

The ground reaction force of walking: a parameter to identify patients?

Medical Data Science Project

Name of the student:

Ursula TRINLER

July 21, 2021

Contents

1	Introduction	2
1.1	Summary	2
1.2	Gait analysis and ground reaction forces	2
1.3	Discrete Parameters of the GRF and COP	4
1.4	The <i>GaitRec</i> dataset in detail	4
1.5	Research Question	6
2	Data analysis strategy	8
2.1	Preparing and exploring the dataset	8
2.2	Correlation of variables	11
2.3	Fitting different prediction models - GLM, GLMM, GAM	12
2.4	Variable selection and performance analysis	14
2.4.1	Indirectly estimate test error using specific criteria	14
2.4.2	Directly estimate test error using resampling methods	17
3	Conclusion and future approaches	19
4	References	20
5	Appendix	21
5.1	Code for pre-processed time series of GRF and COP to extract discrete parameters	21

1 Introduction

1.1 Summary

This document analyses the *GaitRec* dataset which is freely available [online](#). The dataset has, furthermore, been published by Horsak et al. in Nature, Scientific Data (Horsak et al. 2020a). It contains a large-scale dataset of *ground reaction forces* (GRF) during walking of 211 healthy (HC) and 2084 impaired persons. Included patients had different conditions affecting either the hip (H), knee (K), ankle (A), or calcaneus (C). In total, 75.732 bi-lateral walking trials have been compiled.

Before the *GaitRec* dataset will be described in more detail (Chapter 1.4) and the research question will be explored (Chapter 1.5) and analyzed (Chapter 2) a short background information about GRF and gait analysis will be given.

1.2 Gait analysis and ground reaction forces

Human movement is a complex topic and a growing research field. According to Baker (2007) Aristotle (384 until 322 BC) was one of the first known scientists to analyze human walking. It was not until the late 19th and early 20th century, however, that scientific methods started to be applied to gait analysis simulated by new measurement techniques.

Two main areas are mainly covered in human movement analysis: *kinematics* - the science of movement (without considering the forces which are responsible for the movements), and *kinetics* - the science of forces producing these movements. One such force acting on the body is the GRF which is the response of the ground to the foot contact. This GRF can be measured using so called force plates, which are normally embedded into the floor.

If a person would simply stand on such a plate it would measure the person's weight in Newton. Due to Newton's third law ("every action has an equal and opposite reaction") the measured force is the same force the person is "putting" into the ground but has opposite direction ("Newton's Laws of Motion" 2021). This means that the GRF would point vertically through the body. When the person is walking over the force plate, the GRF can increase due to acceleration forces and its direction is changing according to the center of mass of the person, see Figure 1 left.

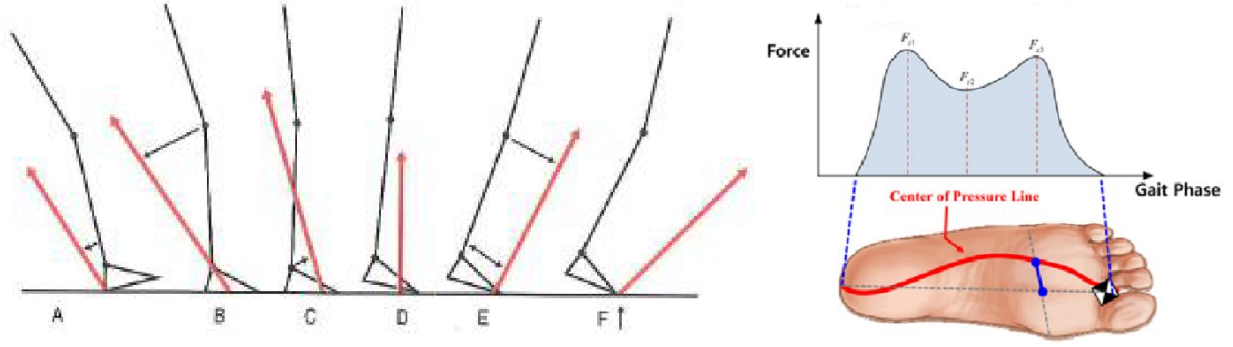


Fig. 1: Progression of GRF under the foot - sagittal view (left) and view from above (right).

The GRF is defined in a 3D space and can be split in three components, which represent the

- anterior-posterior AP (X)
- medial-lateral ML (Y) and
- vertical V (Z)

direction. The X, Y, and Z labeling is arbitrary and can change throughout publications. The X-direction defines the force produced in walking direction, the Y-direction represents frontal forces and the vertical Z- direction defines the force vertically through the body. During the stance phase of walking the GRF changes according to the acceleration and deceleration of the body on the force plate (see Figure 2).

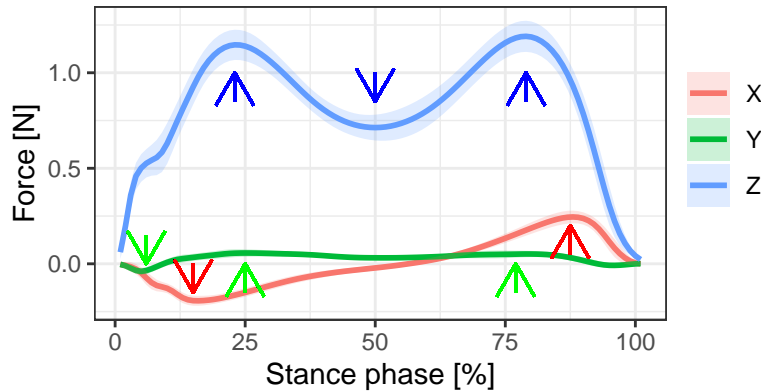


Fig. 2: GRF distribution in X, Y and Z direction during the stance phase of walking. Arrows define discrete parameters of interest.

Additionally to the three force components, another parameter related to the GRF is often analyzed, the so called *center of pressure* (COP). This parameter simply describes, where the force is centered under the foot, i.e., where the most force is concentrated. It is displayed in a 2D space onto the floor. During gait, this COP is traveling from the heel (during initial

foot contact on the floor) to the toes (at the end of stance when the foot leaves the ground). If one would imagine to observe the GRF from above, the COP would draw a line through the foot (see Figure 1 right side, adopted from Lim et al. (2017)).

Like the GRF, the COP can be divided in its components, the

- anterior-posterior AP (X) and
- medial-lateral ML (Y)

direction. The X-direction of the COP describes the propulsion of the COP in walking direction while the Y-direction is the medial-lateral shift of the COP under the foot.

1.3 Discrete Parameters of the GRF and COP

In case the musculoskeletal system is impaired, the shape of the GRF as well as the COP can change (e.g., John et al. (2012)). To understand if these changes have a clinical impact on the patient, specific discrete parameters are analyzed in detail during a clinical gait analysis. These discrete parameters are the maximum and minimum forces of each GRF component during stance phase (see Figure 2, arrows).

The length of the COP line gives additionally information about the walking stability of the patient (Gao et al. 2019). Is the length in X-direction shortened (i.e., distance between most distal to most proximal COP point) it might point to an asymmetric gait and a shorter single leg stance phase (i.e. when the contra-lateral leg is off the ground). An increase in a medial-lateral amplitude of the COP line (i.e., distance of most medial to most lateral point of the COP line) mostly implies insecurity during gait.

Important discrete parameters related to the GRF and COP can be found in Table (1).

