3.48	-0.95	-0.15	-0.84
	MAE (%TW)	5.47	5.13	6.55	6.06	6.17	3.77	1.68	2.04	2.21
	RMSE (%TW)	5.86	5.49	6.92	6.34	6.69	4.62	2.09	2.82	2.89
	R2	-1.80	-0.93	-1.35	-0.44	0.14	0.77	0.78	0.87	0.88

Table 9. Performance accuracy for each subject test set in original wheelchair condition. The ME, MAE and RSME between the measured and predicted normal force on the castor wheels is in percentage of the total weight (%TW).
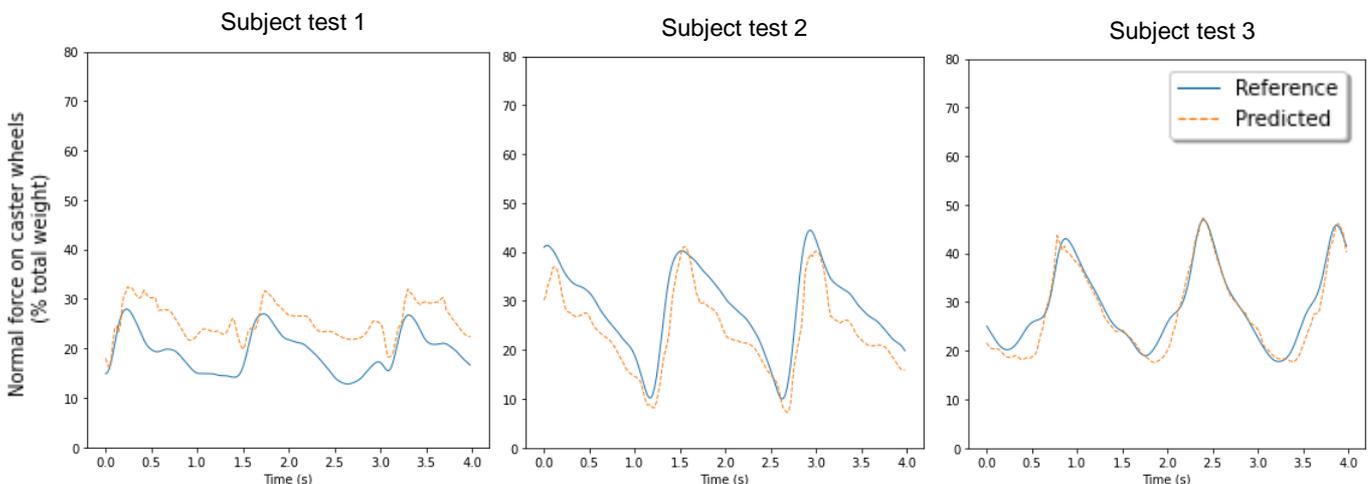

Figure 11. Example of normal force on the castor wheels during a few propulsion cycles with fast trunk movement in original wheelchair condition over the subject test sets. The blue line indicates the measured and the dotted orange line the predicted force

		Subject test set 1	Subject test set 2	Subject test set 3	Mean over all non-test set subjects ± standard deviation
Wheelchair condition: original	ME (%TW)	5.66	-5.21	-0.65	
	MAE (%TW)	5.72	5.33	1.97	
	RMSE (%TW)	6.12	5.96	2.63	
	R2	-1.19	0.42	0.88	
Wheelchair condition test set 1: +5kg	ME (%TW)	6.69	-1.99	1.59	1.32 ±2.24
	MAE (%TW)	6.79	2.72	2.45	3.04 ±1.40
	RMSE (%TW)	7.26	3.49	3.47	3.70 ±1.46
	R2	-1.50	0.82	0.81	0.61 ±0.53
Wheelchair condition: +15kg	ME (%TW)	6.01	-0.17	0.48	
	MAE (%TW)	6.04	1.80	1.98	
	RMSE (%TW)	6.38	2.35	2.86	
	R2	-1.10	0.90	0.79	
Wheelchair condition test set 2: -1.75bar	ME (%TW)	4.73	-3.44	0.16	-0.31 ±1.81
	MAE (%TW)	4.99	3.52	1.86	2.83 ±1.01
	RMSE (%TW)	5.41	4.16	2.60	3.72 ±1.38
	R2	-0.94	0.75	0.87	0.76 ±0.21
Wheelchair condition: -3.5bar	ME (%TW)	3.74	-4.06	0.36	
	MAE (%TW)	4.03	4.17	1.96	
	RMSE (%TW)	4.42	4.73	2.94	
	R2	0.08	0.71	0.85	
Average over each wheelchair condition per subject test set ± standard deviation	ME (%TW)	5.37 ±1.03	-2.97 ±1.75	0.39 ±0.72	
	MAE (%TW)	5.51 ±0.94	3.51 ±1.21	2.04 ±0.21	
	RMSE (%TW)	5.92 ±0.95	4.14 ±1.21	2.90 ±0.31	
	R2	-0.93 ±0.54	0.72 ±0.16	0.84 ±0.03	

