MACHINE LEARNING-BASED ESTIMATION OF 3D KINETICS DURING NORMAL WALKING AND SINGLE-LEG HOP TESTS IN PATIENTS WITH ANTERIOR CRUCIATE LIGAMENT INJURY

Master Thesis

Maaike R. de Bondt

Supervisors:

Ir. S. Krishnakumar

Dr. Ir. B.J.F. van Beijnum

Prof. Dr. J.H. Buurke

Dr. E.H.F. van Asseldonk

August 2023

UNIVERSITY OF TWENTE.

Table of Contents

Abstract

The analysis of quantitative data, particularly kinetics, can offer valuable insights for the monitoring of the rehabilitation process in cases of anterior cruciate ligament (ACL) injuries. However, the conventional methods of collecting kinetic data, using a force plate and an optical motion capture (OMC) system, are time-consuming and require an expensive laboratory setup. This study aims to validate existing machine learning (ML) algorithms based on inertial measurement units (IMUs) from literature and proposes three novel ML algorithms for estimating three-dimensional (3D) ground reaction forces (GRFs) and net knee joint moments during walking and single-leg hop tests. Nine healthy participants and eight ACL patients were assessed during walking and single-hop tests while wearing thirty reflective markers and eight IMUs. The collected dataset was used for the validation and development of the ML algorithms. The algorithms of Leporace and Stetter et al. were validated. Additionally, three novel movement-specific artificial neural networks (ANNs) were designed, and hyperparameter optimization and principal component analysis (PCA) based input selection were performed. The model evaluation based on leave-one-subject-out cross-validation (LOSO CV) among the ACL patients showed rRMSE values of 9.49, 16.28 and 11.79 percent during stance phase of walking for anterior-posterior, medial-lateral and vertical direction of GRF respectively. The most important finding of the present study is that medial-lateral GRFs and all net knee moments are still difficult to accurately estimate during a single leg hop test. However, it does show that ACL patients are a suitable population to develop and evaluate ML algorithms on. The current study serves as a first step in providing quantitative data for monitoring the ACL rehabilitation process.

Index terms

Anterior Cruciate Ligament, ground reaction force, inertial measurement unit, net knee joint moment, machine learning,

Samenvatting (Dutch)

De analyse van kwantitatieve data, in het bijzonder kinetica, kan waardevolle inzichten geven voor het monitoren van het revalidatieproces van voorste kruisband (VKB) letsels. Echter, de conventionele kinetica verzamelende methodes, gebruik makende van een krachtplaat en het optisch vastleggen van beweging, zijn tijdrovend en gebonden aan een dure meetopstelling. Het huidige onderzoek heeft als doel om bestaande machine learning (ML) algoritmen gebaseerd op inertiële meeteenheden te valideren en drie nieuwe ML algoritmen voor te stellen die driedimensionale (3D) grondreactiekrachten en netto knie gewrichtsmomenten bepalen tijdens lopen en eenbenige sprongtesten. Negen gezonde deelnemers en acht VKB patiënten zijn gemeten tijdens lopen en eenbenige sprongtesten terwijl zij dertig reflectieve markers en acht inertiële meeteenheden droegen. De gegevens die hierbij zijn verkregen zijn gebruikt voor het valideren en ontwikkelen van de ML algoritmen. De algoritmen van Leporace et al. en Stetter et al. zijn gevalideerd. Daarnaast zijn drie nieuwe bewegingsspecifieke kunstmatige neurale netwerken (ANNs) ontworpen en optimalisatie van hyperparameters en selectie van input parameters op basis van principale-componentenanalyse (PCA) is uitgevoerd. De evaluatie van het model tijdens de stand fase van het lopen in VKB patiënten leverde rRMSE waarden op van 9.49, 16.28 en 11.79 procent voor respectievelijk de anterior-posterior, mediaal-lateraal en verticale richting van de grondreactiekracht. De belangrijkste vondst uit het huidige onderzoek is dat mediale-laterale grondreactiekrachten en alle netto kniemomenten nog steeds moeilijk accuraat te bepalen zijn tijdens een sprongtest. Desalniettemin laat dit onderzoek zien dat VKB patiënten een geschikte populatie zijn om ML algoritmen voor te ontwikkelen en evalueren. Het huidige onderzoek dient als een eerste stap in het voorzien van kwantitatieve data voor het monitoren van het VKB revalidatie proces.

Index woorden

Voorste kruisband, grondreactiekracht, inertiële meeteenheid, netto kniegewrichtsmoment, machine learning

List of abbreviations

3D = three-dimensional

ACL = anterior cruciate ligament

ANN = artificial neural network

A-P = anterior-posterior

BLC = bony landmark calibration

BW = body weight

IMU = inertial measurement unit

F-E = flexion-extension

GRF = ground reaction force

I-E = internal-external

KJF = knee joint force

LLL = left lower leg

LOSO CV = leave-one-subject-out cross-validation

ML = machine Learning

M-L = medial-lateral

NN = neural network

OHL = own hop landing *(name of a model)*

OHP = own hop push-off *(name of a model)*

OLW = own Leporace walk *(name of a model)*

OMC = optical motion capture

OSHL = own Stetter hop landing *(name of a model)*

OSHP = own Stetter hop push-off *(name of a model)*

OW = own walk *(name of a model)*

PC = principal component

PCA = principal component analysis

RRD = Roessingh research and development

rRMSE = relative root mean square error

T-C = tension-compression

vGRF = vertical ground reaction force

V-V = varus-valgus

Introduction

Numerous sports players, particularly those engaged in soccer or skiing, may perceive an audible popping sensation within their knee during a sudden change of direction, a sudden stop, or an unanticipated landing from a jump [1]. This is a well-known symptom of an Anterior Cruciate Ligament (ACL) injury. To rehabilitate from an ACL reconstruction surgery, exercises, such as squats, jumps and cutting maneuvers, are trained for six to nine months to enhance muscle strength and neuromuscular control [2]. This rehabilitation process is divided into five different goal-based phases [3]. The physiotherapist guides this process from phase to phase by observing the performed movements and monitoring the patient's progress. However, tracking this development is mostly a subjective task, which can make it challenging. Up until now, only kinematics are measured or observed in clinical practice. As the combination of anterior shear force, valgus moment, and internal rotational moment of the knee are believed to be key factors for ACL injury, analyzing knee joint kinetics will add valuable information regarding prevention of re-injuries [4]. Additionally, measuring kinetics will be appreciated, as these quantitative data can easily be compared and can be used to make informed decisions regarding entering the following rehabilitation phase.

There are different ways to obtain objective measurements. Firstly, an optical motion capture (OMC) system can measure kinematics. Secondly, the kinetics can be estimated by inverse dynamics using ground reaction forces (GRFs) measured from force plates. Whilst these methods are the current standards for obtaining kinematics and kinetics, they also have disadvantages. Some disadvantages of OMC systems are their immobility and the time needed to prepare a patient and post-process data, using an expensive setup. This makes OMC systems unsuitable for use at the physiotherapist's practice and makes them unable to track on-field activities.

Inertial Measurement Units (IMUs) are relatively inexpensive and portable for on-site usage [5]. Together with the fact that they give three-dimensional (3D) acceleration, angular velocity, and magnetometer data, this makes them a possible solution to the aforementioned problems to collect kinematics. However, up until now IMUs are mostly used to provide kinematics and estimating kinetics using IMUs is not that common yet.

Although kinetics can be estimated from kinematics in at least three different ways (biomechanical modeling, musculoskeletal modeling, and using machine learning (ML) algorithms), the focus of the current study lies on ML algorithms [6]–[8]. An advantage of ML-based approaches is that little to no a priori knowledge is needed; the model is purely built by using training data and only a few anthropometric measures are needed to personalize the model [9]. Many studies have been performed trying to relate IMU data to GRFs [8], [10]–[19]. Differences between these studies were mainly found in the location and number of IMU sensors, the type of task performed, the type of ML algorithm used, and the type of subjects. Mundt et al. compared three different artificial neural networks (ANNs), concluding that the dataset and the prediction tasks are more important than the type of neural network (NN), since all three networks were able to accomplish a high prediction accuracy [20]. Dorschky et al. suggested that ML models dedicated to one single task will probably outperform ML models which were jointly tuned for different output variables [16]. This indicates the need for task-specific and outcome-specific modeling [9], [16]. Next to this, it is mentioned that an examination of different network architectures is needed [16]. Networks need to avoid computing features of sensor data, such as fully convolutional networks, and need to allow a continuous estimation of kinetics, such as long short-term memory networks [16].

Whereas the studied type of subjects varied from knee osteoarthritis patients to top athletes, only little recent attention was focused on people with pathological gait. To the best of our knowledge, knee arthroplasty, knee osteoarthritis, and hip osteoarthritis are the only types of pathological gait that were studied using IMUs for the estimation of lower body kinetics [21]–[23]. Therefore, the current research will validate existing algorithms on data of ACL patients to widen the research into pathological gait.

Studying ACL patients, it might still be challenging to determine the kinetics of ACL rehabilitation-specific movements. The main reason for this is the relatively small amount of research that has been carried out regarding ACL rehabilitation-specific movements, such as hop tests, in comparison to the number of studies focused on walking or running. The walking movement will still be evaluated in the current research to ensure comparability with existing literature. Additionally, single leg hop tests will be studied since analysis of this ACL-specific movement is a first step towards the ultimate goal of measuring irregular, on-field activities. It is important to assess this movement as it is different from walking and running: the motion is not cyclical and mediolateral forces also play an essential role in the stability during the movement [24].

This research aims to validate existing ML algorithms from literature estimating kinetics based on IMUs against a new dataset of healthy subjects and ACL patients. Additionally, this study aims to propose three novel ML algorithms for estimating 3D GRFs and net knee joint moments during walking and single-leg hop tests. The use of IMUs and ML algorithms is compared to the current standard: force plates and an OMC system.

Methods

Data collection

Measurement data were collected from nine healthy participants (3 M/6 F, height: 173.8 ± 7.8 cm, body mass: 71.0 ± 8.9 kg) and 8 ACL-injured patients (2 M/6 F, height: 173.4 ± 9.4 cm, body mass: 74.2 ± 17.7 kg, ACL leg: 6 R/2 L) at Roessingh Research and Development (RRD) in Enschede, The Netherlands. All participants were informed of the experimental procedures, approved by the local ethical committee at the University of Twente, and provided written consent.

Prior to attaching the data collection equipment, standardized anthropometric measurements were taken (including height, weight, knee and ankle width, inter-anterior superior iliac spine distance, and the length of both legs) to use for further data processing. During the measurements, participants were equipped with 30 reflective markers and 8 IMUs (Xsens, Enschede, The Netherlands) [25]. The IMUs were each mounted on a rig with 3 reflective markers attached. The precise positions of these markers and sensors can be found in Table 8 and Figure 1. The IMU data was recorded at 40 Hz using in-house developed FusionClinician software (RRD, Enschede, the Netherlands). The markers were tracked by an eight-camera Vicon OMC system (four Vero 2.2 MP cameras and four Vantage 5 MP cameras, Vicon Motion Systems, Oxford, UK) that sampled motion at 100 Hz. 3D GRFs were registered using two force plates embedded in the floor (OR6-5-1000, AMTI, Watertown, MA, USA), sampled at 1000 Hz.

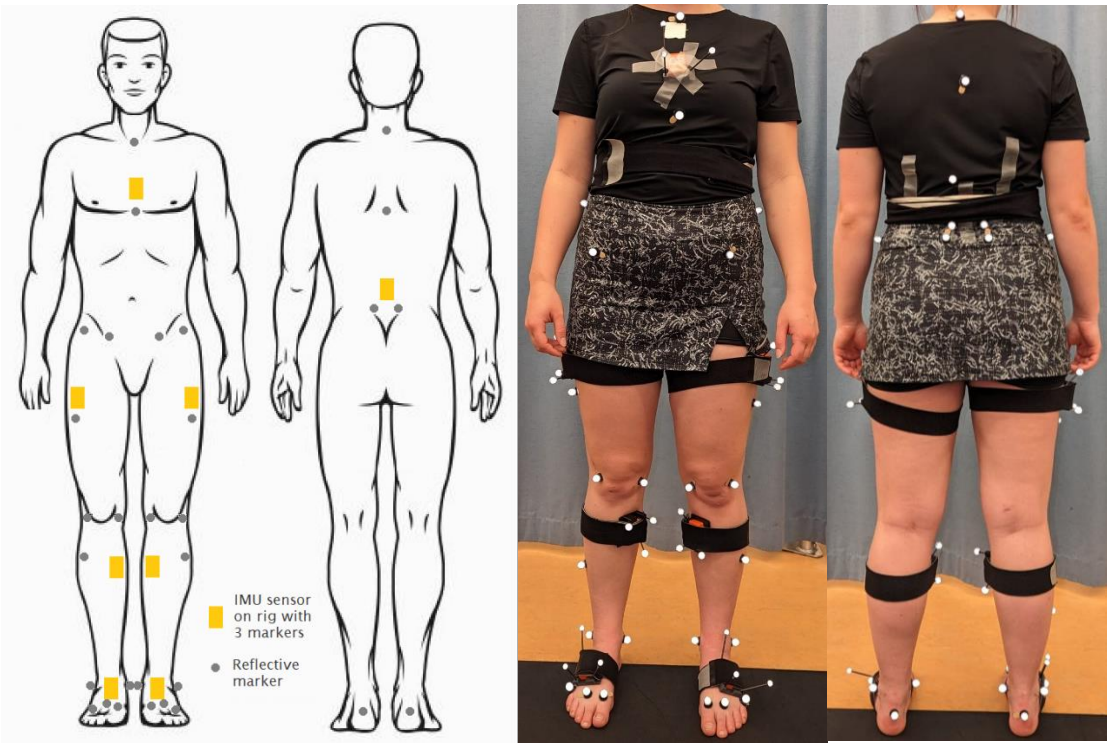

Figure 1 Location of the reflective markers and the IMU sensors. Left: schematic. Right: during experimental set-up.

Each participant started with a segment calibration task, which included bending the upper body, bending the knees, and tipping the toes five times. After this, the participant performed two successful trials of walking and single leg hop tasks. A schematic overview of these movements is shown in *Appendices: Performed movements (page 30)*. Prior to the measurements, several unrecorded trials were performed to ensure familiarization with the tasks. During the walking trial, the participant was instructed to walk as normally as possible, ensuring that each force plate was only hit with a foot once. The single leg hop trial involved the participant placing the jumping leg on the first force plate, jumping off the plate, and landing on the second force plate with the same leg. Trials were unsuccessful, thus excluded, if the participant did not step or land on the force plate with

the entire foot. Both left and right leg were captured as jumping leg of the hop test for both healthy subjects and ACL patients.

Machine learning algorithm selection

The development of the NNs was divided into two main steps. In the first step, algorithms from literature were implemented to validate their value against the acquired dataset. This involved implementing existing algorithms and evaluating their performance on the collected data. A flowchart displaying why specific papers were or were not validated can be seen in *Appendices: Selection of literature models, Figure 14 (page 31)*. In the second step, experimentation with different inputs for the new own developed networks was performed. This step involved tweaking the architecture and training of the NNs to optimize their performance.

Initially, one model was developed for the walking movement: a MLP based on literature to ensure fair comparability of the newly designed model with existing algorithms [18]. This model is called the 'Own Leporace Walk' (OLW) model.

To the author's knowledge, the single leg hop movement has not been modeled yet using a movement-specific ML algorithm. Therefore, literature including multiple movements and/or other landing movements had been studied to create a new algorithm [15], [23], [26]–[30]. The algorithm of Stetter et al. was used as starting point, as it seemed reproducible enough and the knee joint forces (KJFs) were the kinetic parameter of interest [27]. The models built based on this algorithm are called 'Own Stetter Hop Push-off' (OSHP) and 'Own Stetter Hop Landing' (OSHL).

Data processing

Before the algorithms could be trained, the obtained data needed to be processed. Data was collected using three different measurement systems: IMUs, OMC and force plates. The overall data flow of this study is visualized in Figure 2.

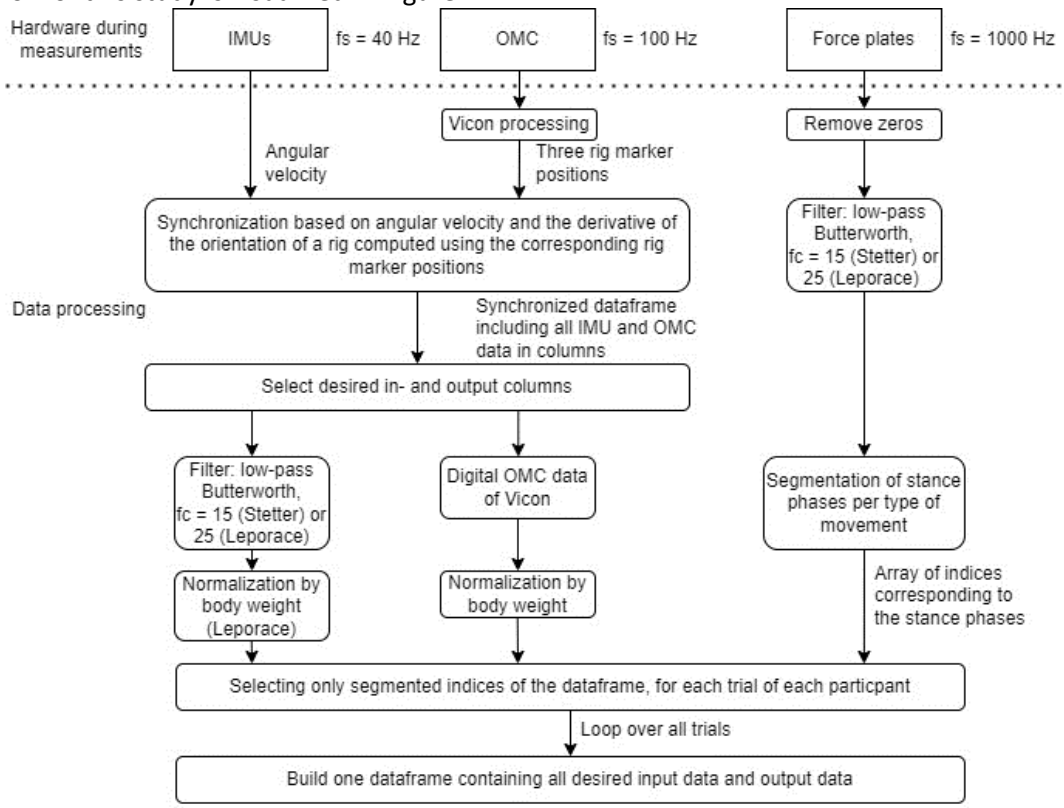

Figure 2 Visualization of data flow. Data was collected using three different measurement systems: IMUs, OMC and force plates. To combine information from the different systems, synchronization was needed. Per algorithm, the desired in- and outputs of the model were selected. Segmentation was executed to select the region(s) of interest within the synchronized dataset.

The raw OMC data were processed using Vicon Nexus Software (2.12.1, Oxford, UK). Prior to any computational steps, gaps in the marker trajectories were filled using spline fill and if needed rigid body fill. The marker trajectories were then smoothed using the Woltring filter (filter mode: MSE, smoothing: 2). After filtering, Vicon's proprietary Dynamic Plug-in Gait Model was run per trial to calculate joint angles and net joint moments (by calculating inverse dynamics using the force plate kinetics) [31]. The c3d files resulting from this last step, containing information on marker locations, GRFs, joint angles and net joint moments, were used to develop the ML algorithms.

The acceleration and angular velocity data from the IMUs were filtered using a zero-lag fourth-order low-pass Butterworth filter with a cut-off frequency of 25 Hz (OLW) or 15 Hz (OSHP and OSHL). The analog GRFs were filtered using a fourth-order, zero-phase Butterworth low-pass filter with a cut-off frequency of 25 Hz (OLW) or 15 Hz (OSHP and OSHL).

To combine information from the different systems, a synchronization step was needed. Synchronization of the IMU and OMC data of all trials was achieved using the angular velocity of the right upper leg IMU and the calculated angular velocity of the corresponding rig. The latter was calculated by determining the orientation of the rig based on the position data of the three rig markers and then taking the derivative of this orientation. Figure 3 and Figure 4 show the background and an example application of these steps.

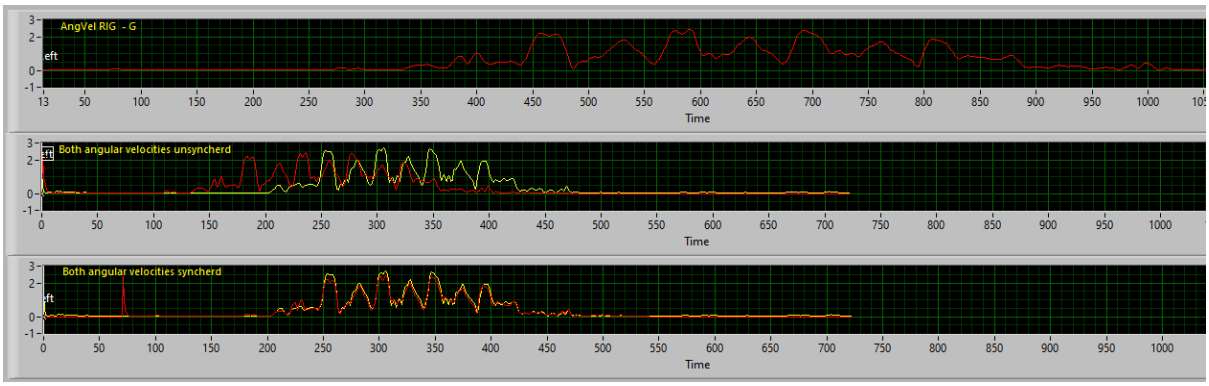

*Figure 3 Visualization of the synchronization method. The calculated angular velocity of the rig is shown in the **upper graph** in red. The **middle graph** displays both the angular velocity of the rig (resampled) in red and the angular velocity measured by the IMU in yellow. The calculated rig angular velocity was resampled to have the same sample frequency as the angular velocity of the IMU. The cross-correlation between these two signals was used to determine the time lag: the highest peak corresponds to the time difference. In the **lower graph**, this synchronization factor was used to shift the Vicon data to suit the IMU data, such that they practically overlap.*

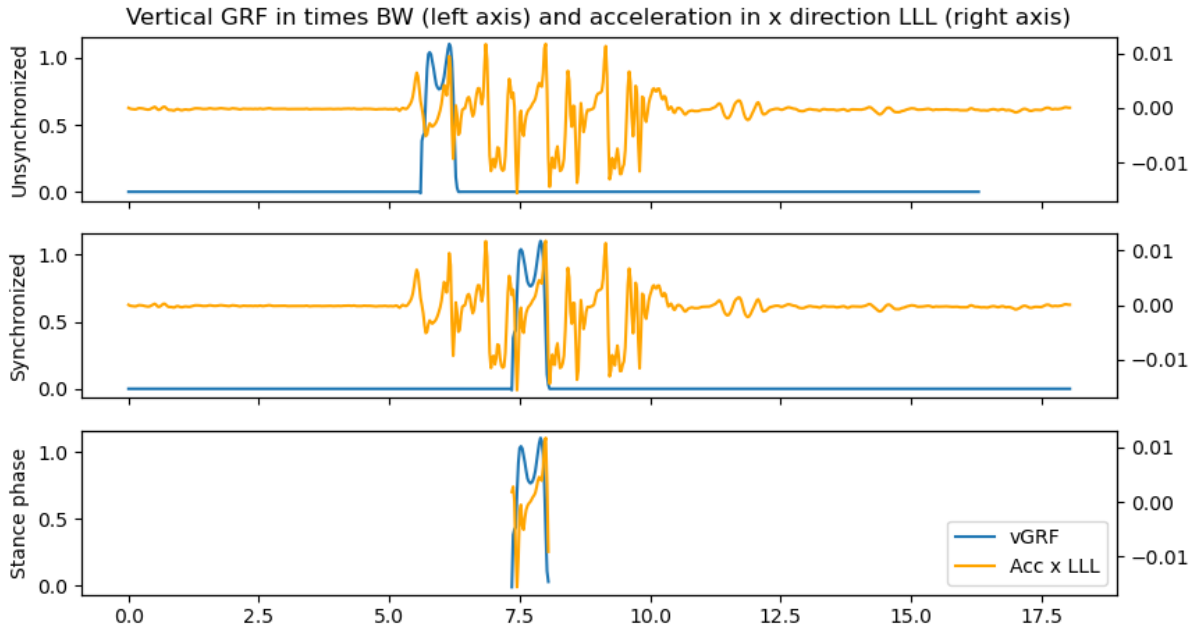

*Figure 4 Illustration of the results of synchronization of one of the trials. Both IMU data (acceleration in x direction of left lower leg (LLL)) and Vicon data (vGRF) are shown (**upper graph**). The Vicon data was shifted with the determined synchronization factor (**middle graph**). After this step, the gait segmentation was performed (**lower graph**).*

Segmentation was executed to select the region(s) of interest within the synchronized dataset. Segmentation of the stance phase of gait was implemented using the methods of Leporace et al. based on a threshold of 10 N in the vertical GRF (vGRF) [18]. Example results of this segmentation are shown in Figure 4 and Figure 5. These stance phases were resampled to 51 samples per stance phase, as Leporace et al. had performed, which allowed for expressing all values as a percentage of stance.