Table 1: Discrete parameters of the GRF during walking

Variable Name	detailed information
F_V_max1	Maximum of vertical GRF in first half of stance
F_V_max2	Maximum of vertical GRF in second half of stance
F_V_min	Minimum of vertical GRF between both maxima
F_AP_max	Maximum of anterior-posterior GRF in second half of stance
F_AP_min	Minimum of anterior-posterior GRF in first half of stance
F_ML_max1	Maximum of medial-lateral GRF in first half of stance
F_ML_max2	Maximum of medial-lateral GRF in second half of stance
F_ML_min	Minimum of medial-lateral GRF at initial foot contact
COP_AP	Length of COP course in anterior-posterior direction during stance
COP_MLh	Length of COP course in medial-lateral direction during stance

1.4 The *GaitRec* dataset in detail

The GaitRec dataset consists of several .CSV files. The main document (“GRF_metadata.csv”) includes main information about the participants and measurement trial (see Table 2). The

remaining files contain the raw and pre-processed GRF and COP time series of the stance phase of walking. The raw data present the measurements purely received from the force plates, i.e., no data cleaning has been undertaken. In the pre-processed files, following steps have been undertaken to clean and tidy up the dataset:

- Force signals were filtered with a 2nd order low-pass butterworth filter with a cut-off frequency of 20Hz to reduce noise,
- time normalized to 100% stance (i.e. 101 data points) to reduce effects temporal differences, and then
- normalized to the body weight of each participant (force divided by the body mass times gravity) to reduce effects due to individual anthropometry.
- Threshold of 80N to avoid inaccuracies in COP calculation at small force values
- Medio-lateral COP coordinates were mean-centered
- Anterior-posterior coordinates zero-centered

This report includes the pre-processed GRF and COP dataset.

Table 2: Participant and trial metadata of the GaitRec dataset

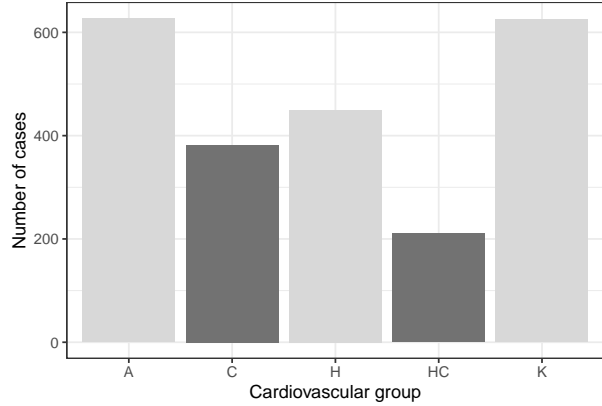
Participant	Walking trial
Sex (male, female)	Shod condition (barefoot, shoe)
Age (years)	Orthopaedic insole (yes, no)
Height (cm)	Walking speed (slow, self-selected, fast)
Body mass (kg) and weight ($\text{kg} \cdot \text{m/s}^2$)	Readmission into hospital (yes, no)
Shoe size (EU)	Session type (initial vs. control measurement)
Affected side (left, right, both)	Session date

As described in Chapter 1.1, participants can be divided into five different groups, one healthy group and four different patient groups. Number of participants per group varied slightly between subgroups (Figure 3 B), however, all exceeded 200 number of cases. Anthropometrics of each group is summarized in Figure 3 A. A more detailed visualization of age and BMI distribution between groups is shown in sub-Figure C and D. While anthropometrics are similar distributed between patient groups, the healthy control group seems to be younger with a smaller BMI than the patients. Running a Kruskal-Wallis-Test (Shapiro Test shows non-normal distribution) and a pairwise Wilcoxon test for Post-hoc analysis, differences can be shown between healthy controls and each patient group (see Figure 3 C and D). Analysis further down below will, therefore, include age and BMI as covariates.

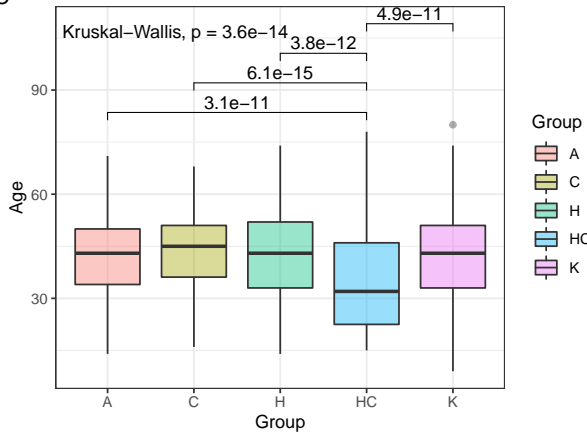
A

Mean (SD) of group anthropometrics					
	A (N = 627)	C (N = 382)	H (N = 450)	HC (N = 211)	K (N = 62)
Age (years)	41.5 (11.4)	43.5 (10.4)	42.6 (12.8)	34.7 (13.9)	41.6 (12.0)
Height (cm)	175.6 (8.0)	178.0 (7.0)	176.5 (8.0)	173.2 (8.5)	175.5 (9.1)
Body mass (kg)	86.9 (18.1)	84.0 (14.7)	82.3 (15.6)	74.0 (15.5)	84.3 (18.7)
BMI(km/m ²)	27.8 (5.0)	26.5 (4.2)	26.4 (4.3)	24.6 (4.4)	27.3 (5.2)

B



C



D

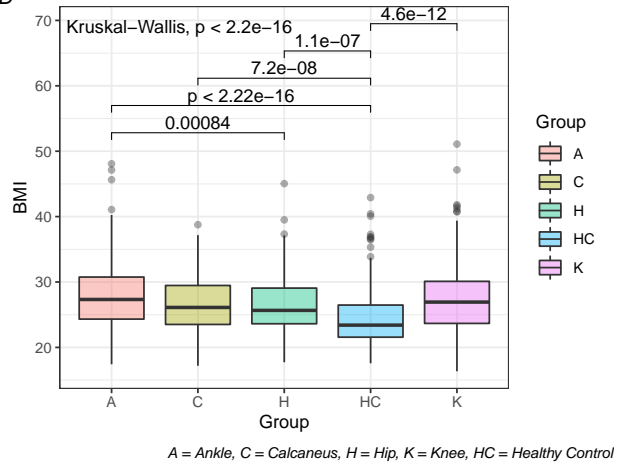


Fig. 3: Anthropometrics and number of cases of each sub-group.

1.5 Research Question

A full gait analysis is time consuming and needs trained personnel. Calculating joint angles and joint moments as well as power needs professional knowledge and results are not intuitive to understand. Therefore, simply using GRF to analyze the walking pattern would be time efficient and can be undertaken by non-trained personnel, while distinctive differences in discrete parameters of the GRF (“higher”, “lower”) can be easily interpreted. The research question of this report is, therefore, if and which discrete parameters of the GRF can be used to distinguish between patients with impairments of the lower limb and a healthy control group. Dependent variable (i.e. primary endpoint) to analyze will be the factorial variable healthy yes or no.

While age and BMI have above been shown to differ between healthy controls and patient groups and are identified as possible covariates, other main factors have been shown to influence gait analysis outcome: shod condition (Oeffinger et al. 1999) and walking speed (Schwartz, Rozumalski, and Trost 2008). Therefore, it is of high interest to understand, which other variables might have an influence on the predictions outcome to be able to

control for these covariates.

2 Data analysis strategy

2.1 Preparing and exploring the dataset

The pre-processed dataset provided by Horsak et al. (2020b) has been further processed by extracting discrete parameters (see Appendix for coding), presented in Table 1, and calculating the mean of each discrete parameter above all walking trials for each participant and each session. In another step, variables' data type have been checked, while variables “sex”, “affected side”, “shod condition”, and “speed” have been changed from numeric to factor. Also, discrete parameters, available for each leg side (left and right) had to be assigned to the healthy or to the affected side.