Table 10. Performance accuracy for each test set, where the ME, MAE and RMSE between the measured and predicted normal force on the castor wheels is in percentage of the total weight (%TW). The grey cells indicates the complete blind test set containing data from the three subjects in the two wheelchair conditions.

3.5 Friction power
The results of the friction power estimations excluding and including the changes in mass distribution over the three test subjects with no, moderate and fast trunk movement during original wheelchair condition is shown in Table 11. The ME, MAE and RSME between the estimated and measured friction power is given in percentage of the mean friction power (%MFP) calculated by the excluding the changing mass distribution method per trunk movement per test subject. The results indicate that the using the study's model to predict the instantaneous friction power by including the changing mass distribution (MAE: 2.71 and 2.27 %MFP) is more accurate than the excluding method (MAE: 3.68 and 4.26 %MFP) for moderate and especially fast movement. The lower RMSE and higher R2 for the method including the changing distribution also support this statement. For visualization, an example of each subject test set during fast trunk movement is given in Figure 12. Nevertheless, during no trunk movement, the power prediction including the changing distribution (MAE: 2.64 %TW) has similar accuracy to the excluding method (MAE: 2.24 %MFP). This makes sense given the fact that the two methods use approximately the same mass distribution on the wheels. It is also logical that the friction power during fast movement is best predicted with the method incorporating the mass distribution, as the higher load on the castor wheels has a large effect on the friction power. This demonstrates that an IMU placed on the trunk will be valuable in predicting the friction power during moderate and fast trunk movement, but not when no trunk movement is applied.

Moreover, the results show that the R2 values of the friction power prediction based on including the changing mass distribution method for test subject 1 are not negative. While the R2 values of the normal force prediction model for the same test subject were negative (see Table 10). This may indicate that the errors made in predicting the normal force on the castor wheels diminish when using this value in the prediction of the friction power.

Original wheelchair condition		Subject test set 1			Subject test set 2			Subject test set 3			Average subject test sets ± standard deviation		
		Trunk movement			Trunk movement			Trunk movement			Trunk movement		
		No	Moderate	Fast	No	Moderate	Fast	No	Moderate	Fast	No	Moderate	Fast
Excl. changing mass distri-bution	ME (%MFP)	-1.44	-2.12	-2.00	0.70	-1.25	-0.83	-1.97	-5.42	-5.89	**-0.90 ±1.15**	**-2.93 ±1.80**	**-2.91 ±2.16**
	MAE (%MFP)	1.70	2.26	2.20	2.34	3.26	4.62	2.69	5.50	5.95	**2.24 ±0.41**	**3.67 ±1.36**	**4.26 ±1.55**
	RMSE (%MFP)	2.06	2.72	2.73	2.99	4.30	5.46	3.33	7.25	7.63	**2.79 ±0.54**	**4.75 ±1.87**	**5.27 ±2.00**
	R2	0.90	0.87	0.75	0.82	0.68	0.74	0.77	0.62	0.27	**0.83 ±0.05**	**0.72 ±0.11**	**0.58 ±0.23**
Incl. changing mass distri-bution	ME (%MFP)	2.14	2.00	2.54	-4.78	-4.89	-2.76	-0.54	-0.08	-0.49	**-1.06 ±2.85**	**-0.99 ±2.88**	**-0.24 ±2.17**
	MAE (%MFP)	2.16	2.02	2.57	4.81	4.92	2.98	0.97	1.17	1.27	**2.64 ±1.60**	**2.71 ±1.61**	**2.27 ±0.73**
	RMSE (%MFP)	2.33	2.17	2.72	5.05	5.37	3.63	1.22	1.64	1.66	**2.87 ±1.61**	**3.06 ±1.65**	**2.67 ±0.81**
	R2	0.89	0.90	0.75	0.76	0.80	0.94	0.97	0.98	0.96	**0.87 ±0.08**	**0.89 ±0.08**	**0.88 ±0.10**