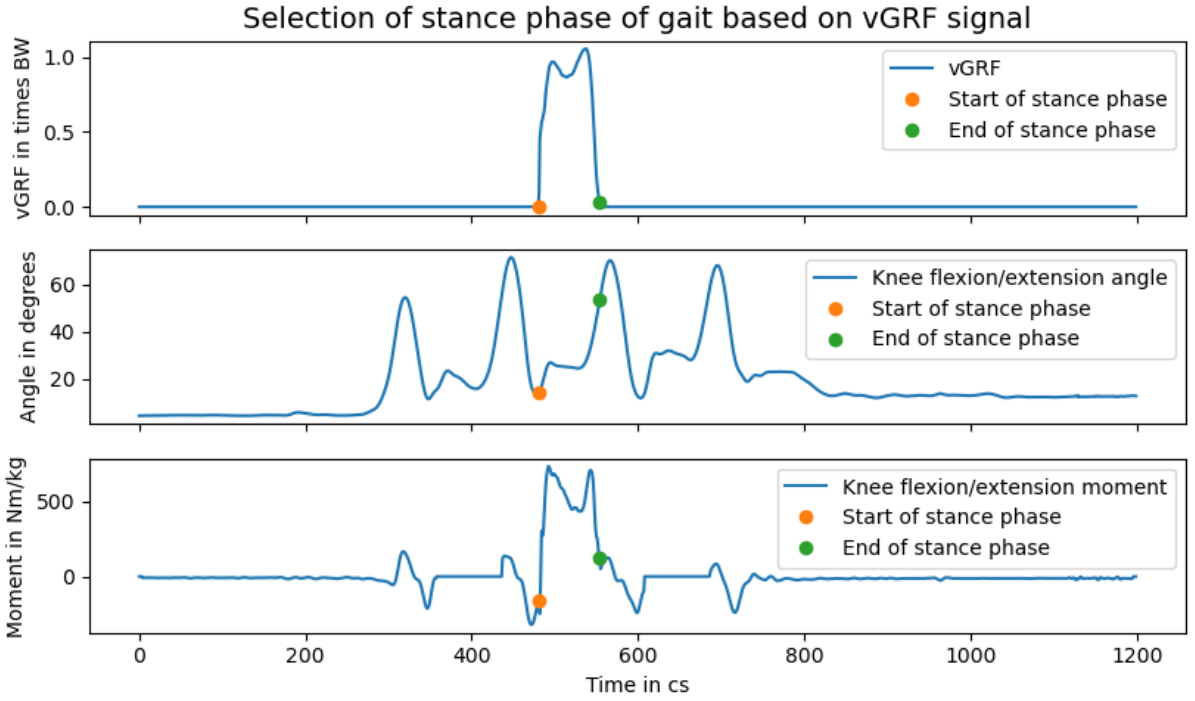

Figure 5 Selection of stance phase of gait (between orange and green dot) based on vGRF signal. This figure shows what parts of the vGRF, knee flexion/extension angle and net knee flexion/extension moment correspond with the stance phase.

For single leg hop tests, two different stance phases were segmented per trial: push-off and landing. Segmentation of both phases during the single leg hop test was performed using a sliding window approach and the threshold value body weight (BW) found by Stetter et al. [27]. The window size for selection of the start of the push-off phase was 150 ms and the window size for selecting the end of the landing phase was 200 ms. The average values of the windows needed to be approximately one time BW. For the push-off and landing phases, the standard deviation of the window values needed to be > 0.05 and < 0.025, respectively. If the deviation was larger than 0.05, the start of the push-off stance phase was found and if the deviation was smaller than 0.025 the end of the landing phase was determined. This way small alterations within the signal were acceptable, but the larger deviations that occur during preparation of the hop and landing of the hop were selected within the segmentations (see Figure 7). The best method for determining the start of the landing phase was chosen by testing two different methods: start landing, which involved looking for the minimum of the second force plate after applying a high pass filter (cut-off frequency=0.5 Hz) on the vGRF signal (see Figure 6), and start landing 2, which used a threshold of 0.05 times BW in the vGRF signal (see Figure 7). Start landing was found to occur at a later instance than start landing 2 in almost every trial, as can for example be seen in Figure 8. Therefore, start landing 2 was implemented for the final segmentation to ensure that important initial contact information was not lost. The end of the push-off phase was selected in a similar way as the not chosen method of the start of the landing phase: the minimum of the high pass filtered vGRF signal of the first force plate (see Figure 6). Both hop stance phases were resampled to 100 samples per stance phase as Stetter et al. also had performed.

Segmentations used as data for the own developed algorithms were resampled to 100 samples as this was easily translated to percentage of stance phase.

Trials could be invalid due to incorrect saving of collected data, an incorrect segmentation of the stance phase or a different measurement set up, which did not allow for proper synchronization. An overview of how many trials were used is shown in *Appendices: Overview of used trials per movement per subject group, Table 9 (page 32)*.

In this study, GRFs were normalized by BW and net joint moments were normalized by BW and height of the participant to enable intersubject comparison [32].

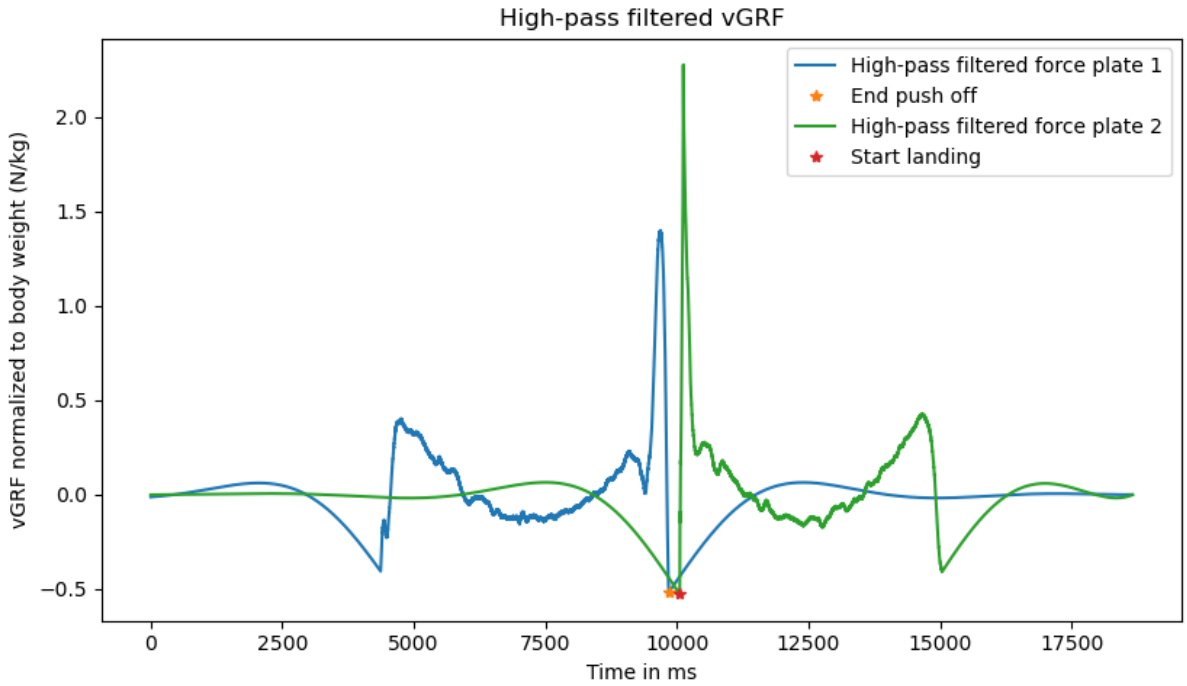

Figure 6 Determining end push-off and start landing by selecting the minima of the high-pass filtered vGRF signal.

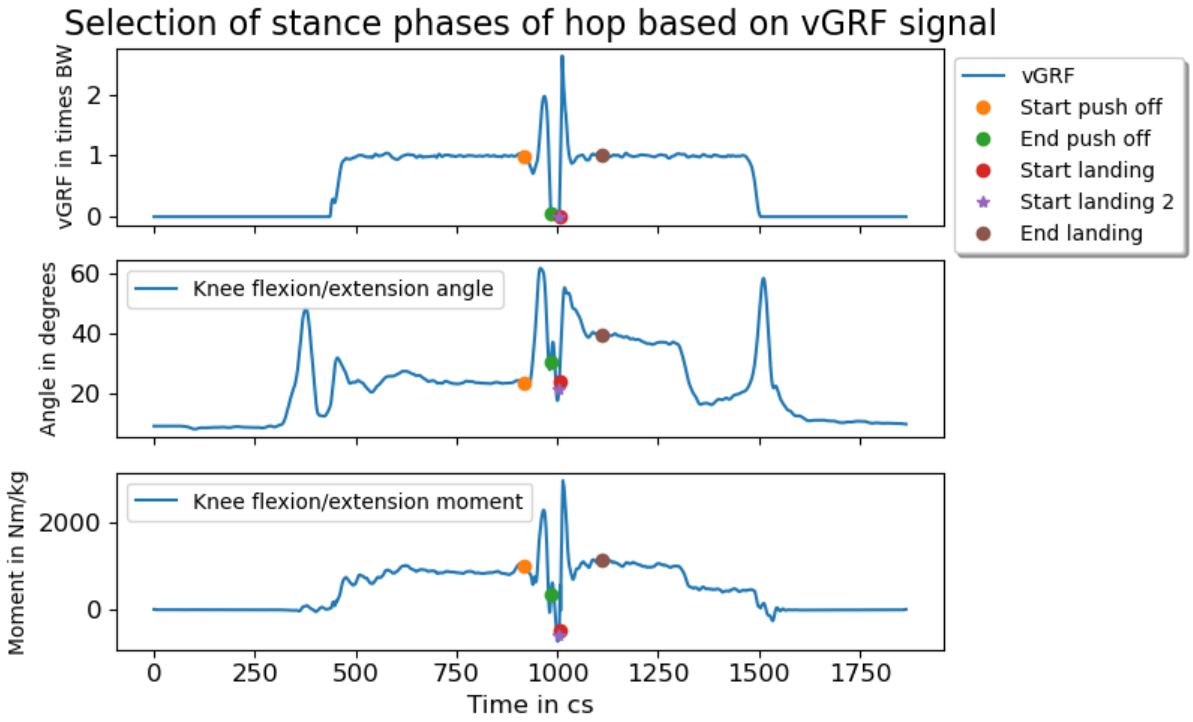

Figure 7 Selection of stance phases of hop based on vGRF signal. This figure shows what parts of the vGRF, knee flexion/extension angle and net knee flexion/extension moment correspond with the stance phases.

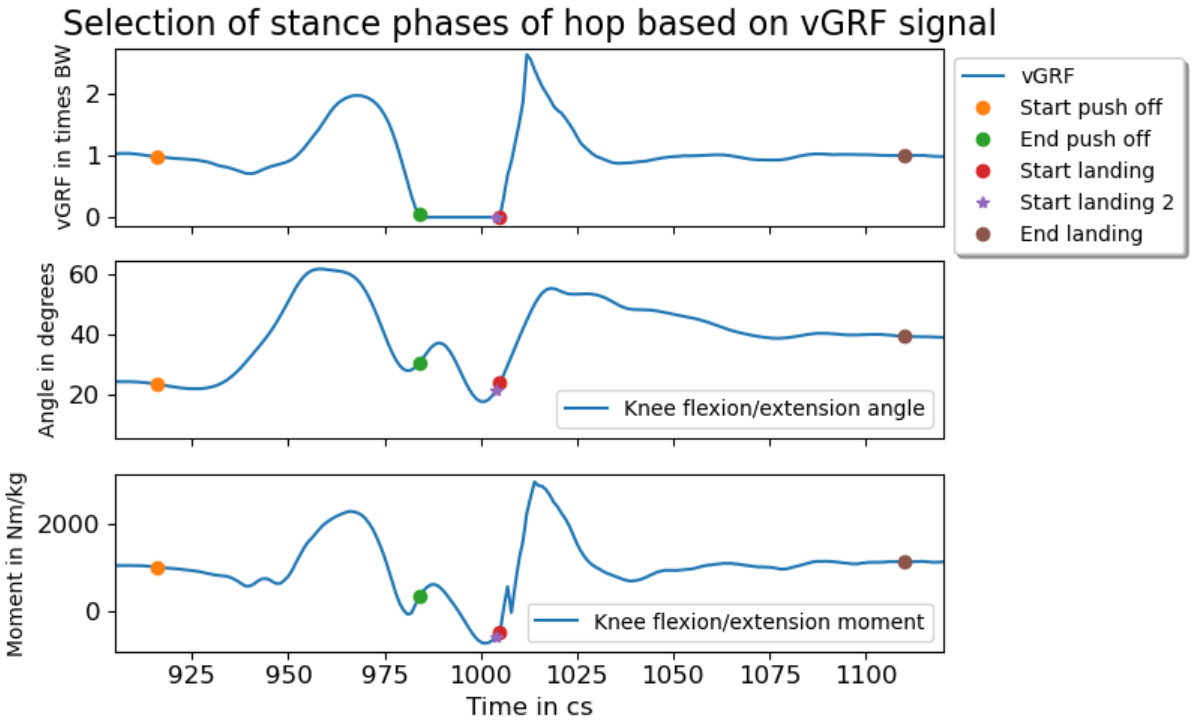

Figure 8 Zoomed in selection of stance phases of hop based on vGRF signal. This figure emphasizes that start landing occurs later than start landing 2.

Development of machine learning algorithms

The NNs developed for this study were built using Python (version 3.9.13, packages: Pyrenn and pyTorch) and relate the IMU signals of a specific movement to the net knee joint moments and 3D GRFs of that movement. To heighten the generalization of the algorithms, all recorded instances of the selected segments were shuffled before being introduced to the ANNs [33].

Although the selected algorithms for validation were partly chosen based on their reproducibility, not all algorithm parameters were known and/or could be implemented as before. The OLW algorithm developed in this paper differed from the original Leporace version in type of optimizer, number of epochs/stop criteria and the type of acceleration used. In this case the Adam optimizer was used, training was done for 1000 epochs and the segment accelerations were implemented instead of the Levenberg-Marquardt backpropagation, a minimum gradient and sensor accelerations. For the OSHP and OSHL algorithms the Adam optimizer was used instead of the Levenberg-Marquardt backpropagation and the models were again trained for 1000 epochs. The activation function of the output layer was unknown from literature but a linear function was implemented for this.

The new developed algorithms (Own Walk (OW), Own Hop Push-off (OHP) and Own Hop Landing (OHL)) used segment accelerations as input and the hidden neurons used a Rectified Linear Unit (ReLU) activation function. The output layer was a linear function and the Adam optimizer was used. A batchsize of 32 was used to speed up training time. One or more dropout layers were included to prevent the model from overfitting, especially because of the relatively small data set.

The choice in input variables strongly affects the predictive ability of a NN [34]. Therefore, a thorough selection of these input variables is essential. Per IMU, the 3D linear accelerations were included in the new algorithms and the percentage of stance phase was added tot the IMU signal matrix. In total, 8 IMUs were used. Therefore, if all IMUs were used as inputs, 25 (8x3 +1) inputs would be available. A preselection of two to three IMUs was made for the new models. All networks were trained from scratch. The IMU signal matrix served as input and the 3D net knee joint moments and 3D GRFs as target output. For the OHP and the OHL algorithm the 3D KJF was also used as an output because this was the only output Stetter et al. had implemented and the outcome of the own developed models now could be compared to the outcome of Stetter et al. [27]. An overview of all these in- and outputs is shown in Table 1.

Leave-one-subject-out cross-validation (LOSO CV) was performed to evaluate how accurate the algorithm works for an unseen participant. This means that the model was trained and validated using data of all but one subject, and it was tested using the data of the remaining subject. The error averaged over all folds corresponds to the prediction error that can be expected for an unseen participant.

Table 1 In- and outputs of the different models.

Model	OLW 1 IMU at lower leg	OW 2 IMUs + 1 joint: lower leg, upper leg and knee	OSHP + OSHL 2 IMUs: lower leg and upper leg	OHP 3 IMUs + 1 joint: lower leg, upper leg, sacrum and knee	OHL 3 IMUs + 1 joint: lower leg, upper leg, sacrum and knee
Input parameters	Acceleration (3D) (normalized by BW)	Acceleration (3D) per IMU or joint (normalized by BW)	Acceleration (3D) per IMU	Acceleration (3D) per IMU or joint (normalized by BW)	Acceleration (3D) per IMU or joint (normalized by BW)
	Velocity (3D) computed from IMU data	Time point of the stance (as % of stance time)	Angular velocity (3D) per IMU	Stance duration (in seconds)	Stance duration (in seconds)
	Displacement (3D) computed from IMU data			Time point of the stance (as % of stance time)	Time point of the stance (as % of stance time)
	First derivative of acceleration (3D) computed from IMU data				
	Stance duration (in seconds)				
	Time point of the stance (as % of stance time)				
Total number of inputs	14	10	12	14	14
Output parameters	GRFs (3D)	GRFs (3D)	KJFs (3D)	GRFs (3D)	GRFs (3D)
		Net knee joint moments (3D)		Net knee joint moments (3D)	Net knee joint moments (3D)
				KJFs (3D)	KJFs (3D)
Total number of outputs	3	6	3	9	9

Hyperparameter optimization

To determine the best hyperparameters of the new models, the Python package Optuna was used [35]. The number of hidden layers, number of neurons per hidden layer, dropout rate per hidden layer, number of training epochs, learning rate and loss function were set as parameters to optimize. Optimization of OW was run by minimizing the sum of six separate Mean Absolute Error (MAE) values (MAE 3D GRF + MAE 3D net knee joint moments). For the OHP and the OHL algorithms, also the MAE of the 3D KJF was taken into account, thus the sum of nine separate MAE values was minimized. The combination with the lowest total MAE after running the optimization for three hours twice, was selected for the model and can be seen in Table 2.

Table 2 An overview of the hyperparameter selection variables and the best scoring (lowest 3D MAE) values.

Hyperparameter	Possible parameter values	Best outcome value		
		OW	OHP	OHL
Number of PCs	[7,8,9] OR [8,9,10,11] (depending on number of original input features)	[7,8,9] → 9	[8,9,10,11] → 11	[8,9,10,11] → 9
Number of hidden layers	[1,2,3]	3	1	3
Number of neurons per hidden layer	[5,10,15,20,25,30,35,40,45,50]	[20,50,45]	10	[10,30,15]
Dropout rate per hidden layer	[0.2,0.3,0.4,0.5]	[0.3,0.3,0.3]	0.5	[0.4,0.4,0.5]
Number of training epochs	[500,1000,1500,2000]	1000	1000	500
Learning rate	[1e-5,1e-1], log steps in between	0.00214688	0.04175844	0.00087434
Loss function	[MSELoss, L1Loss]	L1Loss	MSELoss	MSELoss

Input selection

To prevent the models from overfitting, a principal component analysis (PCA) was performed to only use principal components (PCs) that together explained 80 to 90 percent of the variance. For the new developed algorithms, the number of PCs was included in the hyperparameter optimization (see Table 2). Since the PCs are linear combinations of the initial input parameters, the model still depends on all the input features.

Apart from the hyperparameter optimization and input selection also other techniques were applied to find the best new developed algorithms. An overview of these methods can be found in *Appendices: Model iterations of own developed models (see page 49)*. The exact final architecture of the different networks is shown in Figure 9.

Statistical analysis

To compare the predictions of the different models, the accuracy was assessed for each model using the Pearson's correlation coefficient (r), the root-mean-squared error (RMSE), and the relative RMSE (rRMSE). The Pearson's correlation coefficient is categorized as weak (r<= 0.35), moderate (0.35 <r<=0.67), strong (0.67<r<=0.90) or excellent (r>0.90) correlation [9]. The RMSE and rRMSE were calculated as performed by Ren et al. [36]. All statistical analyses were performed using Python version 3.9.13 (Python Software Foundation).