Within the whole dataset there are 9796 missing values, which are related to body height (and, therefore, BMI) and shoe size (3251 , 3294 respectively). In total, information of 5672 participants are complete, while 3299 participants have at least one missing value. This information is crucial for further analysis where body height (or BMI) and/or shoe size will play a role.

Before one can decide for a proper analysis strategy, the distribution of involved parameters has to be reviewed. Distribution of discrete parameters, which have been calculated, can be seen in Figures 4 to Figure 7. Most of the discrete parameters have a symmetric shape for each group and each leg side (healthy vs. affected), which indicates a normal distribution. However, especially the COP length in medial-lateral direction has a right-skewed tendency.

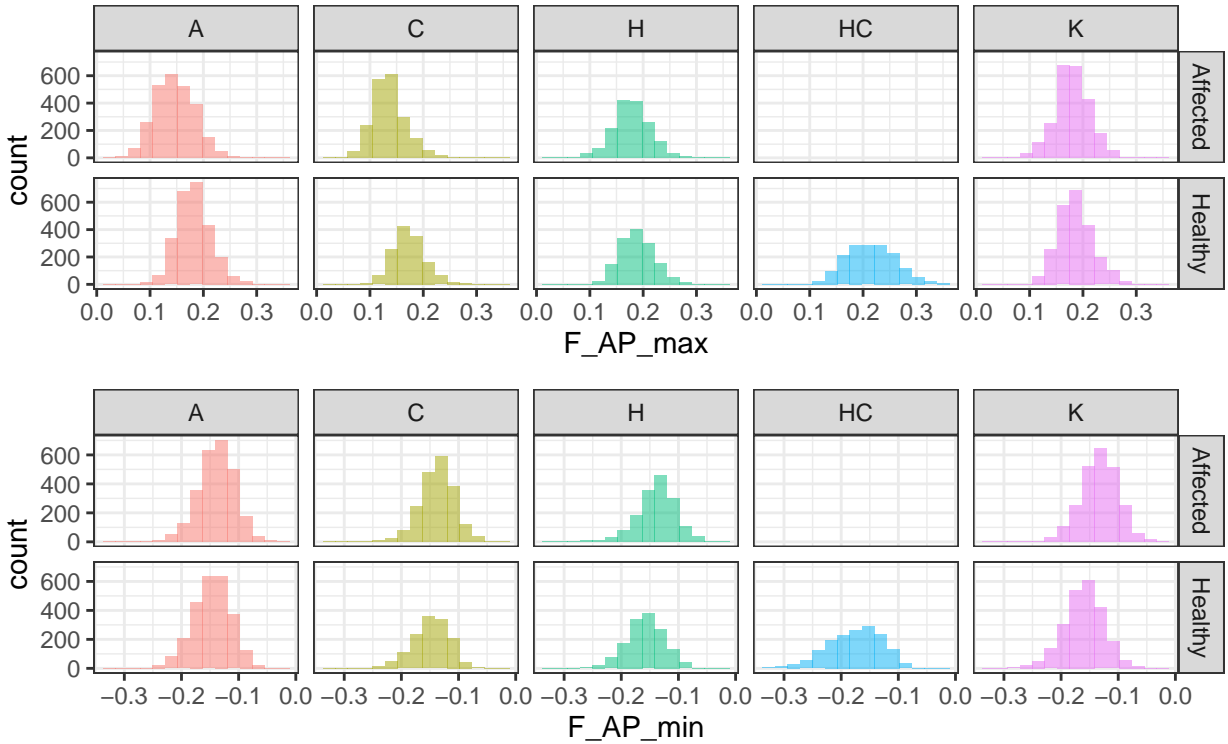


Fig. 4: Distribution of anterior-posterior GRF force divided in group and leg side.

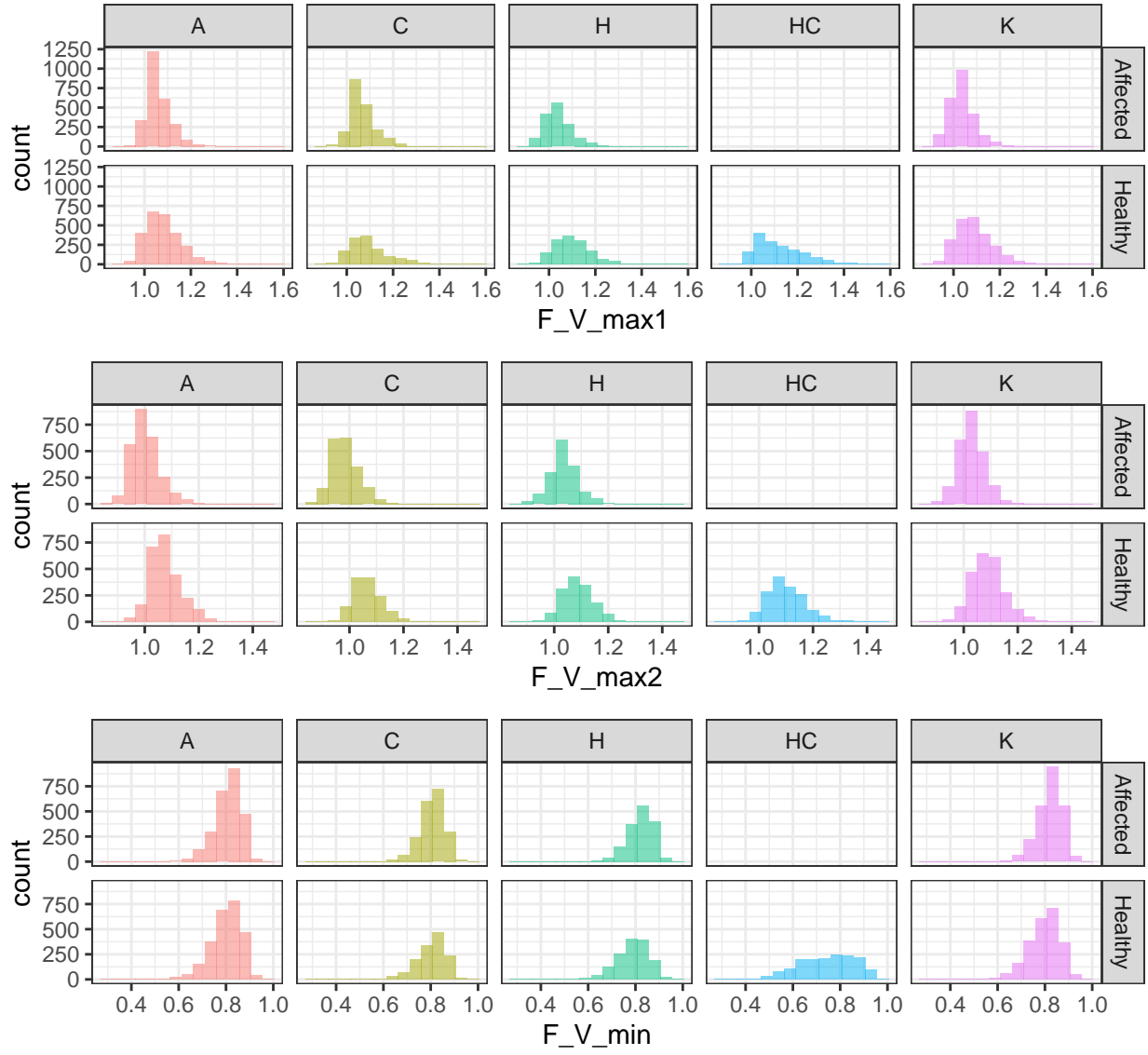


Fig. 5: Distribution of vertical GRF force divided in group and leg side.