Table 11. Performance accuracy estimating the friction power for each subject test set in original wheelchair condition excluding and including the mass distribution. The ME, MAE and RMSE between the measured and friction power is in percentage of the mean measured friction power (%MFP).
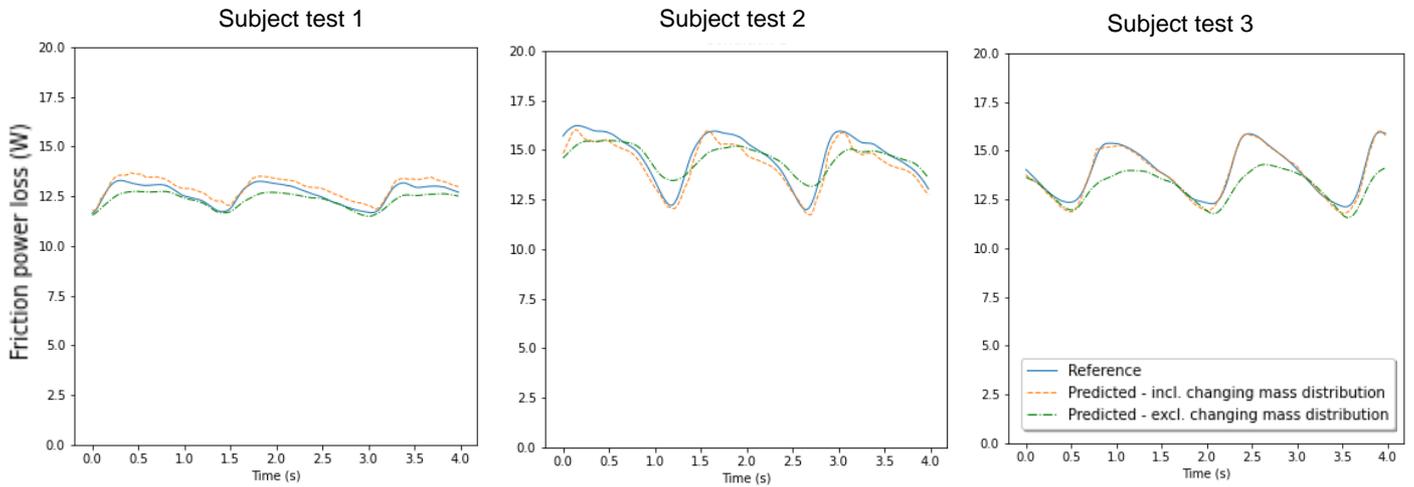

Figure 12. Example of friction power during a few propulsion cycles with fast trunk movement in original wheelchair condition over the subject test sets. The blue line indicates the reference friction power, the orange line the predicted friction power method including the mass distribution and the green line the excluding distribution method.

4 Discussion

The first aim of this master thesis is to explore whether the normal force on the castor wheels can be estimated from IMU data using a machine learning approach. When using the linear velocity and acceleration of the wheelchair and the, linear acceleration of the for- and backward movement of the trunk, an adequate estimation of the normal force (MAE of 3.69% total weight) from an LSTM model can be obtained for unseen subjects. This model is robust for wheelchair settings regarding wheelchair mass and tire pressure, subjects with a high upper body mass ratio (e.g. when missing a leg), for propulsions with no, moderate and fast trunk movement and, for white noise in the input data. The second aim is to evaluate whether incorporating the changing load distribution (including the effect of the trunk inclination) could improve the friction power estimation compared with neglecting the changes in load distribution (excluding the effect of trunk inclination). The method did prove to be more accurate in predicting instantaneous friction power during propulsions with moderate and especially fast trunk movement. Except during propulsions without trunk movement, the friction power prediction methods including and excluding the changing mass distribution gave similar results.