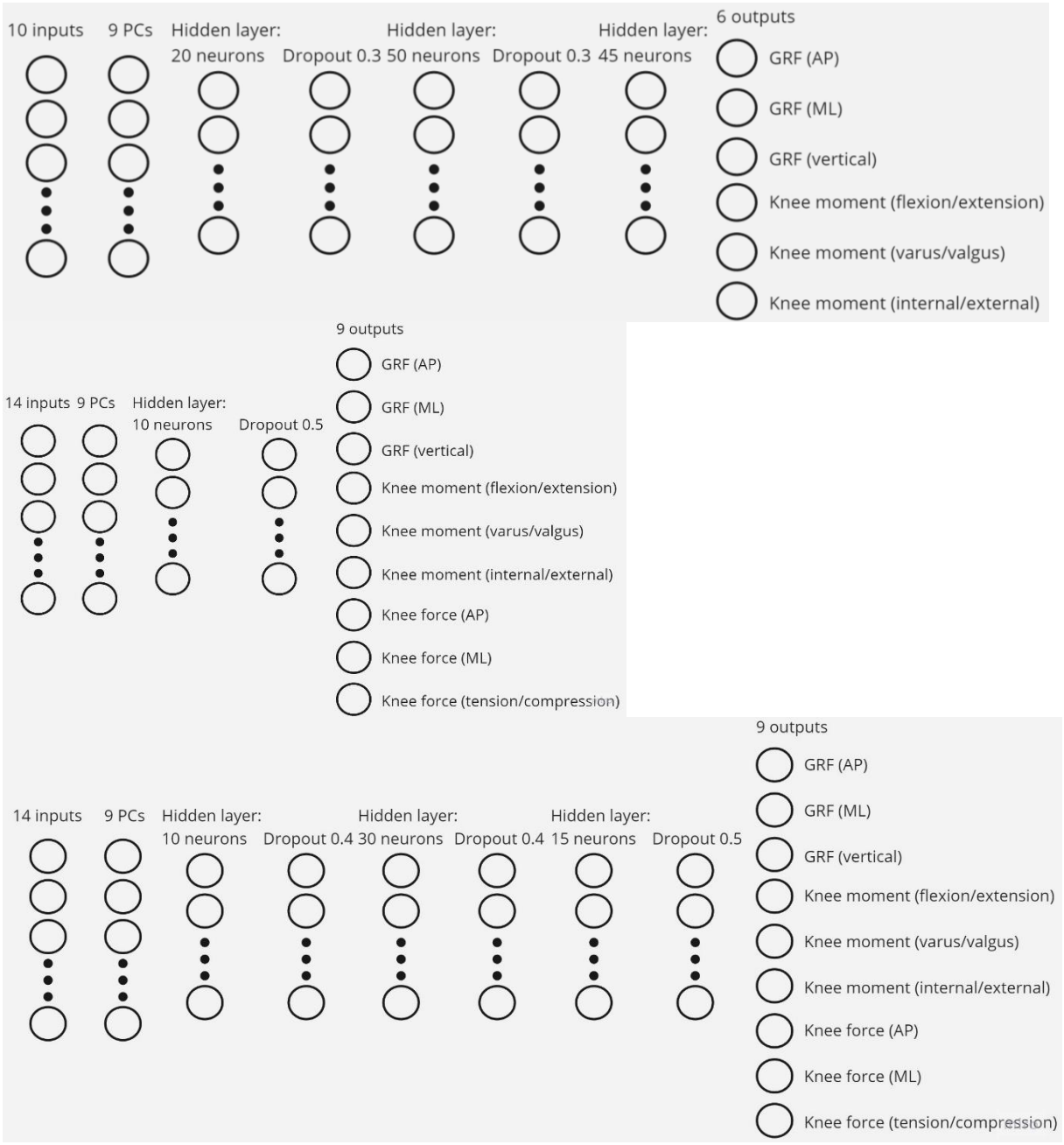

*Figure 9 Architecture of: OW (**upper**), OHP (**middle**) and OHL (**lower**).*

Results

Different ML algorithms are implemented, trained, and tested. The results of the best literature based models and of the new developed algorithms are shown in this section. More detailed results, including reasoning why the presented models were selected, can be found in *Appendices: Methods for and results of obtaining the best validation models (page 33)* and *Appendices: Model iterations of own developed models (page 49)*.

An exploration of the range of output values of both healthy subjects and ACL patients can be seen in *Appendices: Methods for and results of* obtaining the best validation model *(page 33)*. Overall, this range seemed similar between the healthy subjects and the ACL patients, except for a few outliers in both groups.

In Table 3 the outcome performance metrics of the walking models are shown. There were differences between the model performances throughout the LOSO CV of OLW. These can be observed in Figure 10 (the best scoring model) and Figure 11 (the worst scoring model). There are no significant intersubjectgroup performance differences for the OLW algorithm.

The outcome performance metrics on the LOSO CV of the hop models during the push-off and landing phase are visible in Table 4 and Table 5.

*Table 3 Outcome metrics for LOSO CV of right leg of walk models. Standard deviation is shown between brackets. h = trained and tested on healthy subjects (001-1; 002-1; 002-2; 004-1; 004-2; 006-1; 006-2; 007-1; 007-2; 008-2). p = trained and tested on ACL patients (p01-1; p02-1; p02-2; p04-1; p04-2; p05-1; p05-2; p07-1; p07-2; p08-1) . hp = trained and tested on combination of healthy subjects and ACL patients. Any significant intersubjectgroup differences are in **bold**. Light orange: r <= 0.65. Darker orange: r<= 0.35. Grey: not known/not applicable.*

		OLW			OW		
		RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r
GRF	A-P h	0.0490 (0.0095)	9.30 (1.69)	0.9049 (0.0452)	0.0523 (0.0113)	9.81 (1.41)	0.8959 (0.0451)
	A-P p	0.0513 (0.0126)	10.93 (1.81)	0.8886 (0.0524)	0.0504 (0.0107)	10.40 (1.69)	0.8979 (0.0314)
	A-P hp	0.0466 (0.0090)	9.39 (1.49)	0.9132 (0.0203)	0.0526 (0.0144)	10.14 (1.96)	0.8781 (0.0669)
	M-L h	0.0232 (0.0034)	15.61 (2.25)	0.7024 (0.1022)	0.0177 (0.0050)	12.94 (3.11)	0.6996 (0.2192)
	M-L p	0.0188 (0.0055)	15.20 (4.14)	0.7266 (0.0776)	0.0193 (0.0087)	16.02 (6.68)	0.7641 (0.0410)
	M-L hp	0.0209 (0.0039)	15.78 (3.13)	0.7078 (0.0898)	0.0201 (0.0060)	14.70 (3.98)	0.7376 (0.1595)
	Vertical h	0.2156 (0.0382)	13.36 (2.66)	0.6912 (0.1504)	0.2030 (0.0718)	11.97 (3.68)	0.7323 (0.1811)
	Vertical p	0.1982 (0.0281)	12.83 (1.69)	0.7240 (0.0876)	0.1797 (0.0207)	11.94 (1.38)	0.7762 (0.0870)
	Vertical hp	0.2078 (0.0288)	13.03 (1.70)	0.6982 (0.0868)	0.2038 (0.0599)	13.37 (3.46)	0.6902 (0.2058)
	Norm h	0.2173 (0.0382)	13.43 (2.62)	0.6890 (0.1526)			
	Norm p	0.2012 (0.0280)	13.02 (1.77)	0.7148 (0.0898)			
	Norm hp	0.2096 (0.0291)	13.13 (1.74)	0.6943 (0.0886)			
Net knee moment	F-E h				0.0235 (0.0085)	14.47 (3.97)	0.6598 (0.1691)
	F-E p				0.0322 (0.0057)	28.09 (6.71)	0.3859 (0.0822)
	F-E hp				0.0252 (0.0071)	19.12 (6.35)	0.5637 (0.2318)
	V-V h				0.0171 (0.0107)	22.27 (7.64)	0.5896 (0.1502)
	V-V p				0.0166 (0.0039)	23.62 (5.05)	0.6526 (0.0737)
	V-V hp				0.0196 (0.0088)	26.03 (13.89)	0.5885 (0.2637)
	I-E h				0.0047 (0.0009)	14.43 (1.91)	0.8687 (0.1041)
	I-E p				0.0039 (0.0013)	17.61 (1.97)	0.8192 (0.0640)
	I-E hp				0.0046 (0.0017)	18.23 (6.38)	0.7738 (0.1486)

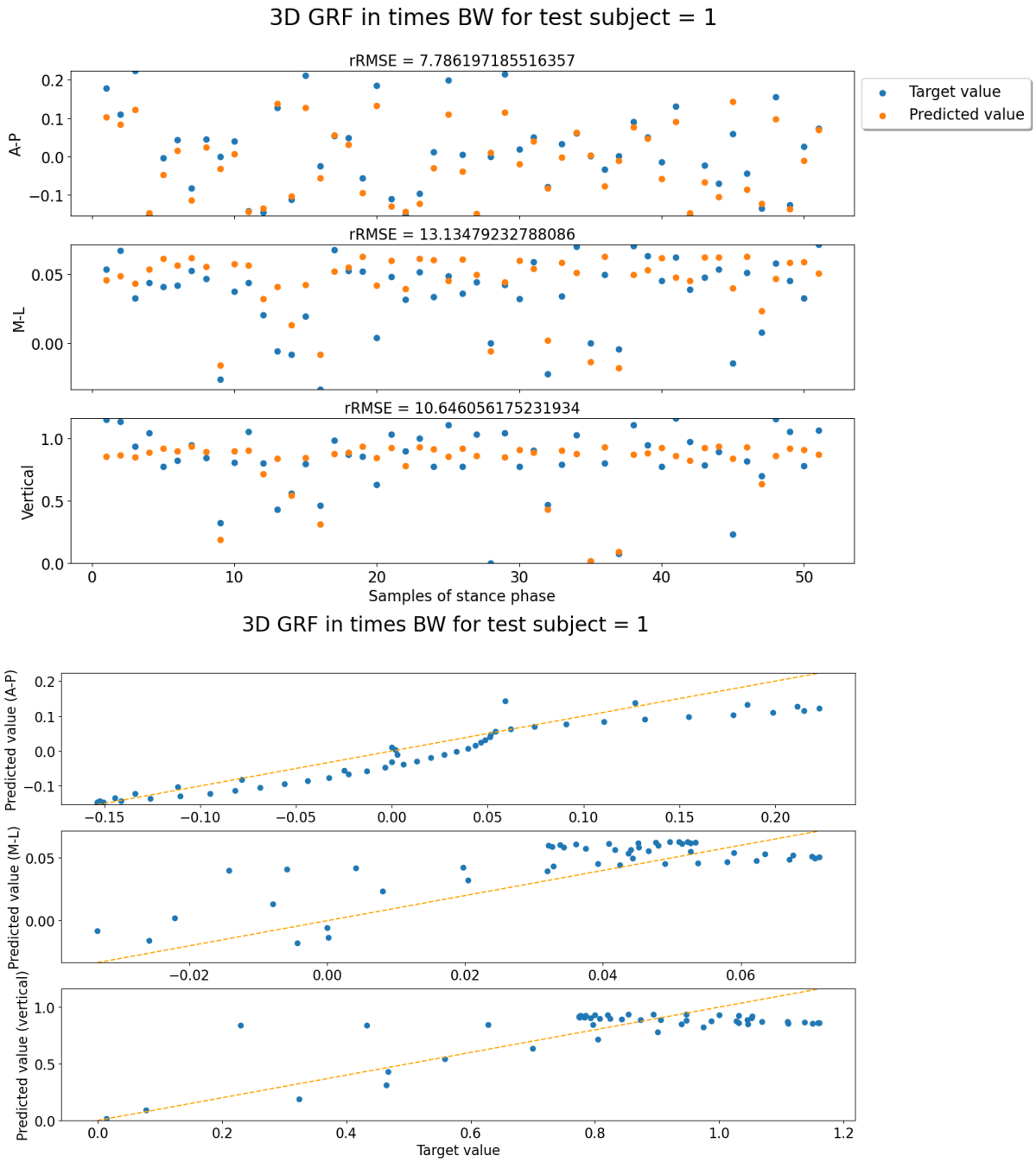

Figure 10 Illustration of predicted versus target GRFs, for the best scoring test subject based on total rRMSE value (subject 1) of the LOSO CV evaluation of OLW.

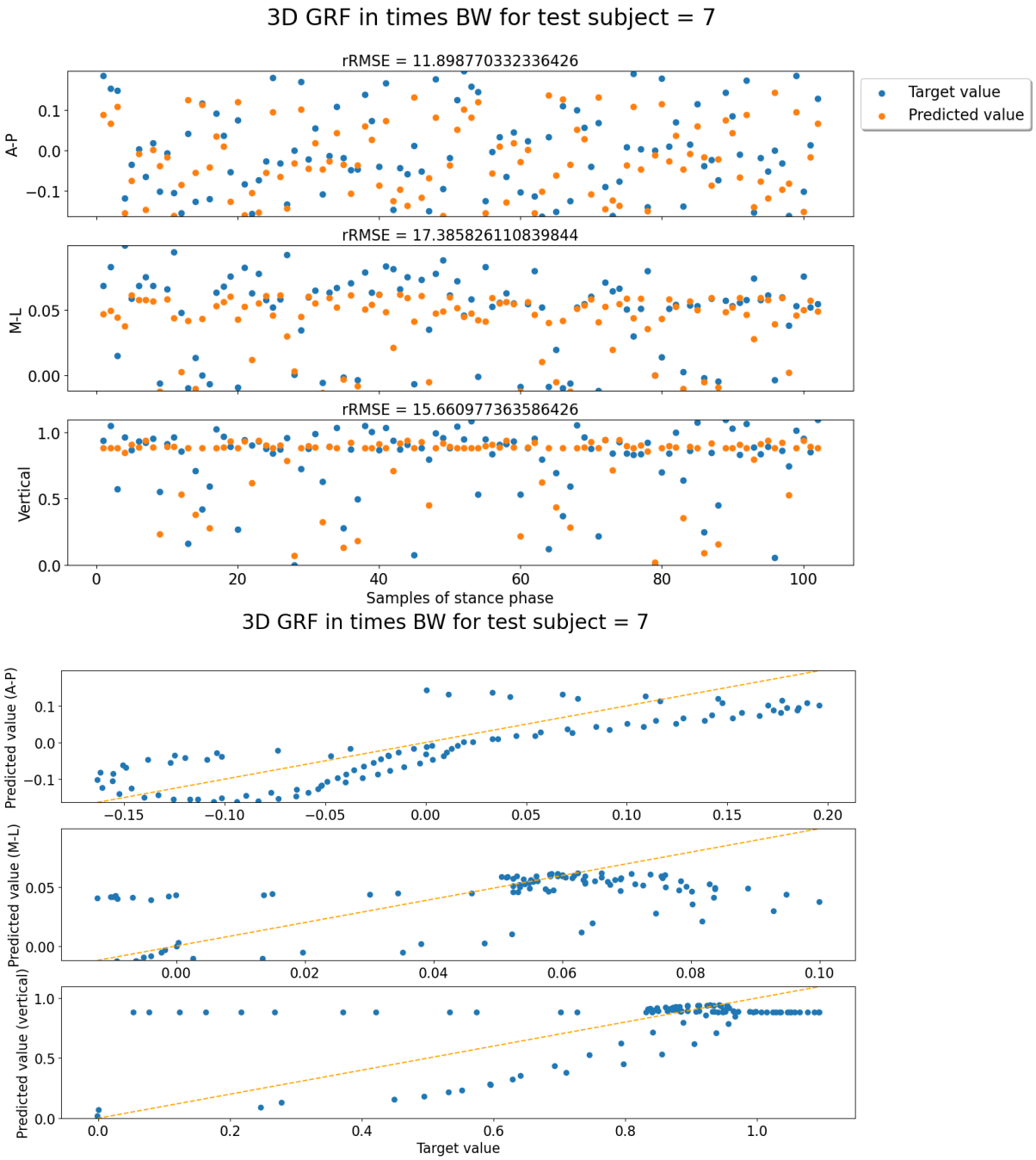

Figure 11 Illustration of predicted versus target GRFs, for the wort scoring test subject based on total rRMSE value (subject 7) of the LOSO CV evaluation of OLW.

*Table 4 Outcome metrics for LOSO CV of hop push-off models of right leg. Standard deviation is shown between brackets. h = trained and tested on healthy subjects (001-1; 002-1; 002-2; 003-1; 003-2; 004-1; 005-1; 005-2; 006-1; 006-2; 007-2; 008-2). p = trained and tested on right ACL leg of patients (p01-1; p01-2; p02-1; p02-2; p05-1; p05-2; p07-2; p08-1). Any significant intersubjectgroup differences are in **bold**. Light orange: r <= 0.65. Darker orange: r <= 0.35. Grey: not known/not applicable.*

		OSHP			OHP		
		RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r
KJF	A-P h	0.1536 (0.0432)	14.28 (6.87)	0.8432 (0.0676)	0.1348 (0.0273)	12.49 (1.46)	0.9170 (0.0192)
	A-P p	0.1681 (0.0411)	17.94 (6.43)	0.8083 (0.0312)	0.1705 (0.0217)	18.79 (4.06)	0.9182 (0.0654)
	M-L h	0.0958 (0.0388)	15.82 (6.39)	0.8292 (0.0604)	0.1663 (0.0860)	28.04 (27.90)	0.7933 (0.2777)
	M-L p	0.1414 (0.0566)	26.21 (10.87)	0.7969 (0.0644)	0.0702 (0.0393)	13.86 (7.68)	**0.8937 (0.0098)**
	T-C h	0.1652 (0.0526)	7.66 (2.74)	0.9226 (0.0410)	0.1734 (0.0432)	8.40 (1.91)	0.8930 (0.0518)
	T-C p	0.1986 (0.0524)	9.84 (2.49)	0.8824 (0.0379)	0.1799 (0.0764)	9.13 (3.96)	0.8623 (0.1541
Net knee moment	F-E h				0.0400 (0.0155)	17.36 (6.63)	**0.8233 (0.0943)**
	F-E p				0.0596 (0.0284)	36.77 (30.72)	0.5338 (0.4530)
	V-V h				0.0498 (0.0188)	34.71 (19.31)	0.4654 (0.3074)
	V-V p				0.0293 (0.0103)	22.99 (10.59)	**0.7919 (0.1285)**
	I-E h				0.0198 (0.0071)	39.08 (13.69)	-0.3023 (0.4344)
	I-E p				0.0114 (0.0074)	29.64 (8.57)	-0.1601 (0.6057)
GRF	A-P h				0.0615 (0.0163)	18.99 (4.64)	0.7274 (0.0946)
	A-P p				0.0542 (0.0175)	17.36 (5.11)	0.7450 (0.1780)
	M-L h				0.0304 (0.0159)	25.57 (7.10)	0.1486 (0.2835)
	M-L p				0.0248 (0.0063)	22.19 (1.70)	0.0033 (0.3695)
	Vertical h				0.2058 (0.0383)	8.36 (1.56)	0.9141 (0.0428)
	Vertical p				0.1977 (0.0607)	8.97 (3.14)	0.8714 (0.1354)

*Table 5 Outcome metrics for LOSO CV for hop landing models of right leg. Standard deviation is shown between brackets h = trained and tested on healthy subjects (001-1; 002-1; 002-2; 003-1; 003-2; 004-1; 005-1; 005-2; 006-1; 006-2; 007-2; 008-2). p = trained and tested on right ACL leg of patients (p01-1; p01-2; p02-1; p02-2; p05-1; p05-2; p07-2; p08-1). Any significant intersubjectgroup differences are in **bold**. Light orange: r <= 0.65. Darker orange: r<= 0.35. Grey: not known/not applicable.*

		OSHL			OHL		
		RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r
KJF	A-P h	0.1778 (0.0369)	13.65 (2.23)	0.7319 (0.2008)	0.1348 (0.0273)	12.49 (1.46)	**0.9170 (0.0192)**
	A-P p	0.1898 (0.0139)	14.39 (3.52)	0.6899 (0.2057)	0.1534 (0.0121)	11.94 (2.15)	0.8561 (0.0371)
	M-L h	0.1449 (0.0498)	21.91 (5.48)	0.5924 (0.2200)	0.1663 (0.0860)	28.04 (27.90)	0.7933 (0.2777)
	M-L p	0.1723 (0.0709)	26.68 (9.66)	0.5717 (0.2005)	0.1131 (0.0517)	21.27 (6.91)	0.7314 (0.0787)
	T-C h	0.2140 (0.1062)	8.47 (3.16)	0.7786 (0.1754)	0.1734 (0.0432)	8.40 (1.91)	0.8930 (0.0518)
	T-C p	0.2073 (0.0463)	8.04 (0.95)	0.8102 (0.0868)	0.1505 (0.0417)	6.55 (1.47)	0.8818 (0.0477)
Net knee moment	F-E h				0.0400 (0.0155)	17.36 (6.63)	0.8233 (0.0943)
	F-E p				0.0657 (0.0230)	26.93 (26.08)	0.6690 (0.1893)
	V-V h				0.0498 (0.0188)	34.71 (19.31)	0.4654 (0.3074)
	V-V p				0.0325 (0.0138)	22.88 (5.13)	0.5734 (0.1142)
	I-E h				0.0198 (0.0071)	39.08 (13.69)	-0.3023 (0.4344)
	I-E p				0.0069 (0.0012)	32.58 (13.29)	0.1615 (0.4077)
GRF	A-P h				0.0615 (0.0163)	18.99 (4.64)	0.7274 (0.0946)
	A-P p				0.0659 (0.0128)	11.19 (2.30)	**0.8750 (0.0315)**
	M-L h				0.0304 (0.0159)	25.57 (7.10)	0.1486 (0.2835)
	M-L p				0.0272 (0.0091)	17.79 (3.29)	0.3553 (0.1565)
	Vertical h				0.2058 (0.0383)	8.36 (1.56)	0.9141 (0.0428)
	Vertical p				0.1650 (0.0334)	6.38 (1.07)	0.8960 (0.0299)

Discussion

The purpose of this study was to analyze how accurate 3D GRFs and/or net knee joint moments can be determined during single leg hop tests and normal walking in ACL patients. The ML models were tested both on a healthy and an ACL patient dataset, consisting of IMU, OMC and force plate data.

The most important finding of the present study was that M-L GRFs and all net knee moments are still difficult to accurately estimate during a single leg hop (r <= 0.65).

For walking, the estimated GRF strongly correlates with the target GRF (r=0.90 A-P; r=0.70 M-L; r=0.73 vertical). Similar to other studies, Pearson's correlation coefficients and rRMSE values were calculated and an overview of the comparison of these values to other research is shown in Table 6. Although the correlation between the predicted and target output is strong, it can be seen that the walking models developed during this study perform worse than those in previous research.

Table 6 Comparison of the prediction error in GRF during gait of previous studies and the validated and developed model (OW and OLW). Grey areas are unknown.