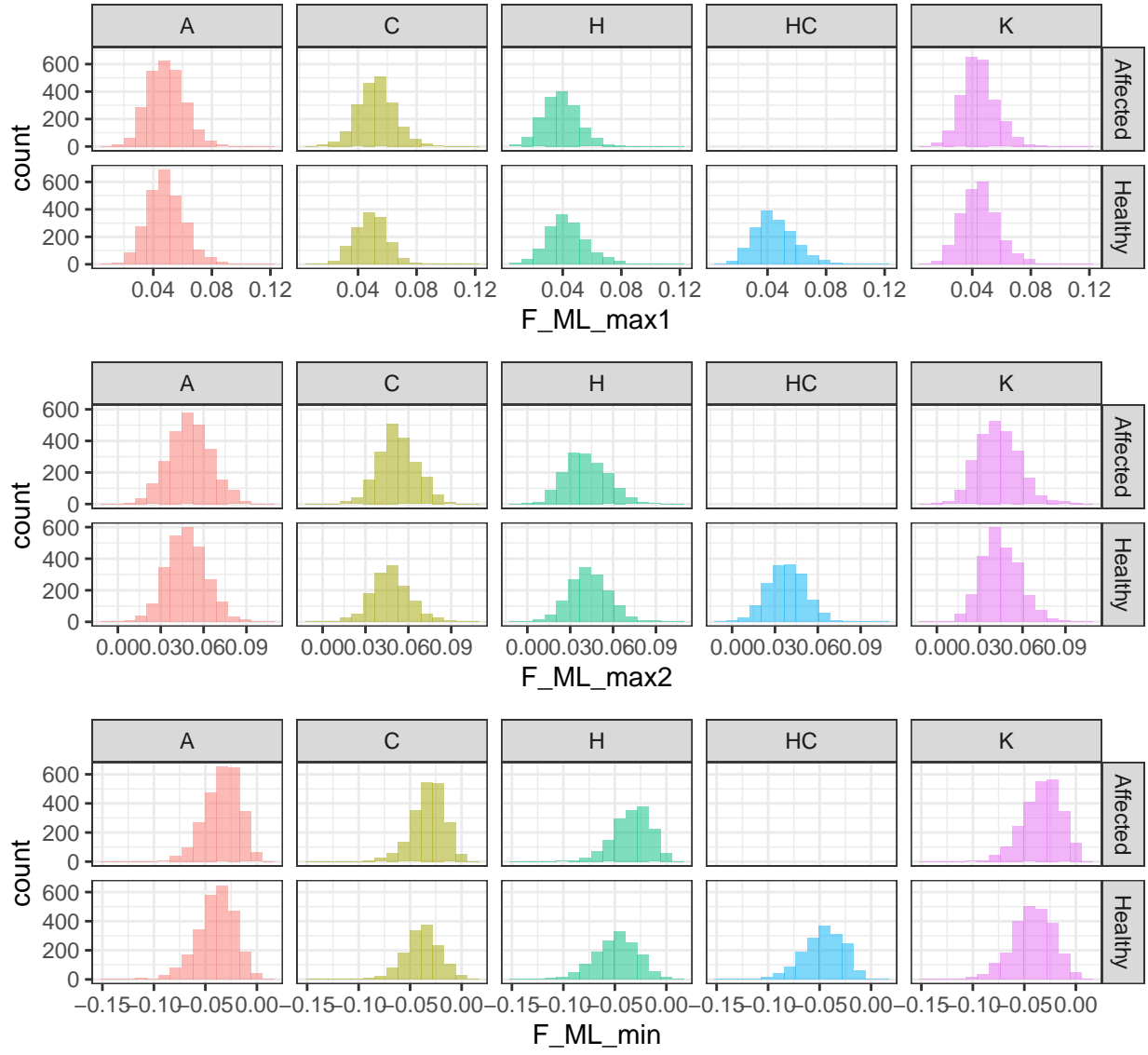


Fig. 6: Distribution of medial-lateral GRF force divided in group and leg side.

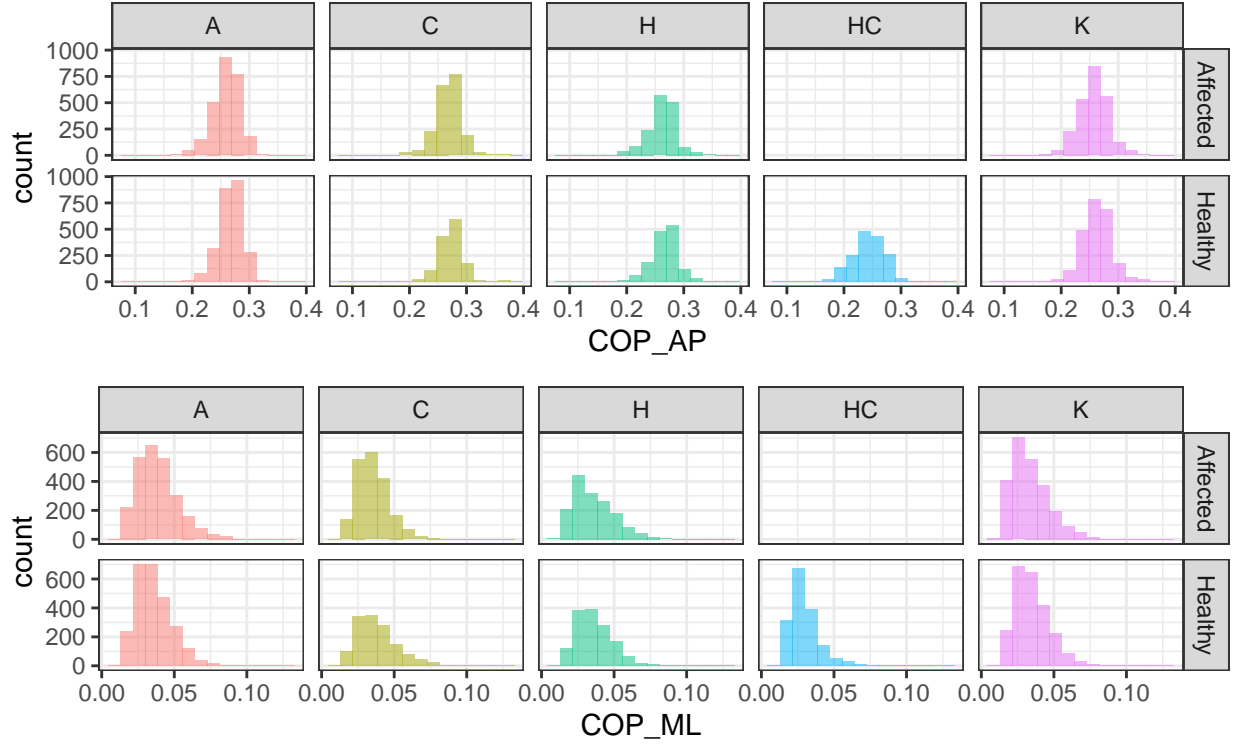


Fig. 7: Distribution of COP lengths divided in group and leg side.

2.2 Correlation of variables

Furthermore, to receive a first overview about possible interactions between variables and to later be able to apply appropriate statistics, pairwise correlation coefficients are calculated for the numeric variables. The correlation matrix in Figure 8 shows mainly low correlation coefficients between variables, however, some values exceed a correlation coefficient of ± 0.5 . Some anthropometric data correlate with each other (e.g., height and shoe size), but also discrete GRF and COP parameters with anthropometrics (e.g., length of COP in anterior-posterior direction and shoe size) or with each other (first maximum and minimum of vertical GRF or minimum of vertical and anterior-posterior GRF).