The results show a over- and underestimation in normal force on the castors for test subjects 1 and 2, respectively. A possible explanation is that these subjects had a lower respectively higher normal force on the castors compared to average normal force in percentage of the total weight. However, the model is proven to be robust for subjects with a high upper body mass ratio. Therefore, the underestimation cannot be related to the higher normal force on the castors compared to average force in percentage of the total weight. Yet, it could be possible that the model is not robust for subjects with a low upper body mass ratio. An other possible reason for over- and underestimation is the occurrence of hardware problems in the load pins. During some of the experiments, the predefined voltage baseline in one or both load pins moved up or down. Dirt or hairs stuck around the load pin caused a higher force measurement and wires loosely connected to load pins resulted in a lower force measurement. Once this was noted, the load pins were cleaned and wires were checked at the beginning of the experiment of each new subject and subsequently the voltage was reset to the predefined baseline (as discussed in Appendix A). If an over- or underestimation in voltage was observed between blocks, the wheels were cleaned and wires reconnected as well. However, this was not observed prior each block within one subject. As a result, the measured force on the castor wheels of some of the 150 experiments could contain incorrect data and it is unknown which experiments were affected. It cannot be determined which experiments in the data set contained incorrect measurements based on MAE and ME values. As the over- and underestimation could also be explained due to the fact that the user employs a different kind of propulsion technique, whereof the model did not learn the relationship correlated to the normal force or its effect on the force cannot be assessed by the current predictor features. Because other explanations for the over- and underestimation could not be excluded, none of the data sets were excluded due to the shift of voltage baseline in the load pins.

As a result, the model was trained and tested on data including some incorrect force measurements. Nevertheless, the training learning curve in Figure 9 does not contain any noisy movements, indicating that the model is trained on a representative trainings data set. Therefore, the training set (42% of the data) contains enough data with correct force measurements to ensure learning between the predictor and target features. However, the validation learning curve does exhibit some noisy movements, indicating that the model is validated on a less representative data set (12% of the data). Therefore, when the evaluation of this model is done on a small batch of the test set, it may also be evaluated on a less representative data set. Hence, the performance accuracy of this model can be seen as a lower bound. Accordingly, the reference friction power, based on this force estimation, can also be seen as an lower bound. When the normal force prediction model is evaluated on a test set with an optimal data set, the errors may be lower. Such an optimal data set of normal force data values could be obtained from the center of mass of the athlete-wheelchair position relative to the castor- and rear-wheel axes by an optical motion capture system, and vertical center of mass acceleration. Subsequently, this method can also more accurately estimate the total normal force compared to the current simplified calculation that disregards the vertical acceleration of the wheelchair-user system. Note that, although this method is non-invasive, it can only measure human movement in a confined space. This method is not feasible when multiple cameras are needed to record movements over a large distance or in a crowded area, which can occur in typical sport situations.

5 Conclusion & Recommendations

This research concluded that two IMUs can be used to non-invasively and cheaply estimate the normal force on the castor wheels in-field. Moreover, this predicted instantaneous normal force can be used to calculate the friction power. This method gives more accurate results during propulsions with moderate and fast trunk movement compared to the excluding changing distribution method. Therefore, this model may has practical relevance during an in-field analysis of the effect of propulsion techniques with different trunk movements on the friction power. Or during an study on different wheelchair frames and its effect on the friction power. As part of a larger context, this research will also contribute to the process of filling the technological gap of in-field monitoring mechanical power via the right-hand-side method during wheelchair sports.

Future coaches, sport scientists, and wheelchair athletes could estimate the instantaneous normal force on the castor wheels and friction power as followed. First, an IMU is placed on the wheel and on the trunk of the user. Secondly, this IMU data is send to a computer, which instantaneous extracts the wheelchair acceleration and velocity and, the for- and backward movement of the trunk. Thirdly, this data is fed into the model (available upon request), which estimates the normal force on the castor wheels in real-time. Lastly, the friction power can be estimated according to the subject body and wheelchair mass and, rolling resistance coefficients of the wheels and instantaneous wheelchair velocity. If these computations could be preformed in a mobile application specialized in displaying sport performance (e.g. Strava™, Garmin™), the user could easily track the real-time friction power during wheelchair propulsion.