	Pearson's correlation coefficient			rRMSE (% BW)		
	A-P	M-L	vertical	A-P	M-L	vertical
OW (healthy)	0.896	0.700	0.732	9.81 (1.41)	12.94 (3.11)	11.97 (3.68)
OLW (healthy)	0.905	0.702	0.691	9.30 (1.69)	15.61 (2.25)	13.36 (2.66)
Leporace et al. 2018 [18]	0.97	0.80	0.98			
Leporace et al. 2015 [37] 1 MLP for 3 outcomes	0.969	0.801	0.968			
Jiang et al. 2020 [38] Interparticipant			0.97			7.15
Dorschky et al. 2020 [16]	0.971		0.980	3.76 (0.31)		9.99 (0.76)
Oh et al. 2013 [34] Single support phase	0.978	0.736	0.988	2.6 (0.7)	26.2 (4.0)	11.3 (2.1)
Lim et al. 2019 [10]	0.98		0.96	6.16-6.70 (1.76-2.99)		6.26-8.21 (1.24-3.63)

To the best of the author's knowledge, this is the first study to present a movement-specific hop test model. Regarding knee force, the models of the current study scored similar to Stetter et al. in AP and tension/compression direction, but significantly higher than Stetter et al. in the M-L direction (see Table 7) [27]. A possible reason for this might be that the current models are movement specific and that the M-L KJFs within a hop test are very different from e.g. the KJFs during walking or fast running that were also included in the trained model of Stetter et al. In the current research, the same trend is visible as Stetter et al. had noticed: There is a drop in the estimation accuracy for M-L forces, both for the walking and the hop models [27].

Table 7 Comparison of the prediction error in KJF during push-off phase of a single leg hop test of Stetter and the validated and developed model (OHP and OSHP).

	Pearson's correlation coefficient			rRMSE (% BW)		
	A-P	M-L	T-C	A-P	M-L	T-C
OHP (healthy)	0.917	0.793	0.893	12.49 (1.46)	28.04 (27.90)	8.40 (1.91)
OHP (patients)	0.918	0.894	0.862	18.79 (4.06)	13.86 (7.68)	9.13 (3.96)
OSHP (healthy)	0.843	0.829	0.923	14.28 (6.87)	15.82 (6.39)	7.66 (2.74)
OSHP (patients)	0.812	0.789	0.882	17.94 (6.43)	26.21 (10.87)	9.84 (2.49)
Stetter et al. 2019 [27]	0.89	0.31	0.92	17.4 (5.5)	45.9 (19.7)	15.4 (6.6)

When comparing the different subject groups (healthy participants vs. ACL patients) there are differences in accuracy visible. However, sometimes this is in the advantage of the patient group and other times in the advantage of the healthy participants. Since there are no decisive differences between the performance of the models of the two subject groups, it can be stated that ACL patients are a proper population to use in developing and testing of ML models and eventually apply these models.

It is striking how in this research, the A-P direction can be estimated most accurate in the OLW algorithm, followed by the M-L direction and the vertical direction scores lowest. This is not as was expected from literature: Stetter et al. had noticed a drop in the estimation accuracy for M-L forces [27]. On the other hand, the OW algorithm does show the expected lower accuracy in estimating the M-L GRFs.

Caution must be exercised in interpreting these results due to the limited amount of data available compared to e.g. Leporace et al. and Stetter et al. [18], [27]. While this study only reached a maximum of two stance phases per subject per leg, Leporace et al. achieved four stance phases per subject per leg, with a doubled number of subjects and Stetter et al. trained their model using 198 trials. To compensate for the relative lack of data, several potential methods can be considered, including data simulation [16], utilizing more input data (e.g., a SINN [39]), predicting a single output per model (reducing the number of trained relationships), using a simplified model with fewer trainable parameters, using a naïve model (e.g. Charry et al. [40]) or developing subject specific models (because the intersubject differences are too large for the dataset size). In the current study a SINN, predicting a single output per model and subject specific models were applied.

All outcome metrics in this study rely on a comparison between the estimated values from the model and the ground truth values. In this study, the kinetics derived from the Vicon model were selected as the ground truth. As there is no authoritative gold standard in this field, we thus studied the model's proximity to the Vicon system. Future research should also focus on obtaining a quantitative measure that does not rely on a 'not golden standard'. An example of such a measure could be dynamic stability. A small research into this measure was performed as side study and can be found in *Appendices: Literature research 'dynamic stability' (page 54)*.

Methodological limitations (experiments)

Certain limitations of this study need to be considered when interpreting the results. Regarding data acquisition, these drawbacks include, but are not limited to, the precision of sensor and marker placement and the execution of the selected movements. Firstly, the placement of sensors and markers was carried out by two researchers with limited prior knowledge of the Plug-In Gait marker placement guidelines. Although the localization of marker locations was completed with care, employing more experienced researchers could enhance the precision of this process. Additionally, the movement of sensors and markers due to fabric movement should be taken into account [41]. Fortunately, a sensor-to-segment calibration was executed on the IMU data, mitigating the impact of uncertainties within sensor placement. Lastly, the performance of the hop test was constrained by a fixed distance between the force plates. This standardized approach aimed to reduce the variation between participants, but it differs from clinical practice where patients are typically instructed to jump as far as possible instead. As a result, the forces and net moments experienced during a restricted hop may not fully represent those observed in a maximal distance hop. Therefore, it can be questioned whether the constrained hop test adequately simulates ACL rehabilitation exercises.

Methodological limitations (data analyses)

Looking at the data analyses part, it can be discussed what time window gives the most optimal segmentation, particularly concerning the single leg hop movement. It is not desired to let the model overfit on the balancing part before and after the actual hop, but it is also of interest to have information regarding the balancing part. Therefore, it is not possible to set fixed timeslots that hold for each subject. In this study, variability in the vGRF was used as a measure for adjustment of position and thus belonging to the push-off or landing phase. However, this method was influenced by filtering of the force plate data. The time of push-off seems to be delayed as a consequence of this [42]. Another segmentation problem is what threshold value to base the segmentation on. For the stance phase of gait, different threshold values are used in literature e.g. 10 N vGRF, 20 N vGRF or 0.05 times body weight [9], [18], [30], [37]. For running, the gold standards for determining IC and TO events are respectively 10 N vGRF and 25 N vGRF [43], but for normal gait there has no such standard been found. This makes gait analysis research less consistent.

Pogson et al. described that it was better to not standardize the length of the signal by rescaling to save the time dynamics [13]. Zero padding would be a solution for this. Since the focus of this study did not lay in time dynamic based models, this is not a big concern for the outcomes of the current study but should be kept in mind for future research.

For the ML models it is best to perform a normalization, as defined by Huang et al.: “a general transformation which ensures that the transformed data has certain statistical properties” [44]. In this study, the input features were transformed per input feature. This means that the size of acceleration in the x direction was scaled independently of the size of the acceleration in the y direction. However, if you transform the input features, thus the three input directions, separately from each other, the actual directions of e.g., acceleration vectors likely change and may not be representative anymore. The ML model might not have difficulties with training this model, but the inputs are not ‘real’ data inputs anymore, which makes it difficult for us humans to see how the ML model created logical connections between them. How tridimensional data can be processed best to use in ML models can still be questioned for this reason.

Clinical implications

Research on prediction of kinetics using IMU data and ML model(s) had mostly been restricted to healthy subjects during the walk or running movement. In contrast to previous work, this study focused on the hop test and ACL patients as well as on walking and healthy subjects. Still, the clinical course of events can only be adapted if more ACL rehabilitation-based research will be performed on this topic.

In this research, contrasting to Stetter et al., De Brabandere et al. and Zhang et al., a different model is developed for each different movement type with the goal of higher intersubject generalization [9], [23], [27], [28]. For clinical practice, this might be received as an extra selection a physiotherapist needs to make before being able to analyze the results. However, this should not impose additional workload to the physiotherapist, as the practitioner needs to know which activity was performed to determine more muscle-specific forces instead of only looking at the net joint moments and GRFs.

Future research

Future investigations should focus more on ACL rehabilitation movements. For example, the side hop movement where sidewards forces play a more pronounced role might be interesting to evaluate. It should also be explored what other movements of the ACL rehabilitation process play a key role in the overall evaluation of the progress within this process. When suitable movements are selected and ML models for these movements are created that can estimate kinetics accurately enough, the outcome of these models can be used as quantitative measures during rehabilitation. The physiotherapist will see graphs of e.g. the knee moment as result of the ML model and the known healthy knee moment is plotted against this graph. These graphs provide an easy way of monitoring progress within a patient and as soon as a database of more patient data has been built, they can be used to compare progress of a patient against a norm and to give more muscle specific exercises as homework to the patient.

In this research, the stance phase was segmented based on force plate data. However, if this model is used in clinical practice, only IMU data will be available. Therefore, IMU based segmentation techniques should be explored more in future research. An example of this can be seen in the research of Benson et al., even though this entailed the running movement [43].

The models designed in this research could be further optimized by implementing biomechanical relations, e.g. the second law of Newton for better estimation of the vGRF [45]. Specific parameters of the formula used for this, could be retuned as well. An additional improvement in performance could be reached by first estimating the kinematics, and consequently determining the kinetics. This two-staged approach was advised by Mundt et al [46]. To improve the joint angle estimation using IMUs, position estimation could be used, e.g. by using the ultrawideband technique or ultrasound [47], [48]. In clinical practice, this advanced joint angle estimation may enhance the accuracy of estimated kinetic parameters.

Conclusion

All in all, two existing models (from Leporace et al. and Stetter et al.) were validated and fine-tuned against a dataset of healthy subjects and ACL patients. On top of that, three new models were developed to be able to predict GRFs and net knee joint moments, not only during walking but also during single leg hop tests. The model evaluation of OW based on LOSO CV among the ACL patients showed rRMSE values of 9.49, 16.28 and 11.79 percent during stance phase of walking for anterior/posterior, medial/lateral and vertical direction of GRF respectively. The most important finding of the present study was that M-L GRFs and all net knee moments are still difficult to accurately estimate during a single leg hop. However, it showed that ACL patients are a suitable population to develop and evaluate ML algorithms on. The current method is a first step of providing quantitative data for monitoring the ACL rehabilitation process. To make this even more ACL-focused, other ACL rehabilitation movements should be modeled as well, especially movements requiring relatively large M-L forces.

Acknowledgements

First, I would like to thank my supervisors (Sanchana Krishnakumar, Bert-Jan van Beijnum, Jaap Buurke, Edwin van Asseldonk), for the feedback and guidance they provided throughout the project. Again, I would like to thank Sanchana. This time for all the subject and patient measurements we performed together because “teamwork makes the dreamwork” and the hours and hours we spent on processing these measurements. Talking of experiments and Vicon processing, I also want to express my gratitude to Leendert Schaake for providing solutions and answers to all our Vicon questions and for changing the force plate configuration halfway during each measurement session. I want to thank Vinish Yogesh for his (acute) assistance during some of the measurements and thanks to Ynze Keulstra for assisting in the patient measurements. Chris Baten, thanks for developing the synchronization software, the bony landmark calibration software, and all shared opinions regarding ML algorithms in general. I also want to thank Frank Wouda and Bouke Scheltinga, for the one-time one-on-one discussions I had regarding the implementation of specific algorithms. Lastly, I need to thank my fellow master students for all lovely tea breaks and/or in-depth discussions.

References

[1] A. M. Joseph, C. L. Collins, N. M. Henke, E. E. Yard, S. K. Fields, and R. D. Comstock, "A Multisport Epidemiologic Comparison of Anterior Cruciate Ligament Injuries in High School Athletics," *J. Athl. Train.*, vol. 48, no. 6, p. 810, Nov. 2013, doi: 10.4085/1062-6050-48.6.03.

[2] "ACL Tear: Symptoms, Recovery & Treatment." https://my.clevelandclinic.org/health/diseases/16576-acl-anterior-cruciate-ligament-tears (accessed Jul. 26, 2022).

[3] R. Cooper, "ACL Rehabilitation Guide," 2018.

[4] T. Purevsuren, B. Khuyagbaatar, K. Kim, and Y. H. Kim, "Investigation of Knee Joint Forces and Moments during Short-Track Speed Skating Using Wearable Motion Analysis System," *Int. J. Precis. Eng. Manuf.*, vol. 19, no. 7, pp. 1055–1060, 2018, doi: 10.1007/s12541-018-0125-9.

[5] A. M. Alanen, A. M. Räisänen, L. C. Benson, and K. Pasanen, "The use of inertial measurement units for analyzing change of direction movement in sports: A scoping review:," *Int. J. Sports Sci. Coach.*, vol. 16, no. 6, pp. 1332–1353, Apr. 2021, doi: 10.1177/17479541211003064.

[6] W. Thompson *et al.*, "Estimation of Lower-body Kinetics from Loading Profile and Kinematics Alone, Without Measured Ground Reaction Forces," Jul. 2018.

[7] E. Dorschky, M. Nitschke, A. K. Seifer, A. J. van den Bogert, and B. M. Eskofier, "Estimation of gait kinematics and kinetics from inertial sensor data using optimal control of musculoskeletal models," *J. Biomech.*, vol. 95, p. 109278, Oct. 2019, doi: 10.1016/J.JBIOMECH.2019.07.022.

[8] M. Lee and S. Park, "Estimation of Three-Dimensional Lower Limb Kinetics Data during Walking Using Machine Learning from a Single IMU Attached to the Sacrum," *Sensors 2020, Vol. 20, Page 6277*, vol. 20, no. 21, p. 6277, Nov. 2020, doi: 10.3390/S20216277.

[9] B. J. Stetter, F. C. Krafft, S. Ringhof, T. Stein, and S. Sell, "A Machine Learning and Wearable Sensor Based Approach to Estimate External Knee Flexion and Adduction Moments During Various Locomotion Tasks," *Front. Bioeng. Biotechnol.*, vol. 8, p. 9, Jan. 2020, doi: 10.3389/FBIOE.2020.00009/BIBTEX.

[10] H. Lim, B. Kim, and S. Park, "Prediction of Lower Limb Kinetics and Kinematics during Walking by a Single IMU on the Lower Back Using Machine Learning," *Sensors (Basel).*, vol. 20, no. 1, Jan. 2019, doi: 10.3390/S20010130.

[11] 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 (Basel).*, vol. 21, no. 4, pp. 1–19, Feb. 2021, doi: 10.3390/S21041553.

[12] 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.

[13] 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.

[14] F. J. Wouda *et al.*, "Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors," *Front. Physiol.*, vol. 9, no. MAR, p. 218, Mar. 2018, doi: 10.3389/FPHYS.2018.00218/BIBTEX.

[15] C. R. Chaaban, N. T. Berry, C. Armitano-Lago, A. W. Kiefer, M. J. Mazzoleni, and D. A. Padua,

"Combining Inertial Sensors and Machine Learning to Predict vGRF and Knee Biomechanics during a Double Limb Jump Landing Task," *Sensors (Basel).*, vol. 21, no. 13, Jul. 2021, doi: 10.3390/S21134383.

[16] E. Dorschky, M. Nitschke, C. F. Martindale, A. J. van den Bogert, A. D. Koelewijn, and B. M. Eskofier, "CNN-Based Estimation of Sagittal Plane Walking and Running Biomechanics From Measured and Simulated Inertial Sensor Data," *Front. Bioeng. Biotechnol.*, vol. 8, p. 604, Jun. 2020, doi: 10.3389/FBIOE.2020.00604/BIBTEX.

[17] Y. Guo *et al.*, "A New Proxy Measurement Algorithm with Application to the Estimation of Vertical Ground Reaction Forces Using Wearable Sensors," *Sensors (Basel).*, vol. 17, no. 10, Oct. 2017, doi: 10.3390/S17102181.

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

[19] K. J. H. Ngoh, D. Gouwanda, A. A. Gopalai, and Y. Z. Chong, "Estimation of vertical ground reaction force during running using neural network model and uniaxial accelerometer," *J. Biomech.*, vol. 76, pp. 269–273, Jul. 2018, doi: 10.1016/J.JBIOMECH.2018.06.006.

[20] M. Mundt, W. R. Johnson, W. Potthast, B. Markert, A. Mian, and J. Alderson, "A Comparison of Three Neural Network Approaches for Estimating Joint Angles and Moments from Inertial Measurement Units," *Sensors 2021, Vol. 21, Page 4535*, vol. 21, no. 13, p. 4535, Jul. 2021, doi: 10.3390/S21134535.

[21] M. Mundt *et al.*, "Prediction of lower limb joint angles and moments during gait using artificial neural networks," *Med. Biol. Eng. Comput.*, vol. 58, no. 1, pp. 211–225, Jan. 2020, doi: 10.1007/s11517-019-02061-3.

[22] C. Wang *et al.*, "Real-Time Estimation of Knee Adduction Moment for Gait Retraining in Patients With Knee Osteoarthritis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 888–894, Apr. 2020, doi: 10.1109/TNSRE.2020.2978537.

[23] A. De Brabandere, J. Emmerzaal, A. Timmermans, I. Jonkers, B. Vanwanseele, and J. Davis, "A Machine Learning Approach to Estimate Hip and Knee Joint Loading Using a Mobile Phone-Embedded IMU," *Front. Bioeng. Biotechnol.*, vol. 8, Apr. 2020, doi: 10.3389/FBIOE.2020.00320.

[24] A. Ancillao, S. Tedesco, J. Barton, and B. O'flynn, "Indirect Measurement of Ground Reaction Forces and Moments by Means of Wearable Inertial Sensors: A Systematic Review," *Sensors (Basel).*, vol. 18, no. 8, Aug. 2018, doi: 10.3390/S18082564.

[25] "Accessories For Vicon Systems ." https://www.vicon.com/hardware/accessories/ (accessed Aug. 03, 2022).

[26] C. Richter Id, E. King, S. Strike, and A. Franklyn-Miller, "Objective classification and scoring of movement deficiencies in patients with anterior cruciate ligament reconstruction," 2019, doi: 10.1371/journal.pone.0206024.

[27] B. J. Stetter, S. Ringhof, F. C. Krafft, S. Sell, and T. Stein, "Estimation of Knee Joint Forces in Sport Movements Using Wearable Sensors and Machine Learning," *Sensors 2019, Vol. 19, Page 3690*, vol. 19, no. 17, p. 3690, Aug. 2019, doi: 10.3390/S19173690.

[28] L. Zhang, D. Soselia, R. Wang, and E. M. Gutierrez-Farewik, "Lower-Limb Joint Torque Prediction Using LSTM Neural Networks and Transfer Learning," *IEEE Trans. NEURAL Syst. Rehabil. Eng.*, vol. 30, p. 2022, doi: 10.1109/TNSRE.2022.3156786.

[29] S. Cerfoglio, M. Galli, M. Tarabini, F. Bertozzi, C. Sforza, and M. Zago, “Machine Learning-Based Estimation of Ground Reaction Forces and Knee Joint Kinetics from Inertial Sensors While Performing a Vertical Drop Jump,” *Sensors (Basel).*, vol. 21, no. 22, Nov. 2021, doi: 10.3390/S21227709.

[30] D. Hendry *et al.*, “An Exploration of Machine-Learning Estimation of Ground Reaction Force from Wearable Sensor Data,” *Sensors 2020, Vol. 20, Page 740*, vol. 20, no. 3, p. 740, Jan. 2020.

[31] Vicon Motion Systems Limited, “Plug-in Gait Reference Guide.”

[32] K. C. Moisio, D. R. Sumner, S. Shott, and D. E. Hurwitz, “Normalization of joint moments during gait: a comparison of two techniques,” *J. Biomech.*, vol. 36, no. 4, pp. 599–603, Apr. 2003, doi: 10.1016/S0021-9290(02)00433-5.

[33] D. S. Komaris *et al.*, “Predicting Three-Dimensional Ground Reaction Forces in Running by Using Artificial Neural Networks and Lower Body Kinematics,” *IEEE Access*, vol. 7, pp. 156779–156786, 2019, doi: 10.1109/ACCESS.2019.2949699.

[34] S. E. Oh, A. Choi, and J. H. Mun, “Prediction of ground reaction forces during gait based on kinematics and a neural network model,” *J. Biomech.*, vol. 46, no. 14, pp. 2372–2380, Sep. 2013, doi: 10.1016/J.JBIOMECH.2013.07.036.

[35] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” 2019.

[36] L. Ren, R. K. Jones, and D. Howard, “Whole body inverse dynamics over a complete gait cycle based only on measured kinematics,” *J. Biomech.*, vol. 41, no. 12, pp. 2750–2759, Aug. 2008, doi: 10.1016/J.JBIOMECH.2008.06.001.

[37] 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,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-November, pp. 2812–2815, Nov. 2015, doi: 10.1109/EMBC.2015.7318976.

[38] X. Jiang, C. Napier, B. Hannigan, J. J. Eng, and C. Menon, “Estimating Vertical Ground Reaction Force during Walking Using a Single Inertial Sensor,” *Sensors*, vol. 20, no. 15, 2020, doi: doi.org/10.3390/s20154345.

[39] F. Wouda, M. Giuberti, N. Rudigkeit, B.-J. van Beijnum, M. Poel, and P. Veltink, “Time Coherent Full-Body Poses Estimated Using Only Five Inertial Sensors: Deep versus Shallow Learning.,” *Sensors*, vol. 19, no. 17, p. 3716, 2019, doi: https://doi.org/10.3390/s19173716.

[40] E. Charry, W. Hu, M. Umer, A. Ronchi, and S. Taylor, “Study on Estimation of Peak Ground Reaction Forces using Tibial Accelerations in Running,” 2013. doi: 10.1109/ISSNIP.2013.6529804.

[41] B. Vanwanseele, “Mini symposium ‘Running Forwards’: wearable trunk accelerometry to monitor and modify running patterns,” 2023.

[42] J. R. Harry, J. Blinch, L. A. Barker, J. Krzyszkowski, and L. Chowning, “Low-Pass Filter Effects on Metrics of Countermovement Vertical Jump Performance,” *J. Strength Cond. Res.*, vol. 36, no. 5, pp. 1459–1467, May 2022, doi: 10.1519/JSC.0000000000003611.

[43] L. C. Benson, C. A. Clermont, R. Watari, T. Exley, and R. Ferber, “Automated Accelerometer-Based Gait Event Detection During Multiple Running Conditions,” *Sensors 2019, Vol. 19, Page 1483*, vol. 19, no. 7, p. 1483, Mar. 2019, doi: 10.3390/S19071483.

[44] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, “Normalization Techniques in Training

DNNs: Methodology, Analysis and Application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, pp. 10173–10196, 2023, doi: 10.1109/TPAMI.2023.3250241.

[45] B. L. Scheltinga, H. Usta, J. Reenalda, and J. H. Buurke, "Estimating Vertical Ground Reaction Force during Running with 3 Inertial Measurement Units," *Avestia Publ. J. Biomed. Eng. Biosci.*, vol. 9, pp. 2564–4998, 2022, doi: 10.11159/jbeb.2022.006.

[46] M. Zago *et al.*, "Estimation of Gait Mechanics Based on Simulated and Measured IMU Data Using an Artificial Neural Network," *Front. Bioeng. Biotechnol. | www.frontiersin.org*, vol. 8, p. 41, 2020, doi: 10.3389/fbioe.2020.00041.