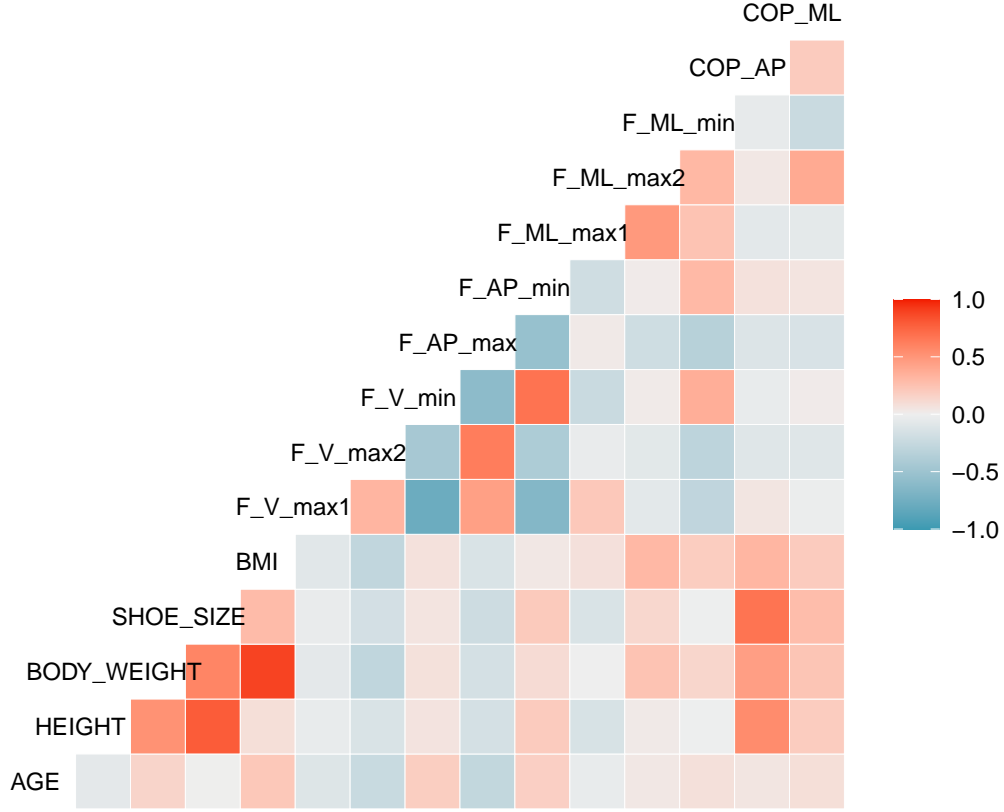


Fig. 8: Correlation matrix of the numeric predictors.

2.3 Fitting different prediction models - GLM, GLMM, GAM

One way of dealing with correlated variables is appropriate variable selection in statistical models. This, among others steps, will be considered while fitting several prediction models. The goal of this report is to build a model that predicts, if a participant has an injury of the lower limb or not (and if yes which injury) while analyzing discrete GRF and COP parameters and account for specific covariates.

In a first step, to gain an overview about the dataset, a classical generalized linear model (GLM) has been fitted onto the whole dataset, using the binomial link function. This allows the dependent variable to be non-normal, which is important for the factorial variable to distinguish between healthy and affected participants. The dependent variable was, therefore, healthy control “yes” or “no”, while discrete GRF and COP parameters as well as sex, age, height, BMI, shoe size and speed were included as independent variables.

The GLM model results in an AIC (Akaike information criterion) of 2983, while some of the discrete GRF and COP parameters seem to distinguish between healthy and affected participant (i.e., significant). Also, sex, age and height seem significantly different between healthy and affected participant ($p < 0.001$). Using this simple model, 95.69% of the participants are classified correctly.

To account for the covariates, which might influence the prediction's outcome (sex, age, height) a generalized linear mixed model (GLMM) has advantages over a classical GLM, as it can adjust for these covariates. With the `lme4` package, GLMM was implemented.

Applying the GLMM model an AIC of 5822 can be reached, which is bigger than the AIC of the GLM model, thus, seems that the simple GLM model is better fitted to the data than the GLMM model. With the GLMM model, 90.41% of the participants are classified correctly.

Sometimes the assumption of a linear relationship of independent variables with the dependent variable is doubtful. Here, a generalized additive models (GAM) can be used, which fit unknown but smooth functions to the independent variables, which are estimated using specific techniques (e.g., restricted maximum likelihood (REML)).

Fitting a GAM model on the whole GaitRec data set, 89.53% of the participants are classified correctly. The results also showed, that all discrete GRF and COP parameters have an approximate significance under $p < 0.05$. The model, however, has only an adjusted R^2 of 0.3281884 which only explains 35.5% of the deviance.

Looking onto these three models, the GLMM seems to be the optimal one, as it accounts for the covariates sex, age and body height. However, these first models were evaluated on the same dataset on which the model was build on, which should be highly avoided, as this causes overfitting. This might also explain the very high number of correctly classified participants.

To overcome the problem of overfitting, the dataset needs to be divided into a train and test dataset - a sub-dataset for training the model and another sub-dataset for testing the model. A predefined randomized partitioning of the dataset has already been provided by the authors, which will be used in this report for further analysis.

With an adapted GLMM model, where the dataset has been divided into train and test data, the AIC value is 3976, which is much smaller than the AIC of the first GLM model without using a training dataset (5822). 90.21% of the participants are classified correctly, which is even slightly higher than with the first GLM model. This model results in high significant differences between healthy and affected for F_AP_max, F_ML_max1, COP_AP and COP_ML ($p < 0.001$) and moderate significance for F_V_max1 and F_ML_min ($p < 0.01$).

Table 3 summarizes the above used models and its outcome.

Table 3: Discrete parameters of the GRF during walking

Model type	data used	AIC	Classification %
GLM	full dataset	2983	95.69
GLMM	full dataset	5822	90.41
GAM	full dataset	—	89.53
GLM	train, test	2148	96.2
GLMM	train, test	3976	90.21

2.4 Variable selection and performance analysis

Another step to undertake is to select the most important discrete GRF and COP parameters, which can distinguish between healthy and affected persons. Also, the fitted model can be analyzed in more detail to understand its performance.

2.4.1 Indirectly estimate test error using specific criteria

One way to distinguish, which discrete GRF or COP parameter is mostly relevant to the classification of healthy or affected participant is to undertake a best fit test using the command `regsubsets()` of the `leaps` package. We allow the models a maximal number of variables of 10 (i.e., all discrete parameters). The summary result of the subset models is shown in Table 4.

