

Predicting Gait Time and TUG Time Using Foot-Tapping Sensor Features

Logan Marshall

Abstract

The objective of this project is to identify safer and more accurate methods for predicting fall risk in adults by examining whether foot-tapping performance is associated with outcomes from two established functional mobility tests: 4-metre gait speed and the Timed Up and Go (TUG) test. The primary research question is: *Which foot-tapping feature set best predicts gait time and TUG time?* A variety of modelling approaches were applied, including stepwise regression, regularization methods, ensemble learning, kernel-based models, and dimensionality-reduction techniques. Visual analyses, including Bland–Altman plots, were also used to evaluate model agreement. Overall, the nondominant foot dataset produced the strongest predictive models. A linear model containing the nonzero lasso-selected coefficients achieved the best performance for gait time, with an average prediction error of 0.795 seconds. For TUG time, the random forest model yielded the lowest error, with an average prediction error of approximately 2.391 seconds. These findings demonstrate the feasibility of using high-dimensional wearable-sensor data to support accurate and interpretable predictions of clinically relevant mobility assessments.

Introduction

The 4-metre gait speed test requires a participant to walk a straight 4-metre path at their usual comfortable pace from a static start on a flat surface, using their walking aid if needed. Timing begins when the first foot crosses the start line and ends when the first foot crosses the 4-metre mark (Bohannon & Wang, 2019). Gait speed is a strong predictor of mobility status and fall risk in older adults.

The TUG test measures the time taken for an individual to rise from a standard armchair, walk 3 meters, turn around, return to the chair, and sit down again at a comfortable and safe pace. Participants wear their usual footwear and may use any walking aid without receiving physical assistance. Before timing, a practice trial is performed to ensure familiarity with the task (Podsiadlo & Richardson, 1991).

Data was collected from 85 adults without Parkinson’s disease, approximately three-quarters of whom were classified as older adults (≥ 65 years). Each participant wore foot-mounted sensors equipped with 3-axis inertial measurement units (IMUs), generating six signals in total (three gyroscope axes and three accelerometer axes). The sensor used for both accelerometer and gyroscope measurements was a Trigno wireless biofeedback system. Figure 1 illustrates the positive foot-tapping direction for each axis and provides an example of four sensors positioned within the Trigno system’s module slots used for data transmission (Noroozi et al., 2025) (Delsys Incorporated, 2023). From these measurements, 179 engineered features were extracted in MATLAB, capturing time- and frequency-domain characteristics in each plane, as well as magnitude-based metrics.

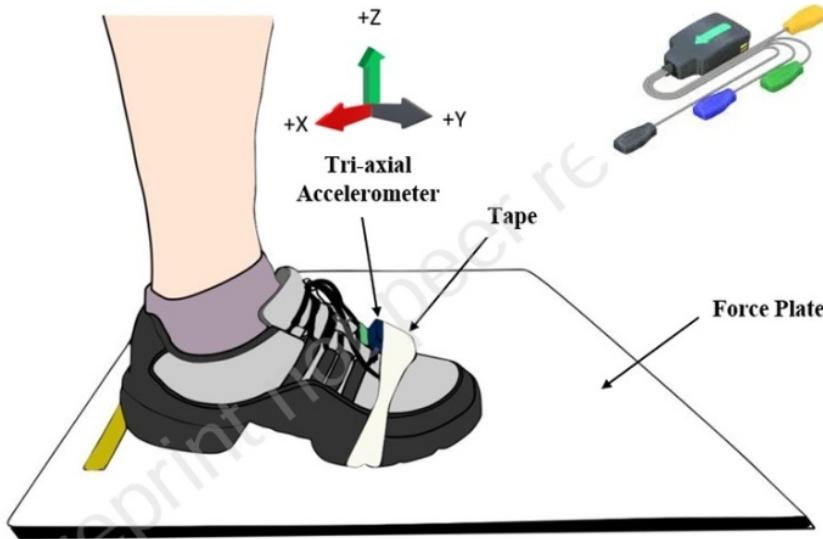


Figure 1: Positive IMU axis directions and example Trigno sensor placement.

Although many of these engineered features are highly correlated, our primary objective is predictive accuracy. While reducing multicollinearity is beneficial, our focus is on minimizing root mean squared error (RMSE) to best assess predictive performance. RMSE is a common regression metric defined as the square root of the average squared difference between predicted and observed values, providing a direct measure of a model’s typical prediction error (James et al., 2013). Multiple modelling approaches are explored in this project, including filter- and wrapper-based feature selection methods, ensemble learning, regularization techniques, kernel methods, and dimensionality reduction strategies, in order to identify the optimal feature subset and model for predicting gait time and TUG time.