[47] S. Gezici and H. V. Poor, "Position estimation via ultra-wide-band signals," *Proc. IEEE*, vol. 97, no. 2, pp. 386–403, 2009, doi: 10.1109/JPROC.2008.2008840.

[48] C. Medina, J. C. Segura, and Á. De La Torre, "Ultrasound Indoor Positioning System Based on a Low-Power Wireless Sensor Network Providing Sub-Centimeter Accuracy," *Sensors*, vol. 13, pp. 3501–3526, 2013, doi: 10.3390/s130303501.

[49] D. Atabay, "Pyrenn." 2016. doi: 10.5281/zenodo.45022.

[50] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," 2019.

Appendices

Participant preparation

Table 8 Overview of the locations of the reflective markers on the body.

Segment	**Marker name**	**Location** [31]
Upper body	STRN	Mid sternum
	CLAV	Clavicular notch
	C7	7th cervical vertebra
	T10	10th thoracic vertebra
	IST(1,2,3)	Rig markers
Pelvis	RASI/LASI	Anterior superior iliac spine
	RPSI/LPSI	Posterior superior iliac spine
	RPEL/LPEL	Lateral hip
	IPV(1,2,3)	Rig markers
Upper leg	RTHI/LTHI	Mid-lateral thigh
	RKNE/LKNE	Lateral epicondyle of the femur
	RMKN/LMKN	Medial epicondyle of the femur
	IRTH(1,2,3)/ILTH(1,2,3)	Rig markers
Lower leg	RTIB/LTIB	Mid-lateral shank
	RHEE/LHEE	Calcaneus
	RANK/LANK	Lateral malleolus
	RMAN/LMAN	Medial malleolus
	IRTB(1,2,3)/ILTB(1,2,3)	Rig markers
Feet	R1MT/L1MT	1st metatarsophalangeal joint
	R5MT/L5MT	5th metatarsophalangeal joint
	RTOE/LTOE	Over the second metatarsal head, on the mid-foot side of the equinus break between forefoot and mid-foot
	IRF(1,2,3)/ILF(1,2,3)	Rig markers

Performed movements

A visualization of the performed movements can be seen in Figure 12 and Figure 13.

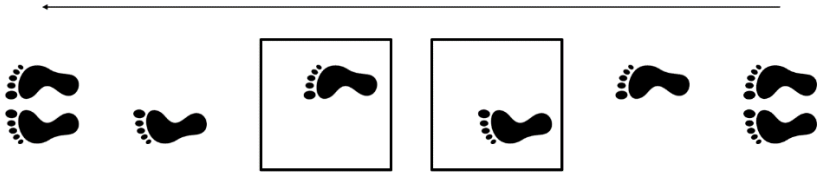

Figure 12 Schematic walking trial.

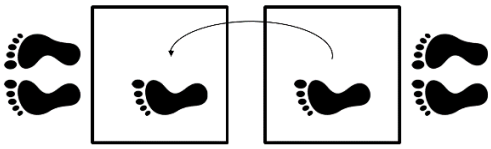

Figure 13 Schematic hop test trial of left leg.

Selection of literature models

To determine the most appropriate ML model for the gait and single leg hop test analysis, all read papers containing kinetic (gait) models were selected for extensive re-analysis. During this analysis, the algorithms were graded on what type(s) of kinetics their output was/were, the accuracy of the algorithm compared to other papers, and the reproducibility of the methods: was the description of the algorithm extensive enough and were no other inputs than IMU data needed for the algorithm to work. This process can be seen in Figure 14. For the hop test algorithm, the accuracy comparison was left out due to the little number of papers to compare with.

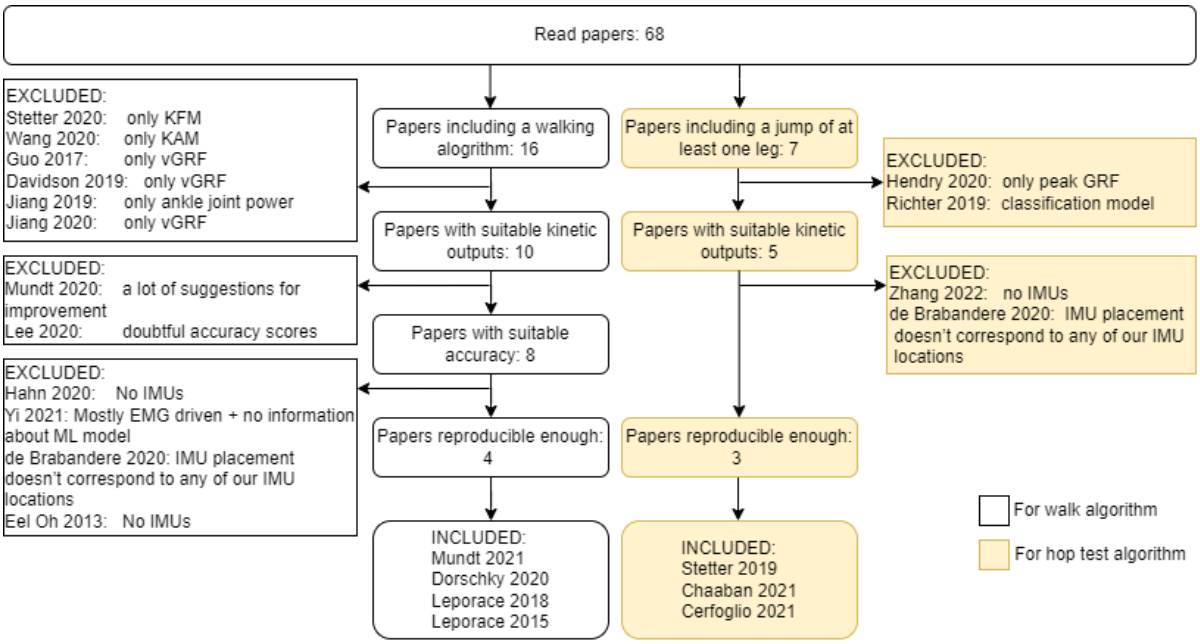

Figure 14 Overview of ML algorithm choices for validation.

Although four and three papers were selected sufficient, only one model of each movement was modeled based on existing literature. The choice of Leporace 2018 et al. was made because the amount of data was too small to model a CNN (as Dorschky et al. did), Mundt et al. had no concluding remarks on which of their model was best and Leporace 2015 was older than the one of 2018. The method of Stetter et al. was preferred above the other two. Chaaban et al. had used a more feature selection based model which was exactly what we wanted to avoid since it takes more preprocessing steps. Cerfoglio et al. did have an interesting approach, but it was not included based on limited time.

Overview of used trials per movement per subject group

Table 9 Overview of used trials per movement per subject group. This is divided into used, segmentation problems, synchronization problems, no ACL leg and saving problems.

	Healthy subjects	ACL patients
Walk (right)	10 used; 5 synchronization problems; 3 saving problems	10 used; 4 no ACL injury on right leg; 2 segmentation problems
Hop left	13 used; 3 synchronization problems; 2 saving problems	NA
Hop right	12 used; 2 segmentation problems; 2 synchronization problems; 2 saving problems	8 used; 4 no ACL injury on right leg; 2 segmentation problem; 2 saving problems

Methods for and results of obtaining the best validation models

For each model, the initial evaluation involved testing its performance on the same dataset on which it was trained. It was expected that the model would perform well in this case, since it was specifically trained to predict the same outcome relations. Secondly, a LOSO CV was conducted to assess the model's generalizability to unseen data. If this step gave promising results in terms of outcome metrics, an additional phase of hyperparameter optimization and PCA was carried out. This optimization and selection process was based on the outcome metrics of the different models. In Figure 15 this model evaluation process is shown.

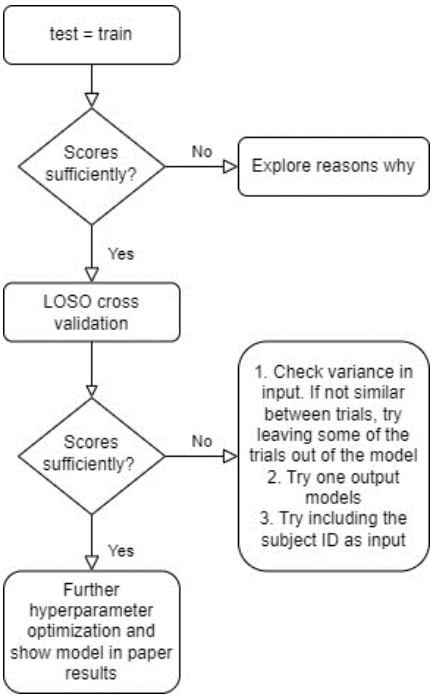

Figure 15 An illustration of the model evaluation process.

Before starting the analysis of the models, the range within the output variables was explored to look for possible differences between and within both subject groups. As can be seen in Figure 16, Figure 17, Figure 18, Figure 19, Figure 20 and Figure 21 for the hop push-off phase, there are a few outliers, but overall there is no significant difference between both groups.

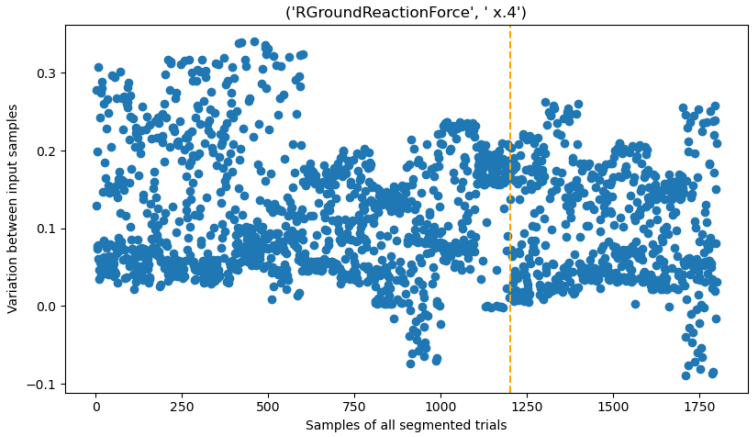

Figure 16 Variation within the A-P GRF during the push-off phase of the right leg hop jump for healthy subjects (<1200 samples) and for ACL patients (>1200 samples).

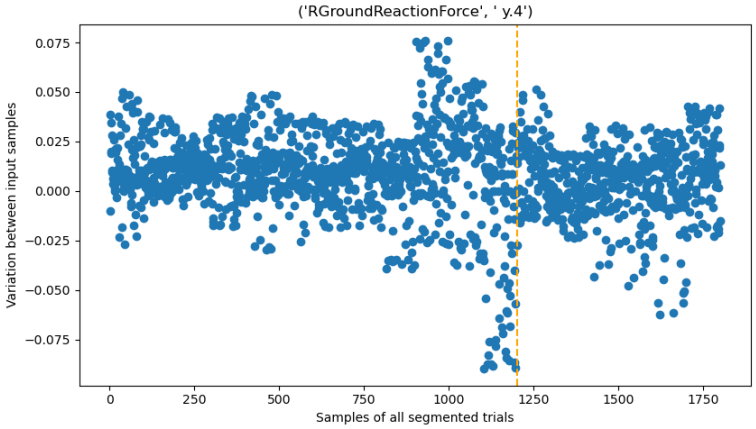

Figure 17 Variation within the M-L GRF during the push-off phase of the right leg hop jump for healthy subjects (<1200 samples) and for ACL patients (>1200 samples).

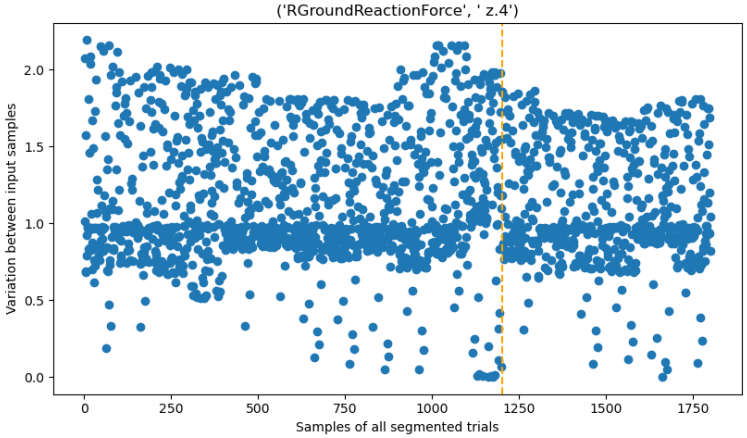

Figure 18 Variation within the vGRF during the push-off phase of the right leg hop jump for healthy subjects (<1200 samples) and for ACL patients (>1200 samples).

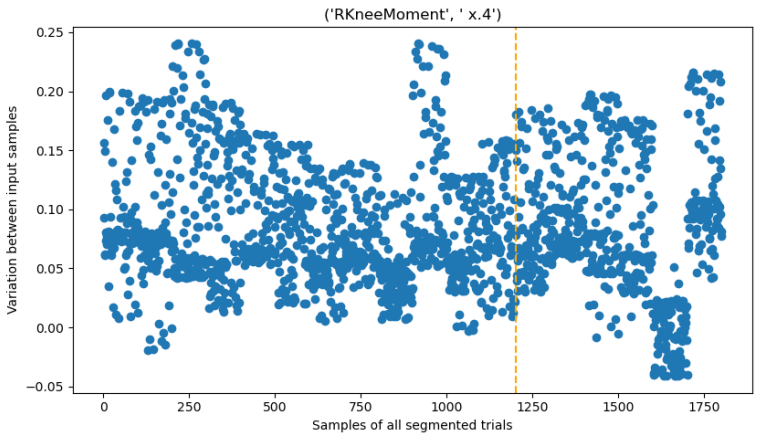

Figure 19 Variation within the flexion/extension net knee joint moment during the push-off phase of the right leg hop jump for healthy subjects (<1200 samples) and for ACL patients (>1200 samples).

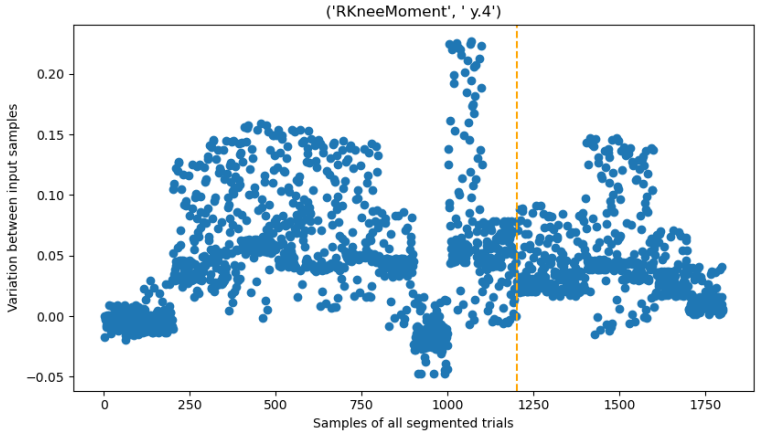

Figure 20 Variation within the varus/valgus net knee joint moment during the push-off phase of the right leg hop jump for healthy subjects (<1200 samples) and for ACL patients (>1200 samples).

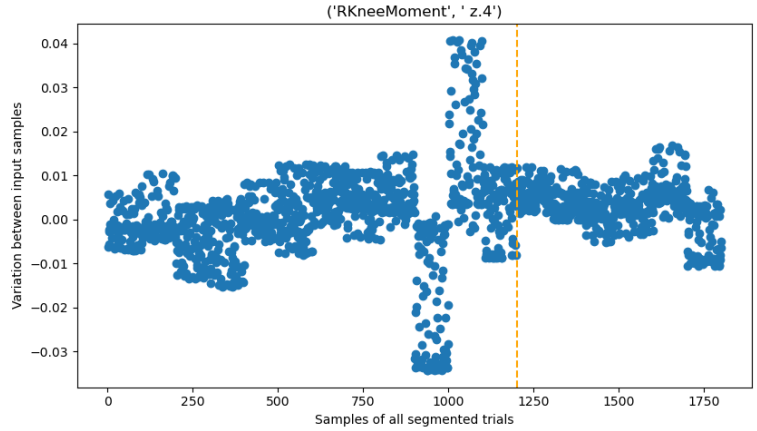

Figure 21 Variation within the internal/external net knee joint moment during the push-off phase of the right leg hop jump for healthy subjects (<1200 samples) and for ACL patients (>1200 samples).

Gait models

Leporace model: pyrenn-based

To mimic the methods of Leporace et al. [18] the Levenberg-Marquardt back-propagation algorithm was implemented using Python's pyrenn package [49]. All inputs, outputs, stance phase selection parameters and network architecture were implemented as in [18]. The maximum number of epochs was set to 1000 and default settings were used for the termination error, damping factor, damping constant and other pyrenn parameters. As depicted in Figure 22, the predicted value is not an accurate estimation of the target value, with all predicted values falling outside the range of target values. Furthermore, it is notable that retraining the pyrenn-based model yielded visibly different results each time, as evident from the comparison between Figure 23 and Figure 24. For these observations highlight the unreliability of the pyrenn-based model, it has been excluded from further evaluation.

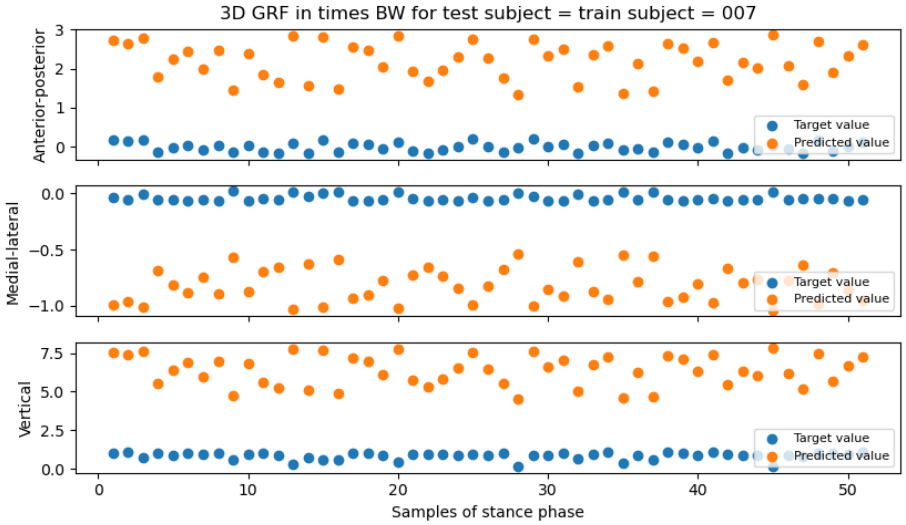

Figure 22 3D GRF predictions of the pyrenn-based model.

Figure 23 3D GRF predictions of the pyrenn-based model, after rescaling y-axes to the range of target values.

Figure 24 3D GRF predictions of the pyrenn-based model, after rescaling y-axes to the range of target values. This time retrained using the same in- and output values for training.

Leporace model: pyTorch-based

Unfortunately, no other easy to implement Levenberg-Marquardt algorithms could be identified. Consequently, a change in the type of optimizer was necessary compared to the methods employed by Leporace et al. Keeping all other parameters as much the same as possible, the Adam optimizer was then implemented, using Python's pyTorch package [50]. Examples of predicted values generated by this pyTorch-based model, after being trained and tested on the same dataset, are illustrated in Figure 25 and Figure 26. Although retraining the model (Figure 26) gave slightly varying results, the overall predicted values remained near the target values. The outcome metrics of this model are presented in Table 10.

Figure 25 3D GRF predictions of the pyTorch-based model.

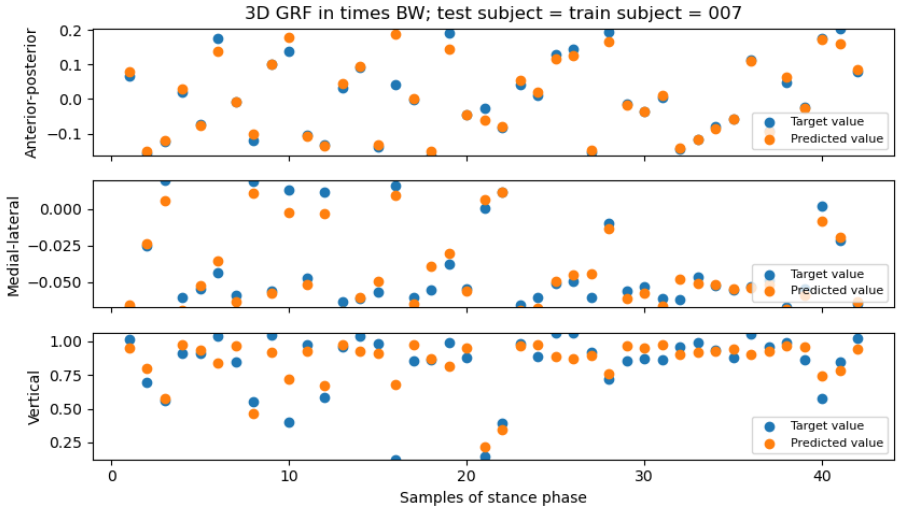

Figure 26 3D GRF predictions of the pyTorch-based model. This time retrained using the same in- and output values for training.

The LOSO CV was performed, where a different subject was used as the test subject each time. The outcome metrics of this model are presented in Table 10 (highlighted in green).

Additionally, the model was evaluated using a separate test set consisting of trials from different subjects. In this way, the model might already have been trained to learn some subject-specific information. As this is a manner of data leakage, this method was purely used to test whether the algorithm is in disadvantage when using it to predict for a new subject. However, the results obtained from this method (using 4 trials, 2 trials or even including the subject ID as input) did not significantly outperform the LOSO CV. Therefore, it can be assumed that the model does not show a disadvantage when predicting for unseen subjects.

Incorporating the subject ID as input data has also been trained and tested using the LOSO CV. The expectation was that this wouldn't show higher performance, as the input subject ID of the test person would be completely new to the model, making it challenging to use the subject ID effectively. The results of this model demonstrated similar outcomes as the normal LOSO CV.

Lastly, a separate model was trained for each output value, enabling the network to learn output-specific relationships rather than training relationships for all three outputs simultaneously. As depicted in Table 10, the A-P RMSE and rRMSE, as well as the A-P and M-L Pearson r, showed better performance than the normal LOSO CV. However, there was no significant difference observed between the scores of both models.