Table 4: Summary result of subset models

	F_V_max1	F_V_max2	F_V_min	F_AP_max	F_AP_min	F_ML_max1	F_ML_max2	F_ML_min	COP_AP	COP_ML
1 (1)				*						
2 (1)				*					*	
3 (1)				*			*		*	
4 (1)			*	*			*		*	
5 (1)			*	*	*		*		*	
6 (1)			*	*		*	*		*	*
7 (1)			*	*	*	*	*		*	*
8 (1)		*	*	*	*	*	*		*	*
9 (1)		*	*	*	*	*	*	*	*	*
10 (1)	*	*	*	*	*	*	*	*	*	*

* indicates which variables is included into which model

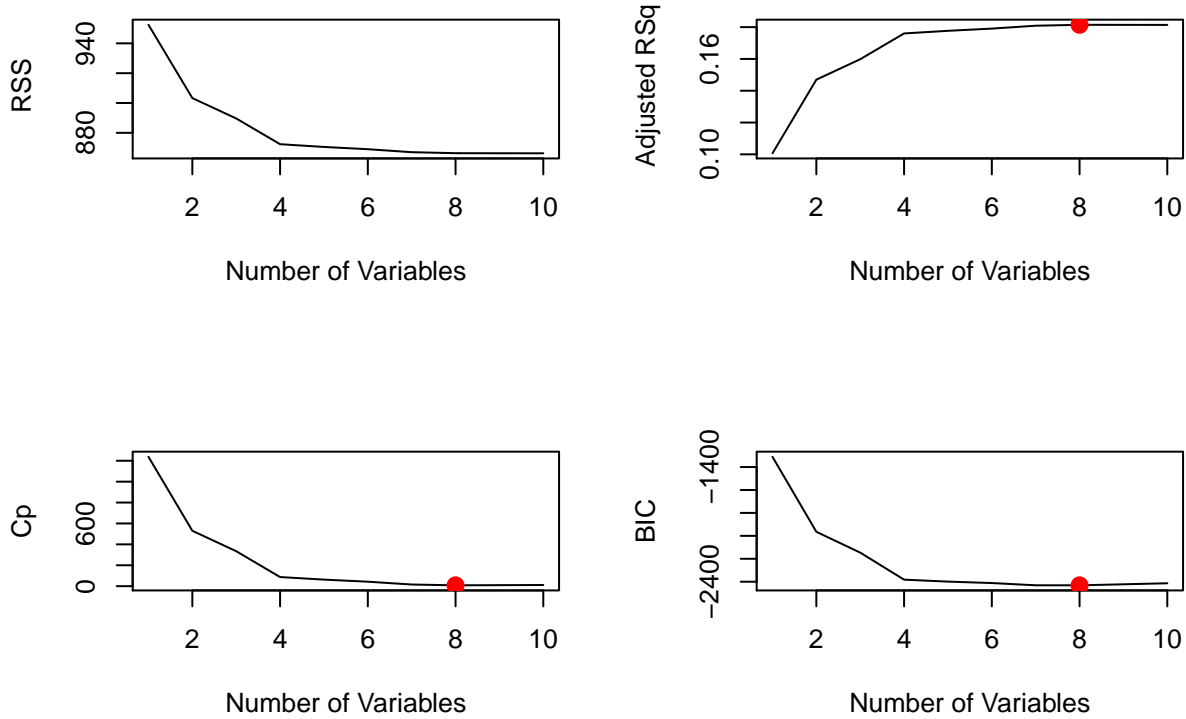


Fig. 9: RSS, adjusted R^2 , Cp, and BIC for all of the models.

Furthermore, the `summary()` function of a `regsubsets` model outputs, among others, the BIC (Schwarz's information criterion) as well as the adjusted R^2 for each subset model. By examine these, one can choose the best overall model. Figure 9 shows the RSS (residual sum of squares for each model), adjusted R^2 , Cp (Mallows' Cp), and BIC for each model. All criteria point to a model with 8 variables for the best model statistic. Therefore, the optimal model according to these criteria is the model with the discrete parameters in line 8 of Table 4.

Another way of visualizing the results of the best GLM fit model is the built-in plot command of `regsubsets`. An example is shown in Figure 10, which displays the selected variables for the best model with a given number of predictors according to the BIC. As darker the shaded areas get as better the model gets. Interestingly, looking at the BIC value, a model with four variables (i.e., discrete parameters GRF vertical min, GRF anterior-posterior max, GRF medial-lateral max2 and COP length in anterior-posterior direction) build already a sufficient model, as the BIC does not increase much anymore with increasing number of variables.

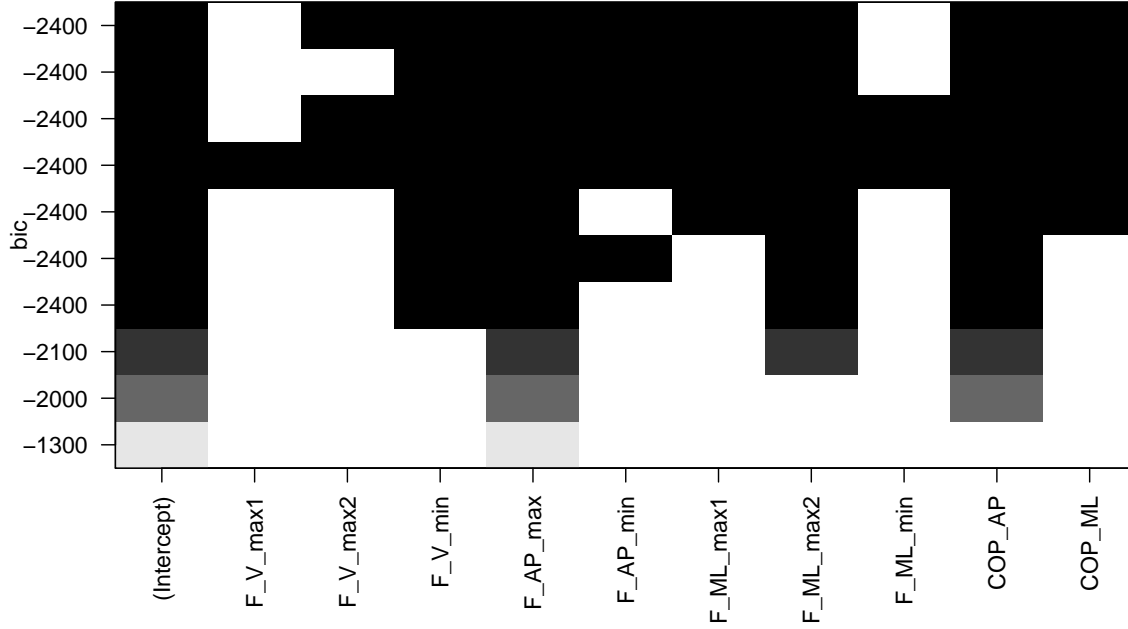


Fig. 10: BIC displayed.

The coefficient estimates associated with this 4-variables model are as follows:

Table 5: Summary result of subset models

Coefficient estimate	
(Intercept)	2.2269772
F_V_min	-0.6738005
F_AP_max	1.1858099
F_ML_max2	-2.5898225
COP_AP	-2.6012367

Adjusting the GLMM model with the 4-variables model yields an AIC value of 4109, which is, however, bigger than the AIC of the GLM model with all discrete GRF and COP parameters (3976). With the 8-variables model, which has been chosen to be the optimal one (Figure 9) an AIC of 3990, close to the AIC of the full model, indicating, that the two missing variables F_ML_min and F_V_max_1 can be neglected.

2.4.2 Directly estimate test error using resampling methods

Resampling methods refit a model using samples which have been extracted from the training data set. One method is the *K-fold Cross-Validation* resampling method. With this method, one can not only select the best model, but also give an overview about the test error of the model. Data is randomly divided in K equal-sized subsets, while K-1 subsets are used for the training and the remaining one is used for testing. This is then repeated with each subset to in the end combine all of the results.

A simple GLM model is used with k-fold=5 for cross-validation, including the 8-variable model from above. To evaluate the model, the Area Under the Curve (AUC) is plotted (Figure 11). An AUC value from around 84% is reached for the GLM model. Furthermore, another resampling method, random forest, is applied to the data with the 8-variable model, resulting in an AUC from about 83% (Figure 11). Summarized, both methods have about the same good measure of separability.