Methods

Data Cleaning and Preprocessing

In this dataset, the number at the end of each feature name indicates the axis along which the signal was measured. Axis 1 corresponds to the X-axis (left–right), with positive values indicating rightward movement; Axis 2 corresponds to the Y-axis (front–back), with positive values indicating forward movement; and Axis 3 corresponds to the Z-axis (up–down), where positive values reflect upward movement. Axis 4 refers to the vector magnitude, which is calculated only for accelerometer features and represents the combined acceleration across all three spatial axes. Features beginning with acc- are derived from the accelerometer, whereas those beginning with gyro- come from the gyroscope. Time-domain features describe how the signal changes over time, while time–frequency features characterize how the signal’s energy is distributed across both time and frequency, similar to what is observed in a spectrogram.

Some features were derived by summing components from decomposed wavelength signals using the Hilbert–Huang Transform (Liu et al., 2020). Additional features were extracted from Continuous Wavelet Transform (CWT) scalograms (Bobić et al., 2018). More detailed descriptions of the features can be found in the feature description file included in the GitHub repository.

The dataset was originally provided in two separate sheets, one containing measurements from participants performing the foot-tapping test with their dominant foot and the other with their nondominant foot. These time-series signals were processed to derive a large set of statistical features describing movement dynamics. Most participants completed both experiments, resulting in 84 observations for the dominant foot and 81 for the nondominant. The difference in sample sizes is due to the removal of faulty or excessively noisy observations to ensure data accuracy and integrity. Irrelevant columns that did not contribute to addressing the research questions were also removed using the **dplyr** package, reducing the total number of variables from 179 to 138 (Wickham et al., 2023). After excluding several categorical or classification features, the dataset consisted solely of numerical predictors suitable for regression analysis.

For analyses involving both the dominant and non-dominant feet, only participants with data available for both feet were retained to ensure consistency across datasets. This filtering step reduced the participant pool from 85 to 80 individuals. In all other analyses focused on a single foot, all available participants were included. The data was then partitioned into training (70%) and testing (30%) subsets using the **caret** package (Kuhn & Max, 2008). Separate partitions were created for the two response variables, Gait Time and TUG Time, to enable independent model development and evaluation. To address multicollinearity among predictors, a correlation-based feature reduction was applied to the training dataset specifically for the stepwise regression models, which assume low intercorrelation among predictors. For each model, a correlation matrix was computed among all predictor variables in the training set. Starting with a high correlation threshold of 0.95, the cutoff was iteratively lowered in steps of 0.05 until stepwise regression was able to execute successfully. This process yielded a final cutoff of 0.8 for gait and 0.75 for TUG. For analyses involving both the dominant and non-dominant feet, stepwise regression could not be applied, as the combination of predictors from both feet created an excessive number of columns relative to the number of observations. Accommodating would require lowering the correlation cutoff to a level that would remove too

many important predictors, potentially discarding critical information. Predictor pairs exceeding these thresholds were reduced using the `findCorrelation()` function from the `caret` package, which removes the variable more highly correlated with the rest of the predictors (Kuhn & Max, 2008). A schematic illustration of the variable filtering process is shown in Figure 2. This feature reduction decreased redundancy, resolved aliasing issues, and reduced dimensionality to a level where the number of predictors was only slightly greater than the number of observations, enabling stepwise regression to run properly. For our other models, this feature reduction step was not necessary.