Leporace model: pyTorch-based and time coherent

Due to the limited amount of available data, the amount of information provided to the model was enhanced by including additional input features based on the time coherent nature of the data. This method, known as a stacked input neural network (SINN), was implemented following a similar approach as Wouda et al. had performed [39]. Specifically, the number of future and past data points considered was set to 2 and the number of samples that were shifted between the samples was also set as 2. For instance, the output of sample 50 was determined based on all previous inputs from sample 46, 48, 50, 52 and 54. It should be noted that the initial and final samples of the signal were excluded due to insufficient past and future samples, respectively. However, this did not pose an issue as the segmented stance phase for all subjects did not occur near the beginning or end of the signal. After the additional time instances were added as input features, all samples were shuffled. Outcome metrics of this approach, evaluated using both the test set that equaled the train set and the LOSO CV, are also reported in Table 10.

Table 10 Outcome metrics of a few different gait models, compared to Leporace's model. Standard deviation is shown between brackets. Green: selected Leporace model.

	RMSE			rRMSE			Pearson Correlation Coefficient (r)		
	A-P	M-L	Vertical	A-P	M-L	Vertical	A-P	M-L	Vertical
Leporace test = train	0.0274	0.0122	0.1366	5.16	8.74	10.38	0.9684	0.9165	0.8132
Leporace test = one trial of four subjects	0.0537	0.0183	0.2722	**8.38**	**12.98**	14.88	0.8794	0.6463	0.5450
Leporace test = one trial of two subjects	0.0591	**0.0167**	0.2990	10.42	13.34	16.74	0.8190	0.7242	0.4495
Leporace test = one trial of two subjects including subject ID	0.0545	0.0170	0.2864	9.74	13.39	16.38	0.8443	0.7386	0.4763
Leporace LOSO CV average	0.0537 (0.0110)	0.0184 (0.0021)	0.2359 (0.0356)	9.99 (1.93)	13.87 (1.98)	14.35 (1.57)	0.8891 (0.0465)	0.7283 (0.1149)	0.6641 (0.1129)
Leporace LOSO CV including subject ID	0.0540	0.0198	**0.2322**	10.06	14.85	**14.17**	0.9016	0.6748	0.6785
Leporace LOSO CV only one output per trained model	**0.0505** (0.0111)	0.0185 (0.0017)	0.2429 (0.0379)	9.31 (1.53)	14.04 (1.83)	14.71 (1.99)	**0.9042** (0.0325)	**0.7397** (0.1114)	0.6346 (0.1307)
Time coherent Leporace (SINN) test = train	0.0273	0.0155	0.0734	5.36	10.14	5.51	0.9694	0.8530	0.9564
Time coherent Leporace (SINN) LOSO CV average	0.05299 (0.0097)	0.0195 (0.0034)	0.2422 (0.0427)	9.76 (1.82)	14.92 (3.08)	14.63 (2.32)	0.8844 (0.0540)	0.6805 (0.1466)	0.6412 (0.1373)
Leporace [18]							0.97	0.80	0.98

Evaluating models using patient data

The first test was to test a model trained on healthy subject data, using patient data. The result of this is visible in Table 11.

Plots of the input variables showed that ACL patients tended to have a longer stance phase than the healthy subjects (see Figure 27).

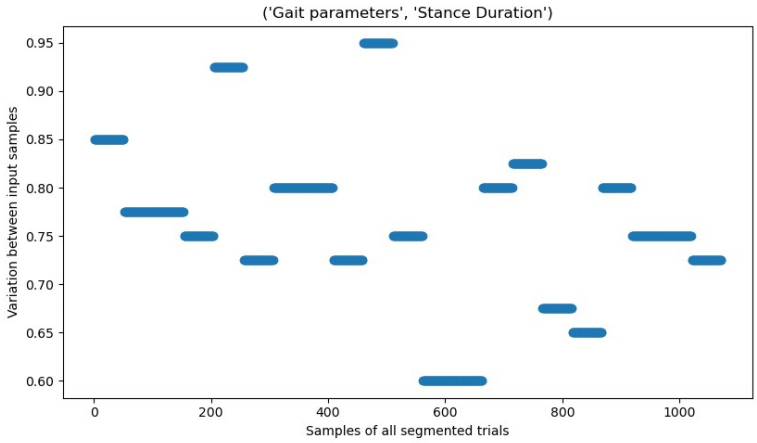

Figure 27 Variation within stance duration of patients (<= 561 samples) and healthy subjects (> 561 samples)

The patient models were run twice: once with the IMU data calibrated using the bony landmark calibration (BLC) as performed using in-house developed software (RRD, Enschede, The Netherlands) and once using stand calibration (the first few seconds of a trial when the subject is standing still). The healthy subject data was created using the latter of the two. When comparing the input of the patient data with the input of the subject data, at first it looked as if there were large differences between the groups (see Figure 28 and Figure 30). However, when testing whether this was a consequence of the different calibration techniques by calibrating the patient data using the stand calibration this time, it was visible that these large difference had become a lot smaller and the data was in a more similar range as the subject data (see Figure 29 and Figure 31). This was the reason why the patient models were also rerun using this new data using the less sophisticated stand calibration technique. The results of the evaluation of the patient models are shown in Table 11. The models of the combined healthy subjects and patient data scored best. This implies that the amount of data is more important than the possible difference between the two groups.

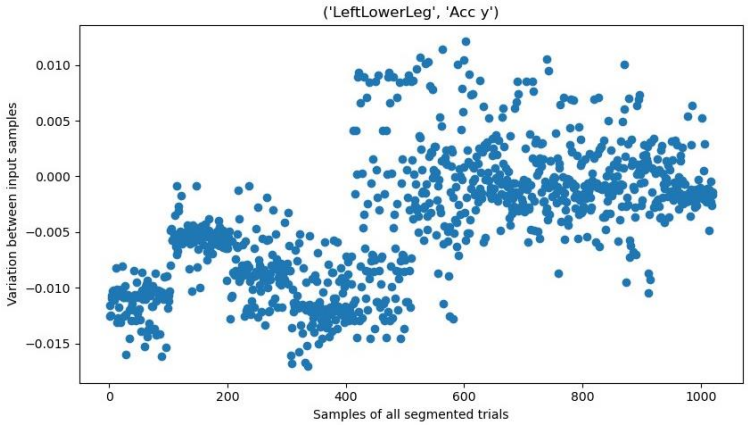

Figure 28 Variation within the acceleration of the LLL in the y direction of patients using the BLC (<= 510 samples) and healthy subjects (> 510 samples)

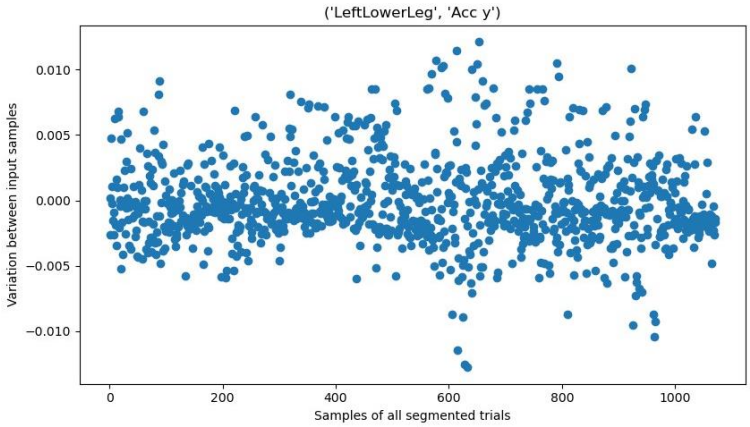

Figure 29 Variation within the acceleration of the LLL in the y direction of patients using the stand calibration (<= 561 samples) and healthy subjects (> 561 samples)

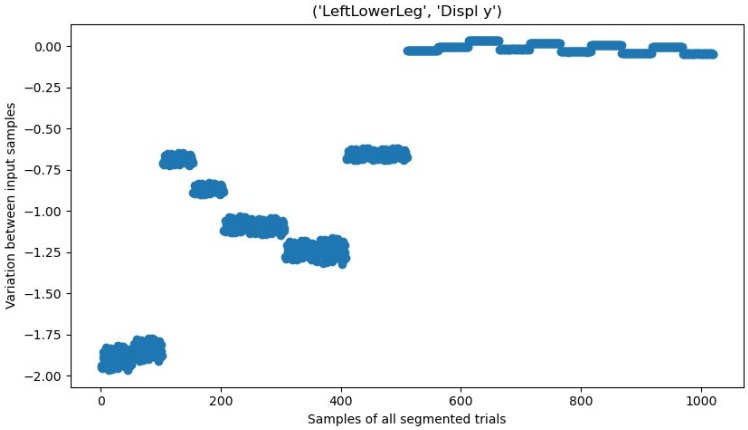

Figure 30 Variation within the displacement of the LLL in the y direction of patients using the BLC (<= 510 samples) and healthy subjects (> 510 samples)

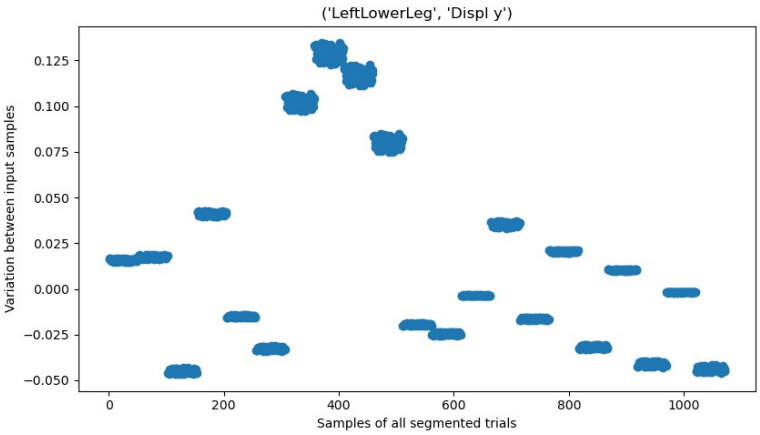

Figure 31 Variation within the displacement of the LLL in the y direction of patients using the stand calibration (<= 561 samples) and healthy subjects (> 561 samples)

Table 11 Outcome metrics of a few different patient gait models. Standard deviation is shown between brackets. In red: results using the BLC to Vicon comparison on ACL patients. In black: results using the stand calibration on ACL patients.

	RMSE			rRMSE			Pearson Correlation Coefficient (r)		
	A-P	M-L	Vertical	A-P	M-L	Vertical	A-P	M-L	Vertical
Pretrained (by healthy subjects) Leporace pyTorch	0.0656	0.0221	0.3006	10.01	12.42	16.95	0.7792	0.4999	0.5056
Patient Leporace LOSO CV	0.0701 (0.0367)	0.0195 (0.0063)	0.3042 (0.1068)	14.75 (6.73)	16.25 (4.64)	19.10 (6.03)	0.6771 (0.3632)	0.6875 (0.1529)	0.6482 (0.1377)
Patient Leporace LOSO CV without 8 and 61 and 12	**0.0525** (0.0152)	0.0217 (0.0051)	0.2509 (0.0597)	12.28 (3.40)	19.33 (5.32)	15.59 (3.32)	0.8473 (0.0727)	0.6867 (0.1803)	0.6989 (0.1597)
Combined patient and subject Leporace LOSO CV	0.0538 (0.0135)	0.0191 (0.0046)	**0.2329** (0.0617)	11.41 (2.43)	16.18 (3.64)	**14.50** (3.50)	0.8632 (0.0643)	0.7099 (0.1350)	**0.7025** (0.1555)
Pretrained (by healthy subjects) Leporace pyTorch	0.0605	0.0198	0.2765	**10.37**	12.39	16.36	0.8145	0.6285	0.5598
Patient Leporace LOSO CV without 8 and 61	0.0571 (0.0205)	0.0206 (0.0042)	0.2576 (0.0914)	12.22 (3.69)	17.31 (3.54)	16.16 (5.09)	0.8187 (0.0994)	0.6965 (0.1065)	0.6662 (0.1442)
Combined patient and subject Leporace LOSO CV	0.0536 (0.0128)	**0.0189** (0.0035)	0.2366 (0.0476)	10.78 (2.26)	**15.37 (3.48)**	14.77 (2.64)	**0.8652** (0.0666)	**0.7222** (0.0872)	0.6681 (0.0924)

Hop models

Stetter model take-off: pyTorch based

After the conclusion that pyrenn was not a suitable package to use, the hop analysis was started using a pyTorch-based model. All inputs, outputs and network architecture were implemented as Stetter et al. had performed [27].

The test=train method performed excellent (r >0.90, see Table 12). This can also be seen in the prediction plot in Figure 32. Therefore, it can be mentioned that this model can learn from training data.

The next step in the model evaluation process was to train and test using the LOSO CV. This gave poorer results, especially the standard deviation of the scores was really high. The difference between the performance of the different trained and tested subject models was high. This could be an indication of a few subjects being unsuitable for this model. This hypothesis was checked by inspecting the variances within the input features. In Figure 33 an example is illustrated where for one trial (samples between 1100 and 1200) a clear distinction from all other samples is visible. For some other input features, e.g., in Figure 34, a large dispersion of samples is seen for some of the trials (samples between 800 and 1000). Based on these observations, these trials (subject 7 and subject 8 hop left 2) were excluded from the dataset and the LOSO CV was run again. In Table 12 it can be seen that this approach only gave a better result for the A-P correlation coefficient. For all other metrics, the performance was worse. On top of that, the standard deviation was still very large and, in some cases, larger than the actual score. Afterwards, models with only one direction of the KJF as output were evaluated. They scored similar to the original LOSO CV model. The model where the subject ID was included as an input seemed to be better than the original LOSO CV model, but the standard deviations were still substantial large so no definite word could be spoken on it. As the performance of the models was still not as hoped or expected, one additional analysis was performed to test the hypothesis that the intersubject differences were too large to teach the model to generalize them by providing the train dataset. This was tested by creating a subject specific model. This was similar to the test=train situation but slightly deviated from this scenario because it was performed with both hop tests of a subject and the train/validation/test split was 70/15/15, indicating that the test set was unseen data. The results of this last evaluation imply that within different subject trials, the variation is not too large to train a model that is generalizable for that specific subject.

All in all, it can be said that the intersubject differences were too large to make a generalizable model for all subjects, but the intrasubject differences are small enough to generalize a subject-specific model.

Side note

When comparing the patient models to the healthy subject models, the decision was made to explore the right leg hop, since six out of the eight patients had their right leg injured. In making this comparison, it soon stood out that the right leg healthy subject models scored significantly better than the left leg models. Three possible reasons are thought of: 1. The mirrored ML forces are more easy to make relations to for the model. 2. For many people, the right leg is their dominant leg. Therefore, they might have better balance with this leg compared to the left leg, which makes the jump more predictable. 3. It is a coincidence that all left hops are performed less good/ there are more intrasubject differences.

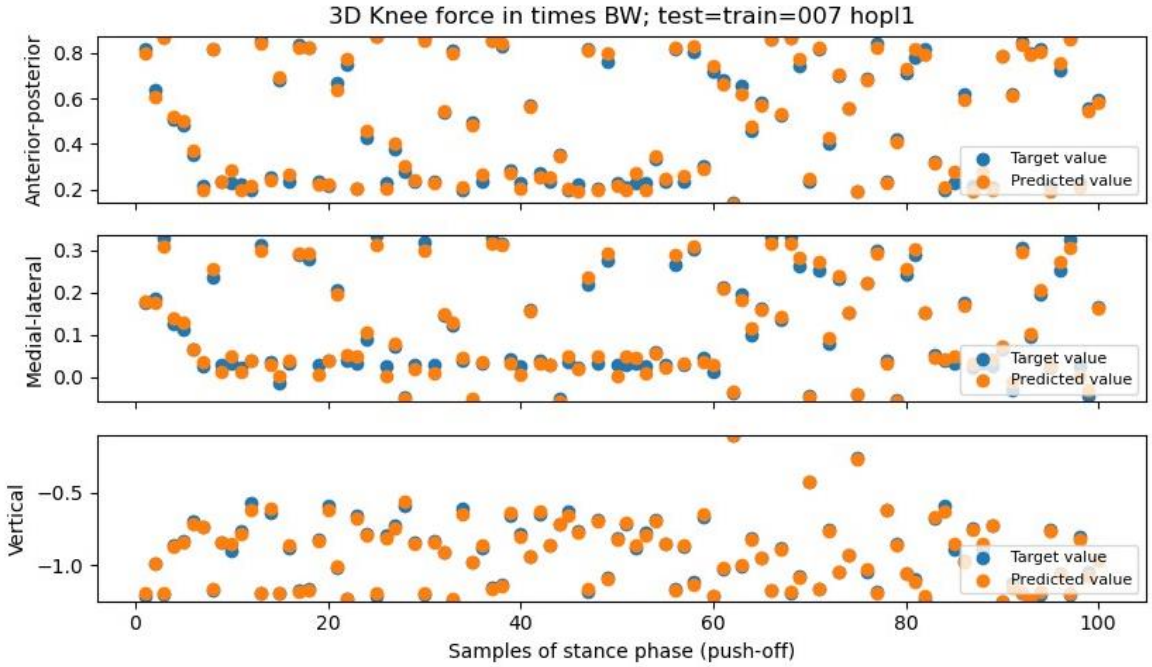

Figure 32 Prediction result of one test trial, when test = train.

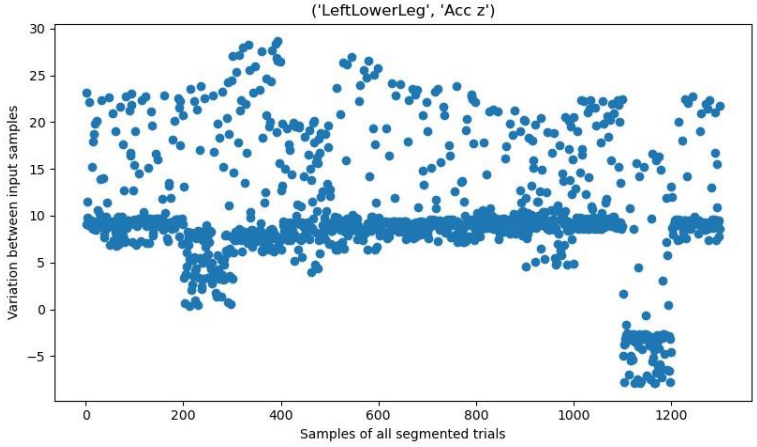

Figure 33 Variance within one of the model inputs: acceleration in the z direction of the LLL.

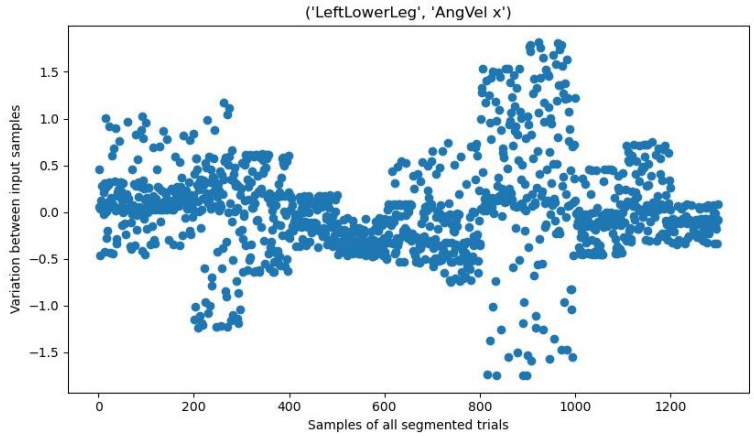

Figure 34 Variance within one of the model inputs: angular velocity in the x direction of the LLL.

Table 12 Outcome metrics of a few different hop push-off models, compared to literature models. Standard deviation is shown between brackets. Light green: very good learning ability. Green: selected OSHP model.

	RMSE			rRMSE			Pearson Correlation Coefficient (r)		
	A-P	M-L	T-C	A-P	M-L	T-C	A-P	M-L	T-C
Stetter test=train (right leg) all subjects	0.1516	0.1710	0.1819	10.22	14.85	6.60	0.7979	0.5316	0.8557
Stetter test = train (left leg)	0.0195	0.0131	0.0159	1.78	2.25	0.94	0.9971	0.9937	0.9977
Stetter test = one trial of five subjects	0.1899	0.2984	0.3013	12.66	20.42	10.94	0.7192	0.1199	0.5368
Stetter test = 15% of subject data	0.0243 (0.0079)	0.0158 (0.0051)	0.0313 (0.0086)	2.15 (0.34)	3.24 (1.60)	1.74 (0.52)	0.9972 (0.0004)	0.9856 (0.0158)	0.9956 (0.0030)
Stetter LOSO CV average (right leg)	0.1536 (0.0432)	0.0958 (0.0388)	0.1652 (0.0526)	14.28 (6.87)	15.82 (6.39)	7.66 (2.74)	0.8432 (0.0676)	0.8292 (0.0604)	0.9226 (0.041)
Stetter LOSO CV average (left leg)	0.2171 (0.0891)	0.3755 (0.1703)	**0.3416** (0.1252)	**18.71 (5.63)**	41.71 (19.36)	**16.58** (4.98)	0.6090 (0.2885)	0.2116 (0.5637)	0.4064 (0.4133)
Stetter LOSO CV without subject 7 and 8 hop left 2 (left leg)	**0.2168** (0.0872)	0.4097 (0.2624)	0.3761 (0.1803)	19.85 (6.29)	43.74 (20.84)	18.19 (7.89)	**0.6556** (0.1988)	0.1936 (0.6665)	0.3639 (0.4521)
Stetter LOSO CV only one output per trained model (left leg)	0.2206 (0.1029)	0.3758 (0.2083)	0.3901 (0.1555)	18.81 (6.44)	41.74 (20.67)	18.60 (6.48)	0.6555 (0.2884)	0.2891 (0.5168)	0.2615 (0.4969)
Stetter LOSO CV including subject ID (left leg)	0.2457 (0.1155)	**0.2779** (0.1360)	0.3879 (0.3088)	20.40 (7.58)	**34.39 (17.47)**	17.57 (10.95)	0.6059 (0.3142)	**0.3158** (0.5766)	0.4568 (0.4515)

	RMSE			rRMSE			Pearson Correlation Coefficient (r)		
	A-P	M-L	T-C	A-P	M-L	T-C	A-P	M-L	T-C
Time coherent Stetter (SINN) test = train	0.0083	0.0068	0.0119	0.65	1.16	0.59	0.9998	0.9990	0.9996
Time coherent Stetter (SINN) LOSO CV average	0.2787 (0.0907)	0.3794 (0.2117)	0.4927 (0.2293)	20.25 (5.41)	43.97 (16.39)	18.48 (7.58)	0.5762 (0.2334)	0.2705 (0.5359)	**0.5687** (0.1972)
Stetter [27]				17.4 (5.5)	45.9 (19.7)	15.4 (6.6)	0.89 (0.25)	0.31 (0.46)	0.92 (0.39)

Table 13 Outcome metrics of a few different hop landing models, compared to literature models. Standard deviation is shown between brackets Grey areas are unknown or not applicable. Green: selected OSHL model.