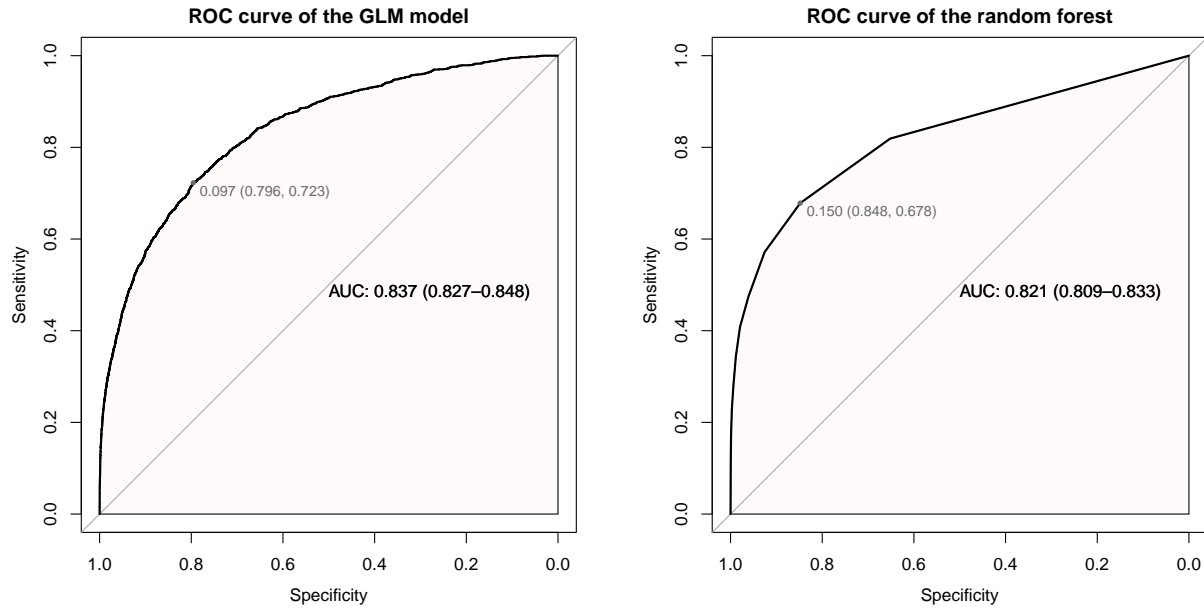


Fig. 11: AUCs for the 8-variable model during 5-fold CV

With the fitted random forest model, variable importance can be analyzed using the Gini index ('varImp' from the 'caret' package). Figure 12 shows the Mean Decreases in Gini Values, indicating higher variable importance with higher value. In the case of the applied model, vertical GRF minimum, anterior-posterior GRF maximum and the length of anterior-posterior COP are from higher importance.

Variable Importance of the Random Forest Model

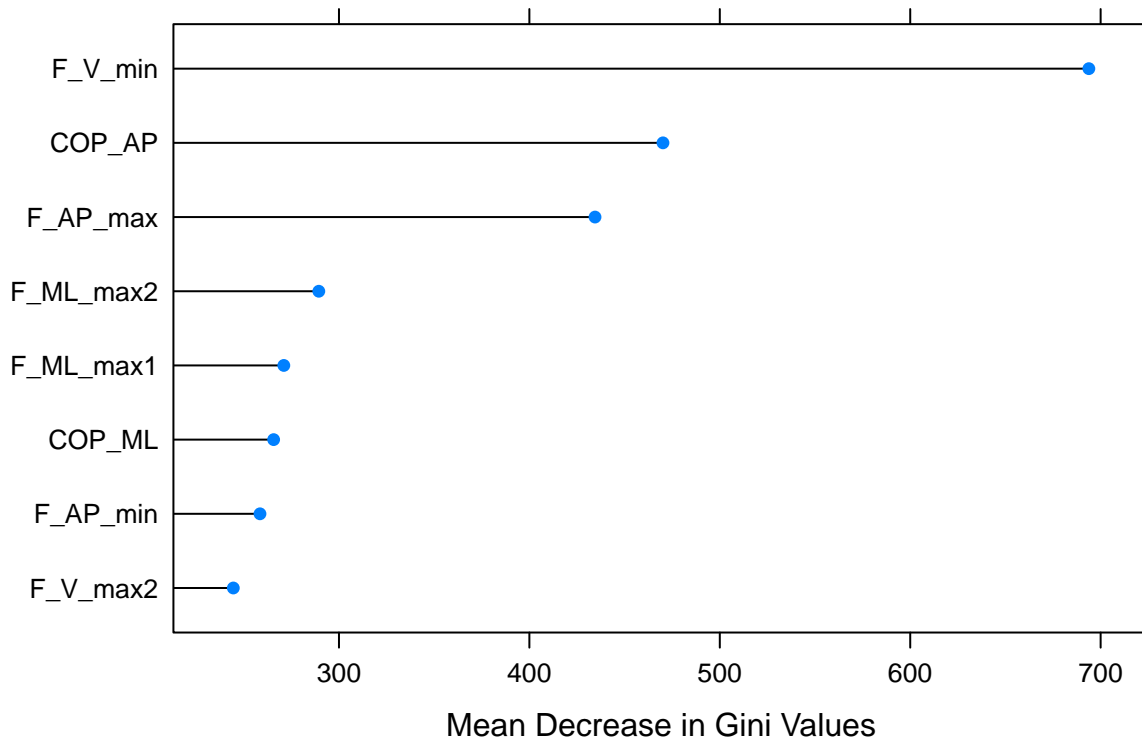


Fig. 12: Random Forest variable evaluation by Gini importance

3 Conclusion and future approaches

Summarizing the above findings, the research question, if healthy and affected persons can be distinguished using force parameters can be answered with yes. It has been shown that the first peak of the vertical GRF and the minimum of the medial-lateral GRF are from minor interest and that a GLMM is most suitable for this dataset, as sex, age and height have an interaction effect on the variable “healthy control” yes or no.

The gini index after using random forest shows that the vertical GRF at minimum between both peaks around mid-stance, the maximal anterior-posterior GRF at the end of stance and the length of the COP in walking direction are of highest importance within these models, which distinguish between healthy and affected persons.

In a next step, an appropriate model should be found to be able to answer the question, if discrete GRF and COP parameters can distinguish not only between healthy and affected persons but also where an affected person has its injury. Also, to further analyze the data set, it would be of interest which influencing factor has an effect on each discrete GRF or COP parameter, i.e., if force parameters (dependent variable) change with age, sex, or BMI as well as with different injuries of the lower limb (independent variables).