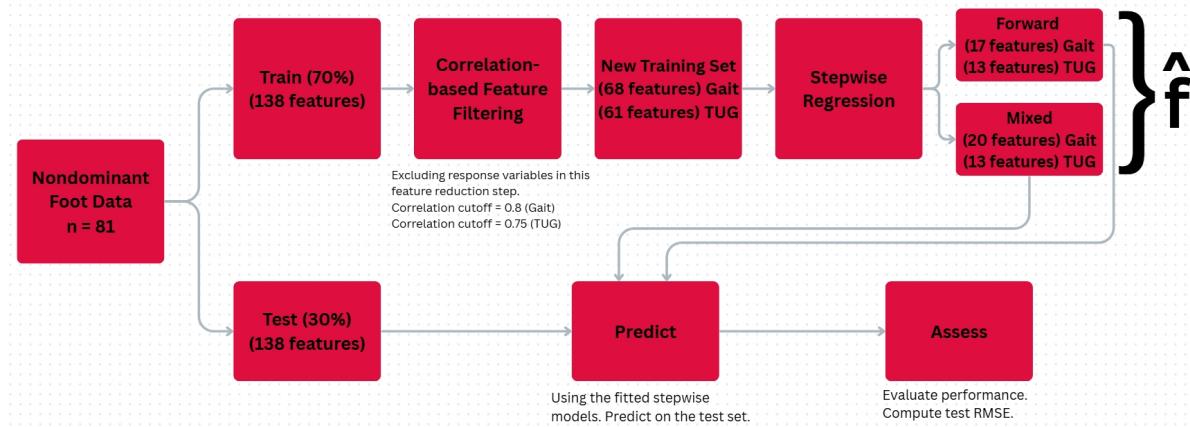


Figure 2: Schematic of the nondominant foot stepwise regression workflow for predicting Gait Time and TUG Time

Merging dominant and nondominant foot features into a single model, as well as modeling the dominant foot separately, did not improve predictive performance. Both approaches produced consistently higher RMSE values than the nondominant only models. As a result, the analyses in this report focus solely on models trained for the nondominant foot.

Stepwise

Applied forward selection and mixed stepwise regression on the pre-filtered dataset using the `step()` function from R's `stats` package, with Akaike Information Criterion (AIC) as the selection criterion (R Core Team, 2024). The model search explored a range from an intercept-only model to a full model containing all predictor variables. Due to the dataset's high dimensionality, more features than observations despite filtering highly correlated variables, methods like backward elimination and recursive feature elimination could not be computed. Forward and mixed selection, which start from a smaller model, were able to handle this scenario without issue.

Lasso

Applied lasso regression to the original 136 predictor variables using the **glmnet** package, setting $\alpha = 1$ and performing 10-fold cross-validation on the training set to identify the optimal regularization parameter (Tay et al., 2023). The optimal lambda was chosen, corresponding to the minimum cross-validated mean squared error. All predictors were automatically standardized by **glmnet** prior to fitting (Tay et al., 2023).

Ridge

Ridge regression was applied to the full set of predictors using the **glmnet** package, setting $\alpha = 0$ to implement pure ridge regularization (Tay et al., 2023). Predictors were standardized by **glmnet** prior to fitting (Tay et al., 2023). 10-fold cross-validation was used to identify the optimal penalty parameter corresponding to the minimum cross-validated mean squared error. Unlike lasso, ridge does not perform feature selection, but it is particularly effective when dealing with multicollinearity or when the number of predictors exceeds the number of observations (Chan et al., 2022).

Elastic Net

An elastic net regression was used to predict the response variable, combining lasso and ridge regularization to balance feature selection and coefficient shrinkage. The model was trained across α values from 0 to 1 (in 0.1 increments), with each configuration evaluated using 100 random 70/30 train and test splits. For each split, a **glmnet** 10-fold cross-validation identified the optimal penalty parameter lambda that minimized cross-validation error (Tay et al., 2023). Test set RMSE was computed for performance assessment, and feature selection was tracked across runs. The alpha with the lowest average RMSE was chosen as optimal. Finally, using a fixed seed for reproducibility, the model was recomputed with the best alpha, and its predictive accuracy was evaluated on both the training and test sets. This approach of elastic net was performed in similar literature (Choi et al., 2021).

Random Forest

A random forest model was trained on the training set using the **randomForest** package, with Gait/TUG time as the response variable and the unused dependent variable removed from the predictors (Liaw & Wiener, 2002). The number of candidate variables considered at each split was set to the default $m = p/3$ where $p = 136$ is the total number of predictors, and 500 trees were grown to ensure the out-of-bag (OOB) error estimate stabilized. When random forest emerged as one of the strongest models based on test RMSE, additional values of m near the default were explored in increments of 5 to assess whether small improvements in predictive accuracy could be obtained. Variable importance was computed using both the percentage increase in mean squared error (%IncMSE) and the increase in node purity (IncNodePurity), and the top ten predictors were visualized using a variable importance plot.