	RMSE			rRMSE			Pearson Correlation Coefficient (r)		
	A-P	M-L	T-C	A-P	M-L	T-C	A-P	M-L	T-C
Stetter test = train (right leg) all subjects	0.1395	0.1599	0.2104	9.86	15.13	7.31	0.7438	0.4159	0.7165
Stetter test = train	0.0066	0.0060	0.0111	0.51	1.02	0.56	0.9998	0.9991	0.9996
Stetter test=train one output per trained model	0.0077	0.0089	0.0171	0.60	1.52	0.86	0.9998	0.9995	0.9996
Stetter LOSO CV average (right leg)	0.1778 (0.0369)	0.1449 (0.0498)	0.2140 (0.1062)	13.65 (2.29)	21.91 (5.48)	8.47 (3.16)	0.7319 (0.2008)	0.5924 (0.2200)	0.7786 (0.1754)
Stetter LOSO CV average (left leg)	0.2660 (0.0986)	0.3038 (0.0732)	0.3502 (0.1012)	17.46 (5.27)	35.27 (13.82)	12.45 (3.17)	0.3331 (0.3230)	0.0231 (0.5485)	0.4495 (0.3292)
Stetter LOSO CV only one output per trained model	0.2161 (0.0396)	0.2246 (0.0638)		20.83 (4.61)	63.82 (58.38)		0.5946 (0.1948)	0.5323 (0.3938)	
Stetter [27]				25.1 (9.4)	38.9 (14.4)	16.7 (7.2)	0.77 (0.53)	0.42 (0.38)	0.84 (0.43)

Evaluating models using patient data

Table 14 Outcome metrics of a few different (right leg) hop push-off models when trained and tested on patient data. Standard deviation is shown between brackets. Grey areas are unknown.

	RMSE			rRMSE			Pearson Correlation Coefficient (r)		
	A-P	M-L	T-C	A-P	M-L	T-C	A-P	M-L	T-C
Stetter pretrained on healthy subjects	0.1583	0.1235	0.1859	11.81	16.58	8.14	0.7852	0.5179	0.8281
Stetter test = train Right ACL patients only	0.1555	0.0919	0.1606	11.27	13.33	6.95	0.7817	0.5936	0.8640
Stetter LOSO CV average	**0.1536** (0.0428)	**0.0961** (0.0388)	**0.1650** (0.0533)	**14.28** (6.84)	**15.87** (6.40)	**7.65** (2.77)	0.8432 (0.0676)	**0.8293** (0.0604)	**0.922** (0.0409)
Stetter LOSO CV average right ACL patients only	0.1779 (0.0573)	0.0983 (0.0543)	0.1953 (0.0436)	18.40 (7.81)	20.08 (7.38)	9.82 (2.03)	0.8127 (0.0288)	0.7894 (0.0738)	0.8819 (0.0419)
Stetter [27]				17.4 (5.5)	45.9 (19.7)	15.4 (6.6)	**0.89** (0.25)	0.31 (0.46)	0.92 (0.39)

Table 15 Outcome metrics of a few different (right leg) hop landing models when (trained and) tested on patient data. Standard deviation is shown between brackets. Grey areas are unknown.

	RMSE			rRMSE			Pearson Correlation Coefficient (r)		
	A-P	M-L	T-C	A-P	M-L	T-C	A-P	M-L	T-C
Stetter pretrained on healthy subjects	0.1851	0.1470	0.2406	10.52	15.41	7.92	0.4949	0.2753	0.5722
Stetter test = train right ACL patients only	0.1677	0.1073	0.1651	10.01	13.95	5.53	0.6218	0.3751	0.8450
Stetter LOSO CV average	**0.1780** (0.0368)	0.1447 (0.0500)	0.2152 (0.1067)	**13.68** (2.29)	**21.87** (5.37)	8.52 (3.17)	0.7317 (0.2008)	**0.5922** (0.2202)	0.7784 (0.1755)
Stetter LOSO CV average right ACL patients only	0.1918 (0.0322)	**0.1188** (0.04980)	**0.1839 (0.110)**	14.53 (4.00)	22.19 (7.08)	**7.58** (0.47)	0.7401 (0.1868)	0.5691 (0.2053)	0.8098 (0.1121)
Stetter [27]				25.1 (9.4)	38.9 (14.4)	16.7 (7.2)	**0.77** (0.53)	0.42 (0.38)	0.84 (0.43)

Model-based discussion

Pyrenn model

Where the pyrenn package was a feasible choice at first sight because of its implementation of the Levenberg-Marquardt backpropagation algorithm, it appeared to not give the expected results. A few hypotheses for these strange results are given. At first, the cause seemed to lie in a lack of data. However, when testing and training on the same trials, the predictions weren't within the range of the target output. Therefore, the amount of data is not the only problem. After a deep dive into the code of the package, it was discovered that there was no validation included in this algorithm. Consequently, no hyperparameter tuning was performed, which can contribute to this low performance. Still, a small amount of data and no hyperparameter tuning cannot be the only elements leading to this low-scoring algorithm. Probably there are more problems with the pyrenn package implementation, but it would have been very time-consuming to find these problems. The quick(er) solution was using other optimizers.

Model iterations of own developed models

Next to the OW, OHP and OHL models mentioned in this paper, also other slightly different models were tried to improve the outcome metrics. One model worked with normalized inputs (centered around zero per input) because this is an assumption that the PCA makes. For another model, the hyperparameters were optimized for a longer time. Instead of the regular 100 trials, now 400 trials with different settings were taken into account. The results of these deviating models are shown in Table 16, Table 17 and Table 18.

In the LOSO CV where the healthy subject group and the ACL patients are combined, we see that for OHL (see Table 18) the performance metrics were averaged compared to the outcome of both separate groups for most of the kinetics. This shows that the benefit of having additional data available to train the model, approximately outweighs the disadvantage of the differences between the two groups. The combined LOSO CV for the two other iterations of OHP model shows the same trend; most outcome metrics are between the outcomes of the two separate groups. The combined LOSO CV for the OHL model with inputs centered around zero shows that in some cases the combined models score better. However, these are not significant differences. The combined OHL presented in the head of the paper shows for many output variables a worse, but not significant, outcome than that of the separate groups. The combined OHP presented in the head of the paper shows for all output variables a worse, but not significant, correlation coefficient than both separate groups. This could indicate that the differences between the groups are too large to generalize for this model. Therefore, it is of additional value that the ACL patient group was analyzed in this study.

The results of the longer optimization model for OHL show that for ACL patients this actually is the best performing hop landing model. However, because this is not the case for the healthy participants, it was not chosen to present in the head of the paper.

Table 16 Additional outcome metrics for LOSO CV for OW model of right leg. Standard deviation is shown between brackets. h = trained and tested on healthy subjects (001-1; 002-1; 002-2; 004-1; 004-2; 006-1; 006-2; 007-1; 007-2; 008-2). Light orange: r <= 0.65. Darker orange: r<= 0.35.

		OW with inputs centered around zero			OW (settings optimized for 400 trials instead of 100)			OW (paper)		
		RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r
GRF	A-P h	0.0548 (0.0140)	10.23 (2.10)	0.8679 (0.0543)	0.5657 (0.0105)	10.90 (1.21)	0.8736 (0.0565)	0.0523 (0.0113)	9.81 (1.41)	0.8959 (0.0451)
	M-L h	0.0219 (0.0077)	14.03 (4.04)	0.7412 (0.1653)	0.0198 (0.0071)	13.04 (3.70)	0.7721 (0.1518)	0.0177 (0.0050)	12.94 (3.11)	0.6996 (0.2192)
	Vertical h	0.2325 (0.0535)	14.18 (2.79)	0.6346 (0.1614)	0.2165 (0.0473)	13.69 (2.46)	0.6856 (0.1502)	0.2030 (0.0718)	11.97 (3.68)	0.7323 (0.1811)
Net knee moment	F-E h	0.0236 (0.0073)	14.77 (3.77)	0.6362 (0.2023)	0.0230 (0.0073)	15.04 (3.18)	0.5882 (0.1825)	0.0235 (0.0085)	14.47 (3.97)	0.6598 (0.1691)
	V-V h	0.0286 (0.0101)	34.73 (15.27)	0.4462 (0.2787)	0.0274 (0.0113)	35.02 (16.72)	0.5517 (0.2856)	0.0171 (0.0107)	22.27 (7.64)	0.5896 (0.1502)
	I-E h	0.0068 (0.0013)	24.37 (5.46)	0.6829 (0.2086)	0.0068 (0.0020)	24.05 (6.86)	0.7063 (0.2062)	0.0047 (0.0009)	14.43 (1.91)	0.8687 (0.1041)

Table 17 Additional outcome metrics for LOSO CV for OHP model of right leg. Standard deviation is shown between brackets. h = trained and tested on healthy subjects (001-1; 002-1; 002-2; 003-1; 003-2; 004-1; 005-1; 005-2; 006-1; 006-2; 007-2; 008-2). p = trained and tested on right ACL leg of patients (p01-1; p01-2; p02-1; p02-2; p05-1; p05-2; p07-2; p08-1). hp = trained and tested on combination of healthy subjects and ACL patients. Light orange: r <= 0.65. Darker orange: r<= 0.35.

		OHP with inputs centered around zero			OHP (settings optimized for 400 trials instead of 100)			OHP (paper)		
		RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r
KJF	A-P h	0.1109 (0.0448)	9.99 (3.49)	0.9320 (0.0165)	0.1033 (0.0479)	9.28 (3.59)	0.9445 (0.0205)	0.1348 (0.0273)	12.49 (1.46)	0.9170 (0.0192)
	A-P p	0.1307 (0.0512)	13.58 (5.23)	0.9407 (0.0329)	0.1289 (0.0484)	13.20 (4.13)	0.9125 (0.0476)	0.1705 (0.0217)	18.79 (4.06)	0.9182 (0.0654)
	A-P hp	0.1079 (0.0495)	10.19 (4.05)	0.9455 (0.0152)	0.1087 (0.0460)	10.39 (4.20)	0.9383 (0.0279)	0.1273 (0.0439)	12.69 (3.07)	0.8972 (0.0463)
	M-L h	0.1755 (0.0812)	25.64 (18.90)	0.6934 (0.2372)	0.1698 (0.0792)	24.94 (17.48)	0.7399 (0.2196)	0.1663 (0.0860)	28.04 (27.90)	0.7933 (0.2777)
	M-L p	0.0750 (0.0388)	14.95 (6.82)	0.8924 (0.0717)	0.1169 (0.0404)	18.77 (5.51)	0.8572 (0.0415)	0.0702 (0.0393)	13.86 (7.68)	0.8937 (0.0098)
	M-L hp	0.1395 (0.0661)	23.70 (17.71)	0.7567 (0.2398)	0.1425 (0.0653)	22.45 (15.55)	0.7676 (0.2158)	0.1522 (0.0460)	28.44 (20.37)	0.7698 (0.2511)
	T-C h	0.1689 (0.0538)	7.70 (2.40)	0.9016 (0.0428)	0.1758 (0.0623)	8.18 (2.90)	0.8897 (0.0568)	0.1734 (0.0432)	8.40 (1.91)	0.8930 (0.0518)
	T-C p	0.1642 (0.0577)	7.58 (2.65)	0.8793 (0.0831)	0.1919 (0.0681)	8.78 (3.02)	0.8207 (0.1271)	0.1799 (0.0764)	9.13 (3.96)	0.8623 (0.1541
	T-C hp	0.1540 (0.0569)	7.08 (2.54)	0.9083 (0.0608)	0.1673 (0.0649)	7.66 (3.07)	0.8851 (0.0775)	0.2228 (0.0636)	12.14 (2.94)	0.7924 (0.0866)
Net knee moment	F-E h	0.0294 (0.0116)	12.71 (5.55)	0.8738 (0.0796)	0.0318 (0.0126)	13.92 (5.51)	0.8738 (0.0514)	0.0400 (0.0155)	17.36 (6.63)	0.8233 (0.0943)
	F-E p	0.0593 (0.0293)	29.42 (23.06)	0.6813 (0.2862)	0.0569 (0.0176)	30.35 (23.53)	0.7276 (0.2845)	0.0596 (0.0284)	36.77 (30.72)	0.5338 (0.4530)
	F-E hp	0.0365 (0.0216)	17.39 (14.17)	0.7925 (0.2237)	0.0378 (0.0199)	18.42 (15.44)	0.8136 (0.2342)	0.0394 (0.0165)	19.64 (12.77)	0.7598 (0.1528)
	V-V h	0.05326 (0.0245)	32.29 (16.94)	0.3898 (0.3213)	0.0490 (0.0252)	28.92 (17.74)	0.5030 (0.3052)	0.0498 (0.0188)	34.71 (19.31)	0.4654 (0.3074)
	V-V p	0.0318 (0.0131)	26.30 (12.39)	0.7173 (0.0732)	0.0313 (0.0094)	22.64 (14.24)	0.7660 (0.1668)	0.0293 (0.0103)	22.99 (10.59)	0.7919 (0.1285)
	V-V hp	0.0417 (0.0217)	29.59 (18.92)	0.5628 (0.2747)	0.0402 (0.0203)	26.22 (16.55)	0.5929 (0.2488)	0.0483 (0.0240)	39.49 (23.97)	0.1753 (0.6555)
	I-E h	0.0140 (0.0079)	36.99 (14.90)	-0.073 (0.4640)	0.0120 (0.0078)	35.24 (12.97)	-0.1678 (0.4392)	0.0198 (0.0071)	39.08 (13.69)	-0.3023 (0.4344)
	I-E p	0.0061 (0.0029)	29.88 (12.86)	-0.0491 (0.5883)	0.0085 (0.0051)	26.36 (14.87)	0.1671 (0.6874)	0.0114 (0.0074)	29.64 (8.57)	-0.1601 (0.6057)

		OHP with inputs centered around zero			OHP (settings optimized for 400 trials instead of 100)			OHP (paper)		
		RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r
	I-E hp	0.0094 (0.0076)	35.47 (14.79)	-0.1068 (0.5610)	0.0096 (0.0066)	30.45 (12.63)	0.2092 (0.4347)	0.0126 (0.0076)	37.25 (18.30)	0.1337 (0.6852)
GRF	A-P h	0.0569 (0.0219)	16.86 (6.97)	0.7092 (0.3304)	0.0524 (0.0188)	15.71 (6.58)	0.7538 (0.2627)	0.0615 (0.0163)	18.99 (4.64)	0.7274 (0.0946)
	A-P p	0.0431 (0.0157)	13.20 (4.03)	0.8321 (0.0936)	0.0454 (0.0078)	13.03 (3.33)	0.8235 (0.0947)	0.0542 (0.0175)	17.36 (5.11)	0.7450 (0.1780)
	A-P hp	0.0527 (0.0171)	16.10 (5.20)	0.7629 (0.1757)	0.0481 (0.0151)	14.51 (5.59)	0.7884 (0.2215)	0.0609 (0.0197)	18.82 (5.35)	0.6815 (0.1625)
	M-L h	0.0228 (0.0116)	20.01 (6.91)	0.3729 (0.3817)	0.0211 (0.0124)	19.25 (7.76)	0.4159 (0.3780)	0.0304 (0.0159)	25.57 (7.10)	0.1486 (0.2835)
	M-L p	0.0184 (0.0033)	20.59 (4.10)	0.1898 (0.2451)	0.0198 (0.0054)	18.05 (3.40)	0.4025 (0.2113)	0.0248 (0.0063)	22.19 (1.70)	0.0033 (0.3695)
	M-L hp	0.0213 (0.0102)	22.44 (5.89)	0.2958 (0.2916)	0.0195 (0.0089)	18.70 (6.02)	0.4460 (0.3176)	0.0234 (0.0111)	23.34 (6.70)	0.1628 (0.4316)
	Vertical h	0.1970 (0.0691)	7.64 (2.66)	0.9155 (0.0338)	0.2021 (0.0712)	7.99 (2.88)	0.9091 (0.0490)	0.2058 (0.0383)	8.36 (1.56)	0.9141 (0.0428)
	Vertical p	0.1676 (0.0473)	6.94 (2.16)	0.8939 (0.0691)	0.2128 (0.0507)	8.63 (2.30)	0.8486 (0.1076)	0.1977 (0.0607)	8.97 (3.14)	0.8714 (0.1354)
	Vertical hp	0.1771 (0.0685)	7.06 (2.64)	0.9207 (0.0455)	0.1936 (0.0758)	7.68 (3.11)	0.8990 (0.0655)	0.2431 (0.0663)	11.15 (2.61)	0.8405 (0.0695)

Table 18 Additional outcome metrics for LOSO CV for OHL model of right leg. Standard deviation is shown between brackets. h = trained and tested on healthy subjects (001-1; 002-1; 002-2; 003-1; 003-2; 004-1; 005-1; 005-2; 006-1; 006-2; 007-2; 008-2). p = trained and tested on right ACL leg of patients (p01-1; p01-2; p02-1; p02-2; p05-1; p05-2; p07-2; p08-1). hp = trained and tested on combination of healthy subjects and ACL patients. Light orange: r <= 0.65. Darker orange: r<= 0.35.

		OHL with inputs centered around zero			OHL (settings optimized for 400 trials instead of 100)			OHL (paper)		
		RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r
KJF	A-P h	0.1320 (0.0582)	11.09 (3.86)	0.8319 (0.1210)	0.1251 (0.0514)	9.66 (3.45)	0.8321 (0.1075)	0.1348 (0.0273)	12.49 (1.46)	**0.9170** (0.0192)
	A-P p	0.1598 (0.0169)	11.74 (2.05)	0.7776 (0.1623)	0.1390 (0.0169)	9.59 (2.52)	0.8826 (0.0624)	0.1534 (0.0121)	11.94 (2.15)	0.8561 (0.0371)
	A-P hp	0.1267 (0.0448)	9.57 (3.31)	0.8523 (0.0819)	0.1272 (0.0427)	8.99 (3.25)	0.8411 (0.1028)	0.1335 (0.0419)	10.76 (2.71)	0.8327 (0.1766)
	M-L h	0.1500 (0.0656)	25.35 (16.82)	0.6806 (0.3529)	0.1489 (0.0568)	21.43 (15.14)	0.6655 (0.3231)	0.1663 (0.0860)	28.04 (27.90)	0.7933 (0.2777)
	M-L p	0.1172 (0.0439)	19.07 (3.54)	0.6957 (0.1858)	0.1169 (0.0481)	18.72 (4.82)	0.7603 (0.0954)	0.1131 (0.0517)	21.27 (6.91)	0.7314 (0.0787)
	M-L hp	0.1314 (0.0569)	20.08 (11.47)	0.7056 (0.2945)	0.1432 (0.0583)	20.08 (12.50)	0.7080 (0.2719)	0.1387 (0.0637)	22.90 (14.04)	0.7283 (0.2739)
	T-C h	0.2033 (0.0700)	9.14 (2.78)	0.7657 (0.0863)	0.2059 (0.0473)	8.25 (2.18)	0.7762 (0.0867)	0.1734 (0.0432)	8.40 (1.91)	0.8930 (0.0518)
	T-C p	0.1369 (0.0369)	5.46 (1.15)	0.8832 (0.0623)	0.1178 (0.0204)	4.33 (0.49)	**0.9185** (0.0242)	0.1505 (0.0417)	6.55 (1.47)	0.8818 (0.0477)
	T-C hp	0.1771 (0.050)	7.32 (2.38)	0.8211 (0.1025	0.1737 (0.0620)	6.60 (2.68)	0.8487 (0.1023)	0.1855 (0.0551)	8.13 (2.05)	0.8131 (0.0850)
Net knee moment	F-E h	0.0502 (0.0178)	13.75 (3.73)	0.7986 (0.1104)	0.0501 (0.0159)	12.57 (3.96)	0.8133 (0.0975)	0.0400 (0.0155)	17.36 (6.63)	0.8233 (0.0943)
	F-E p	0.0714 (0.0199)	28.27 (27.71)	0.4931 (0.2407)	0.0646 (0.0232)	21.57 (19.93)	0.6364 (0.2188)	0.0657 (0.0230)	26.93 (26.08)	0.6690 (0.1893)
	F-E hp	0.0538 (0.0180)	15.81 (12.79)	0.7276 (0.2284)	0.0504 (0.0184)	13.75 (11.06)	0.7688 (0.1826)	0.0519 (0.0171)	16.10 (10.93)	0.7483 (0.1766)
	V-V h	0.0461 (0.0209)	25.54 (16.33)	0.5623 (0.4382)	0.0462 (0.0195)	21.75 (14.72)	0.5741 (0.3978)	0.0498 (0.0188)	34.71 (19.31)	0.4654 (0.3074)
	V-V p	0.0319 (0.0094)	20.33 (3.43)	0.5379 (0.1762)	0.0325 (0.0127)	19.99 (5.05)	0.6638 (0.1071)	0.0325 (0.0138)	22.88 (5.13)	0.5734 (0.1142)
	V-V hp	0.0409 (0.0169)	20.45 (10.90)	0.6049 (0.3545)	0.0436 (0.0171)	20.64 (11.63)	0.6214 (0.3313)	0.0418 (0.0191)	23.13 (12.93)	0.6249 (0.3468)
	I-E h	0.0069 (0.0041)	27.69 (13.12)	0.1842 (0.4628)	0.0072 (0.0040)	26.35 (12.91)	0.0830 (0.2670)	0.0198 (0.0071)	39.08 (13.69)	-0.3023 (0.4344)
	I-E p	0.0090 (0.0015)	25.90 (6.42)	0.1856 (0.2353)	0.0070 (0.0009)	30.42 (14.11)	0.2894 (0.3376)	0.0069 (0.0012)	32.58 (13.29)	0.1615 (0.4077)

		OHL with inputs centered around zero			OHL (settings optimized for 400 trials instead of 100)			OHL (paper)		
		RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r	RMSE (BW)	rRMSE (% BW)	r
	l-E hp	0.0071 (0.0036)	24.27 (11.81)	0.1540 (0.2128)	0.0071 (0.0041)	24.11 (12.92)	0.2781 (0.2738)	0.0072 (0.0043)	28.06 (15.08)	0.3190 (0.3648)
GRF	A-P h	0.0761 (0.0198)	13.03 (4.03)	0.8757 (0.0402)	0.0713 (0.0161)	10.90 (2.47)	0.8600 (0.0609)	0.0615 (0.0163)	18.99 (4.64)	0.7274 (0.0946)
	A-P p	0.0631 (0.0181)	10.06 (3.35)	0.8574 (0.0979)	0.0483 (0.0106)	7.43 (1.81)	0.9220 (0.0477)	0.0659 (0.0128)	11.19 (2.30)	**0.8750** (0.0315)
	A-P hp	0.0605 (0.0174)	9.56 (3.35)	0.9050 (0.0460)	0.0646 (0.0204)	9.43 (2.79)	0.8925 (0.0470)	0.0712 (0.0213)	11.74 (3.66)	0.8788 (0.0466)
	M-L h	0.0273 (0.0133)	16.82 (4.03)	0.1849 (0.2462)	0.0272 (0.0133)	15.71 (4.61)	0.2160 (0.1384)	0.0304 (0.0159)	25.57 (7.10)	0.1486 (0.2835)
	M-L p	0.0269 (0.0071)	15.58 (2.02)	0.3967 (0.1300)	0.0263 (0.0096)	15.48 (3.26)	0.4233 (0.1634)	0.0272 (0.0091)	17.79 (3.29)	0.3553 (0.1565)
	M-L hp	0.0265 (0.0111)	15.19 (3.66)	0.3800 (0.1893)	0.0262 (0.0105)	14.07 (3.84)	0.4182 (0.1415)	0.0277 (0.0111)	16.54 (4.41)	0.2233 (0.1956)
	Vertical h	0.2245 (0.0877)	9.00 (3.20)	0.8160 (0.0886)	0.2267 (0.0597)	8.11 (2.36)	0.8157 (0.0857)	0.2058 (0.0383)	8.36 (1.56)	0.9141 (0.0428)
	Vertical p	0.1620 (0.0316)	5.74 (0.91)	0.8823 (0.0752)	0.1243 (0.0212)	4.11 (0.55)	**0.9274** (0.0220)	0.1650 (0.0334)	6.38 (1.07)	0.8960 (0.0299)
	Vertical hp	0.1914 (0.0636)	7.05 (2.69)	0.8539 (0.0923)	0.1930 (0.0723)	6.70 (2.92)	0.8653 (0.0959)	0.2016 (0.0691)	8.00 (2.62)	0.8510 (0.0771)

Literature research 'dynamic stability'

As this was originally part of my project, I wanted to insert the notes I made along the way. This can be seen as a quick start of diving further into a measure to determine dynamic stability in ACL patients.

https://www.sciencedirect.com/topics/medicine-and-dentistry/joint-stability

The functional joint stability accomplished through the integration and complementary relationship between the static and the dynamic components of joint stability is referred to as dynamic joint stability.