Although this model has a feasible robustness for different trunk motions and wheelchair conditions, the robustness needs further examination. The model is mainly built on linear steady-state wheelchair propulsions, only a relative small number of data samples contained acceleration from 1.2 to 1.7 m/s. Nevertheless, wheelchair athletes perform fast accelerations and propulsions with non-linear trajectory during game situations. Moreover, during this study a standard wheelchair is used on a treadmill surface, although in practise a sport wheelchair with cambered wheels on hard or gym court surface is employed. Therefore, future research must validate how accurate this model can predict the normal force on the castor wheels during accelerations and non-linear trajectory and, in a sport wheelchair on court surface during training and game situations. When the custom-made load pins are used in such experiments, it is highly recommended to verify prior and after each experiment if the voltage is equivalent to the predetermined baseline. Moreover, if it is possible in the experimental setup, the reference normal force on the castors and friction power could be more accurately calculated by a optical motion capture system.

The ongoing development in sensors may lead to other innovations that could even measure the centre of mass of the total system. Once the in-field velocity and acceleration of the centre of mass can be accurately estimated from IMUs or from other non-invasive and cheap sensors, it could be used to predict the normal force on the castor wheels and consequently the friction power. Besides the friction power term, the inertia of the centre of mass can also be used for the calculation of the kinetic power to estimate together the mechanical power output balance via the right-hand-side method.

Bibliography

[1]   M. P. van Dijk, R. M. A. van der Slikke, M. A. M. Berger, M. J. M. Hoozemans, and H. E. J. Veeger, "LOOK MUMMY, NO HANDS! THE EFFECT OF TRUNK MOTION ON FORWARD WHEELCHAIR PROPULSION," 2021. [Online]. Available: https://commons.nmu.edu/isbs/vol39/iss1/80

[2]   D. Theisen, M. Francaux, A. Fay, and X. Sturbois, "A New Procedure to Determine External Power Output During Handrim Wheelchair Propulsion on a Roller Ergometer: A Reliability Study," *Int J Sports Med*, vol. 17, no. 08, pp. 564–571, Nov. 1996, doi: 10.1055/s-2007-972896.

[3]   R. M. A. van der Slikke, M. A. M. Berger, D. J. J. Bregman, and D. H. E. J. Veeger, "Wearable wheelchair mobility performance measurement in basketball, rugby, and tennis: Lessons for classification and training," *Sensors (Switzerland)*, vol. 20, no. 12, pp. 1–13, Jun. 2020, doi: 10.3390/s20123518.

[4]   A. R. Lewis, E. J. Phillips, W. S. P. Robertson, P. N. Grimshaw, M. Portus, and J. Winter, "A practical assessment of wheelchair racing performance kinetics using accelerometers," *Sports Biomech*, vol. 20, no. 8, pp. 1001–1014, Nov. 2021, doi: 10.1080/14763141.2019.1634136.

[5]   T. Rietveld, B. S. Mason, V. L. Goosey-Tolfrey, L. H. V. van der Woude, S. de Groot, and R. J. K. Vegter, "Inertial measurement units to estimate drag forces and power output during standardised wheelchair tennis coast-down and sprint tests," *Sports Biomech*, 2021, doi: 10.1080/14763141.2021.1902555.

[6]   R. J. K. Vegter, S. de Groot, C. J. Lamoth, D. H. Veeger, and L. H. v. van der Woude, "Initial Skill Acquisition of Handrim Wheelchair Propulsion: A New Perspective," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 104–113, Jan. 2014, doi: 10.1109/TNSRE.2013.2280301.

[7]   E. van der Kruk, F. C. T. van der Helm, H. E. J. Veeger, and A. L. Schwab, "Power in sports: A literature review on the application, assumptions, and terminology of mechanical power in sport research," *J Biomech*, vol. 79, pp. 1–14, Oct. 2018, doi: 10.1016/j.jbiomech.2018.08.031.

[8]   V. G. de Vette, D. (H. E. J.) Veeger, and M. P. van Dijk, "Using Wearable Sensors to Estimate Mechanical Power Output in Cyclical Sports Other than Cycling—A Review," *Sensors*, vol. 23, no. 1, p. 50, Dec. 2022, doi: 10.3390/s23010050.

[9]   J. Shepherd, D. James, H. Espinosa, D. Thiel, and D. Rowlands, "A Literature Review Informing an Operational Guideline for Inertial Sensor Propulsion Measurement in Wheelchair Court Sports," *Sports*, vol. 6, no. 2, p. 34, Apr. 2018, doi: 10.3390/sports6020034.