4 References

- Baker, Richard. 2007. “The History of Gait Analysis Before the Advent of Modern Computers.” *Gait & Posture* 26 (3): 331–42. <https://doi.org/10.1016/j.gaitpost.2006.10.014>.
- Gao, Chia-Chi, Jen-Suh Chern, Chun-Ju Chang, Po-Liang Lai, and Chi-Wen Lung. 2019. “Center of Pressure Progression Patterns During Level Walking in Adolescents with Idiopathic Scoliosis.” Edited by Kirsten Tulchin-Francis. *PLOS ONE* 14 (4): e0212161. <https://doi.org/10.1371/journal.pone.0212161>.
- Horsak, Brian, Djordje Slijepcevic, Anna-Maria Raberger, Catherine Schwab, Marianne Worisch, and Matthias Zeppelzauer. 2020a. “GaiTRec, a Large-Scale Ground Reaction Force Dataset of Healthy and Impaired Gait. Sci Data.” *Scientific Data* 7: 143–51. <https://doi.org/10.1038/s41597-020-0481-z>.
- . 2020b. “GaitRec, a Large-Scale Ground Reaction Force Dataset of Healthy and Impaired Gait.” *Scientific Data* 7 (1): 143. <https://doi.org/10.1038/s41597-020-0481-z>.
- John, Chand T., Ajay Seth, Michael H. Schwartz, and Scott L. Delp. 2012. “Contributions of Muscles to Mediolateral Ground Reaction Force over a Range of Walking Speeds.” *Journal of Biomechanics* 45 (14): 2438–43. <https://doi.org/10.1016/j.jbiomech.2012.06.037>.
- Lim, Dong-Hwan, Wan-Soo Kim, Ho-Jun Kim, and Chang-Soo Han. 2017. “Development of Real-Time Gait Phase Detection System for a Lower Extremity Exoskeleton Robot.” *International Journal of Precision Engineering and Manufacturing* 18 (5): 681–87. <https://doi.org/10.1007/s12541-017-0081-9>.
- “Newton’s Laws of Motion.” 2021. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Newton%27s_laws_of_motion&oldid=1032605773.
- Oeffinger, Donna, Bridgett Brauch, Shelley Cranfill, Cary Hisle, Cyndi Wynn, Ramona Hicks, and Sam Augsburg. 1999. “Comparison of Gait with and Without Shoes in Children.” *Gait & Posture* 9 (2): 95–100. [https://doi.org/10.1016/S0966-6362\(99\)00005-3](https://doi.org/10.1016/S0966-6362(99)00005-3).
- Schwartz, Michael H., Adam Rozumalski, and Joyce P. Trost. 2008. “The Effect of Walking Speed on the Gait of Typically Developing Children.” *Journal of Biomechanics* 41 (8): 1639–50. <https://doi.org/10.1016/j.jbiomech.2008.03.015>.

5 Appendix

5.1 Code for pre-processed time series of GRF and COP to extract discrete parameters

```
# read in data
metaData <- read.csv('GRF_metadata.csv')
GRF_F_V_r <- read.csv('GRF_F_V_PRO_right.csv')
GRF_F_V_l <- read.csv('GRF_F_V_PRO_left.csv')
GRF_F_ML_r <- read.csv('GRF_F_ML_PRO_right.csv')
GRF_F_ML_l <- read.csv('GRF_F_ML_PRO_left.csv')
GRF_F_AP_r <- read.csv('GRF_F_AP_PRO_right.csv')
GRF_F_AP_l <- read.csv('GRF_F_AP_PRO_left.csv')
GRF_COP_AP_r <- read.csv('GRF_COP_AP_PRO_right.csv')
GRF_COP_AP_l <- read.csv('GRF_COP_AP_PRO_left.csv')
GRF_COP_ML_r <- read.csv('GRF_COP_ML_PRO_right.csv')
GRF_COP_ML_l <- read.csv('GRF_COP_ML_PRO_left.csv')

### Extract variables of interest and include with meta data

# Variable for each row
GRF_F_V_r_max1 <- apply(GRF_F_V_r[,4:53],1,max)
GRF_F_V_r_max2 <- apply(GRF_F_V_r[,54:104],1,max)
GRF_F_V_r_min <- apply(GRF_F_V_r[,35:70],1,min)
GRF_F_V_l_max1 <- apply(GRF_F_V_l[,4:53],1,max)
GRF_F_V_l_max2 <- apply(GRF_F_V_l[,54:104],1,max)
GRF_F_V_l_min <- apply(GRF_F_V_l[,35:70],1,min)
GRF_F_AP_r_max <- apply(GRF_F_AP_r[,4:104],1,max)
GRF_F_AP_r_min <- apply(GRF_F_AP_r[,4:104],1,min)
GRF_F_AP_l_max <- apply(GRF_F_AP_l[,4:104],1,max)
GRF_F_AP_l_min <- apply(GRF_F_AP_l[,4:104],1,min)
GRF_F_ML_r_max1 <- apply(GRF_F_ML_r[,27:54],1,max)
GRF_F_ML_r_max2 <- apply(GRF_F_ML_r[,68:98],1,max)
GRF_F_ML_r_min <- apply(GRF_F_ML_r[,4:14],1,min)
GRF_F_ML_l_max1 <- apply(GRF_F_ML_l[,27:54],1,max)
GRF_F_ML_l_max2 <- apply(GRF_F_ML_l[,68:98],1,max)
GRF_F_ML_l_min <- apply(GRF_F_ML_l[,4:14],1,min)
GRF_COP_AP_r_length <- (apply(GRF_COP_AP_r[,4:104],1,max) -
                        apply(GRF_COP_AP_r[,4:104],1,min))
GRF_COP_AP_l_length <- (apply(GRF_COP_AP_l[,4:104],1,max) -
                        apply(GRF_COP_AP_l[,4:104],1,min))
GRF_COP_ML_r_length <- (apply(GRF_COP_ML_r[,4:104],1,max) -
                        apply(GRF_COP_ML_r[,4:104],1,min))
```

```

GRF_COP_ML_l_length <- (apply(GRF_COP_ML_l[,4:104],1,max) -
                        apply(GRF_COP_ML_l[,4:104],1,min))

# mean for each session
varData <- cbind(GRF_F_V_r[1:3],GRF_F_V_r_max1,
                 GRF_F_V_l_max1,
                 GRF_F_V_r_max2,
                 GRF_F_V_l_max2,
                 GRF_F_V_r_min,
                 GRF_F_V_l_min,
                 GRF_F_AP_r_max,
                 GRF_F_AP_l_max,
                 GRF_F_AP_r_min,
                 GRF_F_AP_l_min,
                 GRF_F_ML_r_max1,
                 GRF_F_ML_l_max1,
                 GRF_F_ML_r_max2,
                 GRF_F_ML_l_max2,
                 GRF_F_ML_r_min,
                 GRF_F_ML_l_min,
                 GRF_COP_AP_r_length,
                 GRF_COP_AP_l_length,
                 GRF_COP_ML_r_length,
                 GRF_COP_ML_l_length)

varData_av <- varData %>% group_by(SUBJECT_ID,SESSION_ID) %>%
  summarize(GRF_F_V_r_max1 = mean(GRF_F_V_r_max1),
            GRF_F_V_l_max1 = mean(GRF_F_V_l_max1),
            GRF_F_V_r_max2 = mean(GRF_F_V_r_max2),
            GRF_F_V_l_max2 = mean(GRF_F_V_l_max2),
            GRF_F_V_r_min = mean(GRF_F_V_r_min),
            GRF_F_V_l_min = mean(GRF_F_V_l_min),
            GRF_F_AP_r_max = mean(GRF_F_AP_r_max),
            GRF_F_AP_l_max = mean(GRF_F_AP_l_max),
            GRF_F_AP_r_min = mean(GRF_F_AP_r_min),
            GRF_F_AP_l_min = mean( GRF_F_AP_l_min),
            GRF_F_ML_r_max1 = mean(GRF_F_ML_r_max1),
            GRF_F_ML_l_max1 = mean(GRF_F_ML_l_max1),
            GRF_F_ML_r_max2 = mean( GRF_F_ML_r_max2),
            GRF_F_ML_l_max2 = mean(GRF_F_ML_l_max2),
            GRF_F_ML_r_min = mean(GRF_F_ML_r_min),
            GRF_F_ML_l_min = mean(GRF_F_ML_l_min),
            GRF_COP_AP_r_length = mean(GRF_COP_AP_r_length),
            GRF_COP_AP_l_length = mean(GRF_COP_AP_l_length),

```

```
GRF_COP_ML_r_length = mean(GRF_COP_ML_r_length),  
GRF_COP_ML_l_length = mean(GRF_COP_ML_l_length))  
  
allData <- cbind(metaData, varData_av[, 3:22])
```