(1) Effects of anterior load carriage on dynamic gait stability during level overground walking

Dynamic gait stability quantifies the kinematic relationship between the human body's center of mass and base of support and has been widely used to assess fall risk.
Dynamic gait stability was determined based on the kinematics.
It was developed based on the Feasible Stability Region (FSR) theoretical framework (Pai & Patton, 1997). The FSR theory defines dynamic gait stability in terms of the COM's relative motion state (i.e., the combination of position and velocity) to its BOS (=Base of Support). The FSR is enclosed by two limits: the boundaries against backward and forward balance loss. These two boundaries were determined analytically using computer simulation assisted by dynamic optimization techniques (Pai Iqbal, 1999; Pai Patton, 1997; Yang, Anderson, Pai, 2007; Yang, Passariello, Pai, 2008). The simulation process sought the minimum (or maximum) COM velocity relative to the BOS, for a series of discrete initial COM positions, which would bring the COM over the heel to avoid a backward balance loss (or to the toe to prevent a forward balance loss) when the body's COM stops moving. Then, each boundary was obtained by spline interpolations over the derived COM velocity-position pairs in the state space. The boundaries have also been verified experimentally by massive data collected during various motor tasks (Wang et al., 2011; Wang et al., 2012; Yang et al., 2018).

Two reasons could account for the discrepancy in the findings between the studies. First, stability in the previous study was quantified by the margin of stability. Although the concept of margin of stability was also established from the kinematic relationship between the COM and BOS, it was proposed based on a simplified inverted pendulum model that used a linear approximation to solve its equation of motion (Hof, Gazendam, & Sinke, 2005). While the margin of stability is highly attractive due to its simplicity, the linearity may not quite accurately characterize the limits of stability at a gait speed range like that during walking (Hof et al., 2005). The FSR-based dynamic gait stability was analytically derived based on a 7-link human gait model and could more accurately quantify the COM-BOS kinematic relationship during gait than the margin of stability (Yang et al., 2007). Second, like the concept of dynamic gait stability, the margin of stability originally used the edge of the BOS as the reference point to determine the stability value. However, the reference point was the geometric center of the BOS (Alamoudi et al., 2018). As the centroid of the BOS is always different from its edge, this variation could also contribute to the difference in the findings regarding the front load carriage's effect on gait stability. The results in the current study would provide a precise picture of gait stability during anteriorly-loaded walking.

(2) Quantifying the Dynamic Stability of Gait Patterns in People with Hallux Valgus

To assess complex gait dynamics, we quantified the potential changes in gait stability by using the maximum Lyapunov exponent (MLE). Angular displacements of the ankle, knee, and hip in the sagittal plane during walking were investigated to calculate the MLE for gait stability based on foot conditions (i.e., barefoot, flat shoes, and high heels).

The maximum Lyapunov exponent (MLE), one such non linear dynamics analysis method, estimates the local stability of a system [21, 23, 24]. The MLE can be used to represent gait stability by

quantifying the ability of the human dynamic system to attenuate small perturbations in the gait trajectory that occur over time [25–29]. For example, people without neurological diseases or musculoskeletal disorders use their body to dampen variability and improve gait stability, in order to maintain a stable gait pattern even in situations involving disturbances. A larger MLE results in faster divergence and indicates lower dynamic stability [21]. The MLE has also been used to investigate gait stability of specific groups, such as the elderly and patients with knee arthritis [1, 25]

The Lyapunov exponent is a variable that quantifies the distance between two points in proximity as they move away over time [24]. For the Lyapunov exponent, the small perturbations arising from the difference between stride lengths are traced [22]. A Lyapunov exponent exists for each moving dimension in the analyzed gait trajectory. When the divergence rate is high, the Lyapunov exponent has the highest value; this value is called the Maximum Lyapunov Exponent, i.e., .max [20, 24]. Data were analyzed without filtering and resampling to accurately represent the variability within the system. Kinematic data pertaining to the lower extremity of each subject were analyzed to obtain the time series data over 30 gait cycles [22].

The higher MLE values at the knee joint showed that, for people with HV, the lowest local dynamic stability is observed at the knee joint, as compared to that at the lower limb joints.

(3) Effects of anteriorly-loaded treadmill walking on dynamic gait stability in young adults

Dynamic gait stability, defined by the Feasible Stability Region (FSR) theory, quantifies the kinematic relationship between the body's center of mass (COM) and base of support (BOS). FSR-based dynamic gait stability has been used to evaluate the fall risk.

The FSR characterizes the position-velocity relationship between the body's COM and its BOS to sustain balance [9] (Appendix).

Dynamic gait stability is defined as the shortest distance from the COM's motion state to the FSR's lower boundary [9].

Compared with other metrics quantifying body stability (e.g., maximum Floquet multipliers [10–12], Lyapunov exponents [11–13], the margin of stability [14]), FSR-based dynamic gait stability is more closely related to fall risk given that balance control is a process to preserve an appropriate COM-BOS relationship [15,16].

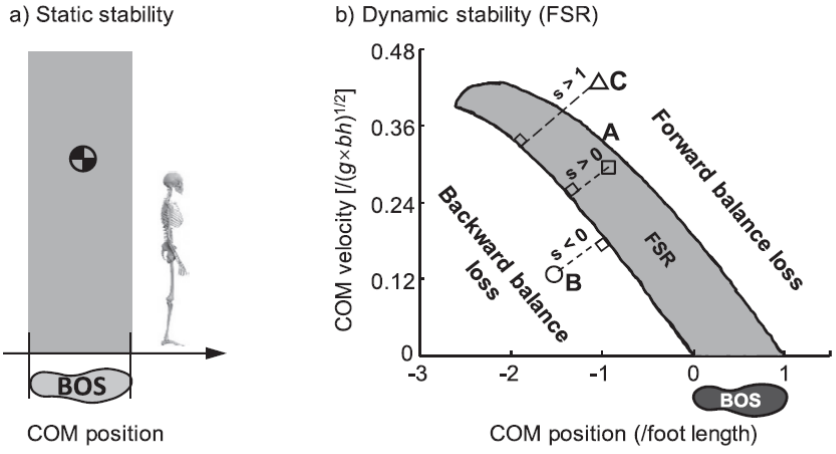

Fig. A1. The demonstration of a) the static stability limits and b) dynamic stability based on the Feasible Stability Region (FSR) theory. The static stability only considers the relative position of the COM to the BOS. Whenever the COM projection is inside the BOS, one is stable. The FSR takes both the position and velocity of the COM relative to BOS into consideration. The two components of the system's COM motion state (i.e., its anteroposterior position and forward velocity) are calculated relative to the rear of the BOS. When the COM's motion state is above (point A)/below (point B) the lower limit of FSR, the dynamic stability is positive/negative, indicating a stable/unstable state against backward falling. When the COM's motion state is above the upper limit of the FSR (point C), the dynamic stability is greater than one indicating an unstable state against forward balance loss. Its value (s) is quantified as the shortest distance (the dashed lines) from the COM motion state to FSR's lower limit.

(4) Seven-Weeks Gait-Retraining in Minimalist

Dynamic Postural Stability Index (DPSI),

we analyzed the Dynamic Postural Stability Index and its directional components: the medial-lateral, anterior posterior, and vertical stability indices (Wikstrom et al., 2005).

For both assessments, participants completed a double leg jump, from 70 cm away from the center of the platform, landed on their dominant foot, and stabilized as quickly as possible (Wikstrom et al., 2005). Upon landing, participants remained in a single limb stance for 20 seconds but only the first 3 seconds after landing were used for further analysis (Wikstrom et al., 2005).

The Dynamic Postural Stability Index (DPSI) and its directional components: the medial-lateral (MLSI), anterior-posterior (APSI), and vertical (VSI) stability indices, were calculated from the GRF signals as the dispersion of forces from the center of pressure in each of the axes (Wikstrom et al., 2005).

It is important to note that these dynamic postural stability variables do not have specific units because they are dimensionless. Thus, higher values indicate worse stability and lower values indicate better stability.

Dynamic postural stability was measured by asking participants to jump over a known distance and land on the ground with one leg stabilizing as quickly as possible. This task quantifies the ability to move from a highly unstable situation to a controlled situation. Dynamic postural stability variables were calculated from the GRF signals as the dispersion of forces in each of the axes, where higher values of each variable were associated with poor stability and lower values with high stability.

(5) Dynamic gait stability in patients with idiopathic normal pressure hydrocephalus with high and low fall-risk

Dynamic stability was defined as the ability to maintain an extrapolated center of mass within the base of support at heel contact, with the distance between the two defined as the margin of stability. Conscious motor control was assessed by the Movement-Specific Reinvestment Scale.

Thus, it can be surmised that dynamic stability during gait in iNPH differs from that in healthy individuals. Based on the simple inverted pendulum model, dynamic stability during gait can be represented by the ability to maintain an extrapolated center of mass (COM) within the base of support(BOS), with the distance between the two defined as the margin of stability (MOS) (Hof et al., 2005). The MOS can reflect the degree of dynamic stability during gait in various gait impairments, including neurological disorders such as Parkinson's disease, multiple sclerosis, and post-stroke patients (Hak et al., 2013; Peebles et al., 2016, 2017; Stegemller et al., 2012; Urakami et al., 2021). However, no previous studies have experimentally evaluated dynamic gait stability parameters, such as the MOS, in patients with iNPH. Moreover, it is unclear whether there is difference in dynamic stability during gait between patients with iNPH with high and low fall-risk.

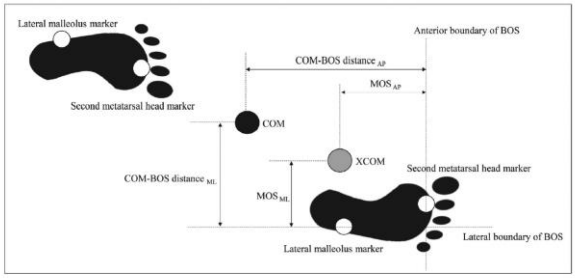

Fig. 1. COM-BOS distance AP/ML was calculated as the distance from the COM (filled circle) to the anterior/lateral boundary of the BOS, defined by the leading second metatarsal head/lateral malleolus marker (open circle). MOS AP/ML was calculated as the distance from the XCOM (gray circle), which indicates the COM position reflecting the COM velocity, to the anterior/lateral boundary of the BOS (Urakami et al., 2021).

(6) Effect of dynamic stability on a step task in ACL deficient individuals
Not useful

(7) A diagonal landing task to assess dynamic postural stability in ACL reconstructed females
Previous research has used time to stabilization (TTS) from forward landing tasks to assess dynamic postural stability in ACL reconstructed (ACLR) athletes in order to identify impaired sensorimotor control and mechanical stability. The purpose of the present study was to compare TTS values from a forward land and a diagonal land to determine if diagonal landing TTS values are more sensitive to dynamic postural stability deficits in female ACLR athletes.

Time to stabilization (TTS) is a functional measurement of neuromuscular control and dynamic postural stability. TTS scores assess an athlete's ability to transfer from a dynamic to a static situation on one leg. Longer TTS values have been reported in athletes with chronic ankle instability as well as athletes who have undergone ACL reconstruction surgery when compared to healthy controls [20–22]. These studies considered TTS scores from a forward landing task.
Recently it has been suggested that dynamic postural stability tests should include landings that challenge stability in the medial–lateral (M–L) direction, thus emphasizing an increased requirement for frontal and transverse plane neuromuscular control, as well as the anterior–posterior (A–P) direction [23]. Injury mechanisms are more likely to include a M–L component to the landing, which the forward landing protocol does not challenge. This is in agreement with recent research which suggests that ACL injury is most likely the result of multiplanar neuromuscular control deficits [24,25].

These findings suggest that TTS values from a diagonal landing are more sensitive at detecting dynamic postural stability deficits in an ACLR population compared to TTS values from a forward landing.

(8) The Effects of Balance Training on Static and Dynamic Postural Stability Indices After Acute ACL Reconstruction
Before and after the interventions, overall, anteroposterior, and mediolateral stability indices were measured with a Biodex Balance System in bilateral and unilateral stance positions with the eyes open and closed.

The overall stability index (OSI), anteroposterior stability index (APSI), and mediolateral stability index (MLSI) were recorded with a Biodex Balance System (SD 950-340, Biodex Medical Systems, Inc., Shirley, NY, USA).
The Biodex system has a circular deck with a 55 cm diameter located 20 cm above the floor inside its body, which is able to tilt 20 degrees from the horizontal position to all sides. The overall stability index shows the variance in plate deviation from the horizontal plane. The anteroposterior and mediolateral indices show the deviation of the plate from the horizontal position in the sagittal and frontal planes, respectively. The scores for the indices show the level of deviation from the horizontal position, so the lower scores indicate better balance (Akbari et al., 2014a, 2014b, Manual of Biodex System).

(9) Dynamic stability in the anterior cruciate ligament deficient knee
however, research has shown that passive anterior knee laxity is unrelated to knee stability during dynamic activities [5, 11,18, 29].

During low level activities, such as walking, ACL deficient individuals who are highly successful at stabilizing their knees dynamically (copers) display no alteration in their gait pattern [22]. If copers are included in the mix of all subjects, then genuine differences in movement patterns will be obscured.

(10) Dynamic Knee stability Current theory and implications for clinicians and scientists
The ability of the knee joint to remain stable when subjected to the rapidly changing loads it withstands during activity is referred to as dynamic knee stability.
Dynamic knee stability is the result of the integration of articular geometry, soft tissue restraints, and the loads applied to the joint from weight-bearing and muscle action.

Dynamic knee stability is the result of several factors, including articular geometry, soft tissue restraints, and the loads applied to the joint from weight-bearing and muscle action.

(12) The effect of the stability threshold on time to stabilization and its reliability following a single leg drop jump landing.
The TTS (time to stabilization) was defined as time between impact and the intersection of the processed signal with the threshold, after which it remained below the threshold for the subsequent 0.5 s.

Finally, in order to define the phase that primarily concerns the impact of the landing, we established the mean 'time to bodyweight' (TTBW) (based on 492 trials). This is the intersection of the RAW V signal with the bodyweight (100%) directly following the impact peak(Fig. 1). In further data processing, we ignored thresholds yielding TTS values shorter than mean TTBW.

Based on minisymposium on 3 February 2023

Adding an arch support to a heel lift improves stability and comfort during gait
The displacement and velocity in the medial-lateral COP, especially during the forefoot contact phase and the foot flat phase, are associated with foot stability [6].
The COP trajectory during gait reflects the dynamic action of the foot [16]. The medial shift of the COP during forefoot contact phase in the arch support condition suggests a transfer of loading. There fore, the medial shift of the COP trajectory caused by the arch support might indicate better stability.

Energy cost of running instability evaluated with wearable trunk accelerometry
Wearable trunk accelerometers provide a new level of analysis for dynamic stability of human locomotion. Accelerometers have improved from an accuracy, sensitivity, and computing power standpoint and have enabled more sophisticated analyses of motion. When mounted to the lower trunk, accelerometry unobtrusively estimates CoM motion and thus allows for several aspects of dynamic stability to be captured. These stability aspects, whether it function vertically, i.e., body weight support; mediolaterally (ML), i.e., side-to-side balance control; or anteroposteriorly (AP), i.e., braking and propulsion could more directly test various biomechanical hypotheses underpinning running economy.
Several linear and nonlinear stability aspects are worthy of investigation.

1. higher amplitudes or variations of trunk accelerations expressed as the acceleration root mean square (RMS) could reflect excessive changes in momentum that are energetically wasteful (14).

2. the dominant autocorrelations of acceleration waveforms could empirically test whether the ability to maintain a global consistency either between steps or strides are influential on economy.
3. sample entropy of trunk accelerations accounts for the complexity of the trunk acceleration signal waveform and could assess whether overall fluidity of a runner's gait pattern is related to economy (1).

We experimentally evaluate these hypotheses using simple and nonlinear measures including 1) the RMS, 2) interstep, 3) interstride regularity, and 4) the sample entropy of waveforms of each acceleration axis (vertical, M-L, and A-P), each of which express unique aspects of dynamic stability during running.

Individual selection of gait retraining strategies is essential to optimally reduce medial knee load during gait

Influence of outdoor running fatigue and medial tibial stress syndrome on accelerometer-based loading and stability

Various types of dynamic loading and dynamic stability measures have since been reported in the literature, and thus these terms need to be more specifically defined. **Dynamic stability** is defined operationally as the ability to maintain optimal variability, symmetry, regularity, or complexity of tri-axial trunk acceleration patterns while running.

These measures were quantified firstly by the ratio of each linear acceleration axis root mean square (RMS) relative to the resultant vector RMS to capture variability in accelerations [22]; secondly by the regularity of steps and strides using the primary and secondary dominant unbiased autocorrelation coefficients to indicate consistency between steps and strides, with perfect regularities equivalent to one [20] (of note, this autocorrelation procedure is also used to acquire step frequency using the time lag of the primary dominant autocorrelation coefficient [13,20]); and thirdly using sample entropy as a non-linear measure to capture complexity of unfiltered acceleration waveforms, with values typically in range of 0–2 for physiological systems, and higher values indicating less periodicity or more unpredictability [24]. In contrast to the aforementioned measures, sample entropy was analysed from unfiltered accelerations so as not to mask or remove any dynamical properties or variability present within the system that could be physiological meaningful [24,25]. Detailed equations and algorithm inputs for the computation and extraction of these dynamic stability measures are shown in the APPENDIX. All formulas and explanations are included in appendix (also saved in the same folder as the article itself).

Surface effects on dynamic stability and loading during outdoor running using wireless trunk accelerometry

Dynamic postural stability parameters were quantified from tri-axial (vertical, M-L, A-P) accelerations firstly using the ratio of each linear acceleration axis root mean square (RMS) relative to the resultant vector RMS to capture variability [21]; secondly using step and stride regularity (unbiased autocorrelations procedure) to capture symmetry and consistency of running steps and strides respectively, with perfect regularity equivalent to one [25]; and thirdly using sample entropy from raw accelerations to capture the waveform predictability, with higher values indicating less periodicity or more unpredictability [27].

Wireless Tri-axial Trunk accelerometry detects deviations in dynamic center of mass motion due to running-induced fatigue

Tri-axial trunk accelerometry measures were examined using the acceleration root mean square (RMS), the RMS ratio of each axis to the resultant vector [19], step regularity and stride regularity [18], and sample entropy [25] of accelerations.

The acceleration root mean square (RMS) was calculated for each axis independently and gives an overall indication of variability of acceleration dispersion [18,19]. Next, the acceleration RMS ratio,

an indicator of the proportion of accelerations in each axis contributing to the overall movement, was calculated as the RMS of each axis relative to the resultant vector RMS [19,20].
Step and stride regularity of accelerations were computed using the unbiased autocorrelation procedure previously described by Moe Nilssen et al., [18]. Representative unbiased autocorrelation patterns of all three acceleration axes are shown in Fig 1. Step regularity, the first dominant autocorrelation peak (Ad1 in Fig 1), indicates a correlation between consecutive steps and is therefore considered the symmetry index. Since mediolateral trunk accelerations produce both positive and negative accelerations that represent left–to-right lateral trunk motion, step regularity values for the mediolateral direction are always negative (Ad1 in Fig1B). The absolute value for mediolateral step regularity was therefore used for analysis. Step frequency was computed from the vertical axis of the sequence of trunk accelerations using samples per dominant period of the autocorrelation peak and sampling frequency of the accelerometer as inputs [18] (D1 in Fig 1A). Stride regularity (Ad2 in Fig 1), the second dominant autocorrelation peak, represents a correlation between consecutive strides and can be considered as a regularity index. After normalization to the zero lag component, the maximum value (most periodic, most regular) for both step regularity and stride regularity is one.
Lastly, the sample entropy of accelerations was determined using the non-linear mathematical algorithms previously described in detail by Richman and Moorman [25] and quantifies the uncertainty or unpredictability of the accelerometry time series [34], with a larger value indicating a less periodic and less predictable or periodic pattern. In contrast to the aforementioned measures, sample entropy was analyzed from unfiltered accelerations so as not to mask or remove any dynamical properties or variability present within the system [25]. In the literature, there are two contrasting approaches in human gait analysis to select the data string length parameter for sample entropy, either according to a fixed number of samples (time) [35,36], or according to a fixed number of gait cycles [37]. In contrast to its predecessor statistic (approximate entropy), sample entropy values are more robust to shorter data strings and become stable at data strings exceeding over 2000 samples [34]–all of our trials were beyond this length to acquire 20 consecutive running steps (minimum was 2700 samples). Thus, we selected the “fixed-step” approach, also enabling consistency in number of steps selected from the sacral marker trajectory. Therefore, input parameters for our sample entropy calculation were firstly, a time series sample length (N) equivalent to 20 running steps (typical data string between 2700 to 3300 data points), secondly, a series length (m) of 2 data points, and thirdly, a tolerance window (r) normalized to 0.2 times the standard deviation of individual time series [34].