[10]   R. S. Alcantara, W. B. Edwards, G. Y. Millet, and A. M. Grabowski, "Predicting continuous ground reaction forces from accelerometers during uphill and downhill running: a recurrent neural network solution," *PeerJ*, vol. 10, Jan. 2022, doi: 10.7717/peerj.12752.

[11]   J. McGrath, J. Neville, T. Stewart, and J. Cronin, "Upper body activity classification using an inertial measurement unit in court and field-based sports: A systematic review," *Proc Inst Mech Eng P J Sport Eng Technol*, vol. 235, no. 2, pp. 83–95, Jun. 2021, doi: 10.1177/1754337120959754.

[12]   R. M. A. van der Slikke, M. A. M. Berger, D. J. J. Bregman, A. H. Lagerberg, and H. E. J. Veeger, "Opportunities for measuring wheelchair kinematics in match settings; reliability of a three inertial sensor configuration," *J Biomech*, vol. 48, no. 12, pp. 3398–3405, Sep. 2015, doi: 10.1016/j.jbiomech.2015.06.001.

[13]   M. Z. Uddin *et al.*, "Estimation of Mechanical Power Output Employing Deep Learning on Inertial Measurement Data in Roller Ski Skating," *Sensors*, vol. 21, no. 19, p. 6500, Sep. 2021, doi: 10.3390/s21196500.

[14]   C. Sauret, P. Vaslin, F. Lavaste, N. de Saint Remy, and M. Cid, "Effects of user's actions on rolling resistance and wheelchair stability during handrim wheelchair propulsion in the field," *Med Eng Phys*, vol. 35, no. 3, pp. 289–297, Mar. 2013, doi: 10.1016/j.medengphy.2012.05.001.

[15]   M. P. van Dijk, M. Kok, M. A. M. Berger, M. J. M. Hoozemans, and D. H. E. J. Veeger, "Machine Learning to Improve Orientation Estimation in Sports Situations Challenging for Inertial Sensor Use," *Front Sports Act Living*, vol. 3, Aug. 2021, doi: 10.3389/fspor.2021.670263.

[16]   S. Komaris *et al.*, "A Comparison Of Three Methods for Estimating Vertical Ground Reaction Forces in Running," Mar. 2020.

[17]   T. Yiu, "Understanding Random Forest," *Towards Data Science*, Sep. 2021. https://towardsdatascience.com/understanding-random-forest-58381e0602d2 (accessed Aug. 01, 2022).

[18]   F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," Jan. 2012.

[19]   D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, 6th ed. Nashville, TN: John Wiley & Sons, 2014.

[20]   G. Leporace, L. A. Batista, L. Metsavaht, and J. Nadal, "Residual analysis of ground reaction forces simulation during gait using neural networks with different configurations," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015, vol. 2015-Novem, no. August 2015, pp. 2812–2815. doi: 10.1109/EMBC.2015.7318976.

[21]   G. Leporace, L. A. Batista, and J. Nadal, "Prediction of 3D ground reaction forces during gait based on accelerometer data," *Research on Biomedical Engineering*, vol. 34, no. 3, pp. 211–216, Sep. 2018, doi: 10.1590/2446-4740.06817.

[22]   M. Pogson, J. Verheul, M. A. Robinson, J. Vanrenterghem, and P. Lisboa, "A neural network method to predict task- and step-specific ground reaction force magnitudes from trunk accelerations during running activities," *Med Eng Phys*, vol. 78, pp. 82–89, Apr. 2020, doi: 10.1016/j.medengphy.2020.02.002.

[23]   A. K. Jain, Jianchang Mao, and K. M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer (Long Beach Calif)*, vol. 29, no. 3, pp. 31–44, Mar. 1996, doi: 10.1109/2.485891.

[24]   D. Sharma, P. Davidson, P. Müller, and R. Piché, "Indirect Estimation of Vertical Ground Reaction Force from a Body-Mounted INS/GPS Using Machine Learning," *Sensors*, vol. 21, no. 4, p. 1553, Feb. 2021, doi: 10.3390/s21041553.

[25]   P. Davidson, H. Virekunnas, D. Sharma, R. Piché, and N. Cronin, "Continuous Analysis of Running Mechanics by Means of an Integrated INS/GPS Device," *Sensors*, vol. 19, no. 6, p. 1480, Mar. 2019, doi: 10.3390/s19061480.

[26]   F. Chollet, "Keras," *GitHub*, 2015. https://github.com/fchollet/keras (accessed Aug. 31, 2022).

[27]   Batarow Sensorik GmbH, "M1787 Loadpin Specsheet." 2021.

Appendix

A.1 Load pins

A.1.1 Installation
The customized load pins (MB1787, Batarow Sensorik GmbH, Karow, Germany) measure the force in vertical direction (i.e., normal force) on the castor wheels of a wheelchair. The overall installation of the load pins on the wheelchair used during this study and its components can be seen in Figure A1. The load pin is positioned in the castor wheels (1) and fixed frame with a customized keyplate (2) to prevent load pin rotation during propulsion and maintain vertical force measurement. The wiring of the load pins passes through the hollow bolt and is attached on top to a slip ring (3), which guides the wires as the castor rotates and prevents them from curling. The wires are connected to three 9V batteries (4) placed in a box underneath the seating of the wheelchair and to an IMU (5) located at the frame underneath the seating. The batteries provide a 27V input to the load pins and the IMU monitors the real-time voltage output and has data-logging capabilities.
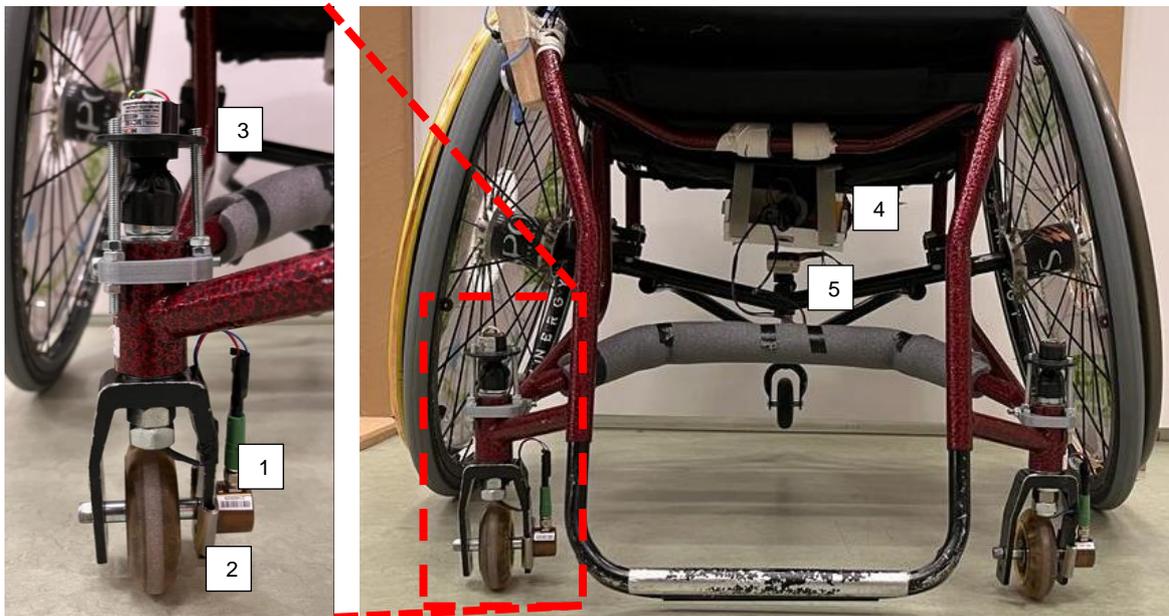

Figure A1. The integration of the load pin system into the wheelchair, with the following components: load pins (1), keyplate (2), slip ring (3), 3x9V batteries (4) and IMU$_{loadpins}$ (5).

A.1.2 Force measurement
The customized load pins have a bored centre, which contains an internal compression force measuring strain gauge (Figure A2). According to the Piezoresistive effect formulated in Equation A1, the resistance (R) of the strain gauges can be calculated with the resistivity of the material or conductivity (p), the length (l) and the area of cross-section (A).

$$R = (p * l)/A \qquad (A1)$$

When the strain gauge is compressed (within the limits of its elasticity such that it does not break or permanently deform) the area of cross-section will increase (A↑), and the length of the strain gauge will become shorter (l↓). Therefore, the resistance of the strain gauge decreases, when the compression force increases.

During the operation of the load pins, a constant voltage is applied across the strain gauge. According to Ohm's law formulated in Equation A2, the current (I) can be calculated with the voltage (V) and resistance (R). This formula

$$I = U/R \qquad (A2)$$

Therefore, when the compression force increases, the resistance decreases and the current increases. Hence, the output current is proportional to the applied compression force and a ratio can be established as described in Equation A3.

$$F/I = F_{max}/I_{max}$$

$$F = (F_{max}/I_{max}) * I \qquad\qquad (A3)$$

The custom-made load pins measure a maximum voltage of 500 N and current of 20 mA[27]. However, the gauge strain uses a superconductor material, making it possible to have a 4mA current without voltage [27]. Moreover, the resistance of the strain gauge is 151 Ohm [27]. By filling in these variables in Equation A2 and A3, the resulting formula for each load pin is given in Equation A4.

$$F_{castorwheel} = (500/ (0.020\text{-}0.004)) * (I\text{-}0.004)$$

$$F_{castorwheel} = (500/ (0.020\text{-}0.004)) * ((U_{output}/151)\text{-}0.004)$$

$$F_{castorwheel} = 207 * U_{output} - 125 \qquad\qquad (A4)$$

The voltage output from the load pins (U_{output}) can be monitored as analogue input by an IMU. Consequently, this signal can be converted into a force measurement using Equation A4. It can be assumed that the total rolling resistance coefficient over the two castor wheels are equal. Therefore, the total rolling friction force of castors can be measured based on the total normal force on both the castor wheels, as shown in Equation A5.

$$F_{total\_castorwheels} = 207 * U_{output} - 250 \qquad\qquad (A5)$$

When placing one load pin in the castor wheels and not applying any compression load on it, the voltage output is 0.6V. However, when mounting the load pin to the castor wheel the voltage output increases. The ground applies compression force in one direction, while the fixed support applies force in the opposite direction (Figure A2). When tightening the nut to mount de load pin to the frame, the fixed support will increase, causing more compression force needed to deform the strain gauge. This leads to an increase in the gradient and offset in Equation 4. Yet, it is essential that the load pins fit tightly in the castors of the wheelchair, because the castor wheels must not jiggle. On the wheelchair used in this study, the voltage given by the load pins when no force is applied, and the castor wheels fit tightly in the frame is around 1.0V. Therefore, prior to any experiment on the treadmill the load pins were cleaned, and wires checked and manually mounted to this voltage. The subjects in a wheelchair performed simulated propulsion motions with no, moderate and fast trunk movement from a stand still for ~2 minutes. The rear wheels were fixed to a dummy plate and the castor wheels rotated in propulsion position (as in Figure A1) were placed on a force plate. The two signals were time synchronized and filtered by a second order low-pass Butterworth filter with cut-off van 6 Hz. Afterwards, first order polynomial is calculated that had the best fit for the load pin data in voltage against the force plate data in Newton. Then the load pin data in voltage was filled in this polynomial, which resulted in load pin data given in Newton. Over all the subjects, the average coefficients of the polynomial are shown in Equation A6.

$$F_{total\_castorwheels} = 228.26 * U_{output} - 442.34 \qquad\qquad (A6)$$

For the first coefficient the standard deviation was ±3.8 and for the second ±9.8. These deviations are assumed to be the result of manual errors while mounting the voltage to 1.0V. Therefore, for each subject the load pin data in voltage is converted to Newton using the subject's individual established polynomial coefficients. The mean absolute error between the load pin data in Newton and force plate data over all the subjects was 3.65±1.75 N, which is 0.74±0.33 % of the total force measured. The uncertainty with 95% confidence interval is ±2,50 [27]. Therefore, the polynomial technique can be considered as accurate.
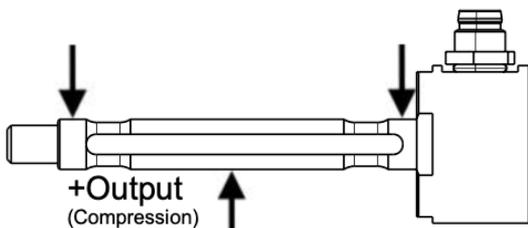

Figure A2. The custom made MB1787 load pin with a bored centre that is subjected to compression force.