Bagging

A bagged ensemble model was trained using the **randomForest** package with the same data preparation procedure as the random forest model, excluding gait or TUG from the predictors (Liaw & Wiener, 2002). The number of candidate variables considered at each split was set to the total number of predictors $m = p$, effectively performing bootstrap aggregation (bagging) without random feature selection. As with the random forest, 500 trees were grown and variable importance was computed.

Boosting

Boosting was implemented using the **gbm** package with a Gaussian distribution and 5-fold cross-validation to determine the optimal number of trees (Ridgeway & Developers, 2024). The model was initially trained with 5000 trees, a learning rate (shrinkage) of 0.01, and an interaction depth of 1, corresponding to simple additive (stump-based) models. These values for learning rate and interaction depth were chosen because they are commonly used in gradient boosting machine studies (Konstantinov & Utkin, 2021). The optimal number of trees for each foot separately was then used to generate predictions on both the training and test sets.

Support Vector Machines

Support Vector Machines (SVM) were implemented using the **e1071** and **kernlab** packages (Meyer et al., 2024) (Karatzoglou et al., 2004). Both linear and radial basis function (RBF) kernels were initially explored, with the RBF kernel selected for its greater flexibility and the inclusion of an additional hyperparameter, gamma (γ), which allowed for finer control over model complexity. A 10-fold cross-validation grid search was performed to identify the optimal combination of gamma and cost (C) values based on the lowest cross-validation RMSE, testing gamma set {0.001, 0.01, 0.1} and C set {0.01, 0.1, 1, 10}. Selected these particular values to include in the set as they were commonly found in literature (James et al., 2013).

Principal Component Regression

Principal component regression (PCR) was performed on the dataset using the **prcomp** function, with all predictor variables scaled and response variables excluded. To determine the number of components to retain, multiple criteria were compared. The Kaiser criterion suggested keeping components with eigenvalues greater than one, while the scree plot provided a visual “elbow” point indicating where the marginal gain in explained variance begins to level off. Additionally, a threshold of approximately 80% cumulative variance explained was considered to ensure that a substantial portion of the variability in the training data was captured, consistent with similar literature (Rahmat et al., 2024). The selected principal components were then used in a linear model to predict the response variable. Model performance was evaluated on the training set and projected onto the test set using the principal component analysis rotation derived from the training data.

Partial Least Squares Regression

Used Partial least squares regression (PLSR) to predict the response variables using the **pls** package with the orthogonal scores algorithm ('oscorespls'), ensuring stable computation of components under high collinearity, scaling all predictors and excluding the response variables (Liland et al., 2024). Unlike PCR, PLSR identifies components that simultaneously capture variance in the predictors and maximize covariance with the response, allowing for a more efficient dimensionality reduction (Boulesteix & Strimmer, 2007).

Bland-Altman Analysis

A Bland-Altman analysis was used to assess the agreement between the predicted values from the model and the corresponding actual measurements. This method is widely applied in similar literature when comparing predicted versus observed TUG performance (Choi et al., 2021). In this study, Bland-Altman plots were generated for both TUG and gait to evaluate the reliability of the developed models.

For each plot, the mean difference (bias) between predicted and actual values was calculated, along with the 95% limits of agreement defined as $\text{bias} \pm 1.96 \times SD(\text{differences})$. The bias is shown as a solid line, and the limits of agreement are shown as dashed lines. These limits represent the expected range within which most prediction errors should fall under normal model performance.

Gait Results

Results for the gait time models are presented in Table 1.

Stepwise

After the correlation filter, 68 features remained for the nondominant foot. Forward and mixed stepwise produced highly similar models. Mixed stepwise selected 20 features and forward selected 17. Both models included a mixture of gyroscope and accelerometer domain features, along with age. Mixed stepwise produced lower training RMSE and higher adjusted R-squared ($R^2_{Train,Adj}$), whereas forward selection yielded the lower test RMSE. Both models showed noticeably higher test RMSE than training RMSE.

Table 1: Nondominant Foot Gait Time Model Summaries (Bold Indicates Lowest Test RMSE)

Model	Number of Features Selected	$R^2_{\text{Train,Adj}}$	$\text{RMSE}_{\text{Train}}$	$\text{RMSE}_{\text{Test}}$
Lasso	5	0.317	0.835	0.773
Lasso Fitted	5	0.331	0.788	0.795
Ridge	136	0.315	0.836	0.747
Elastic Net, $\alpha = 0.1$	27	0.289	0.852	0.774
Forward Selection	17	0.564	0.557	1.078
Mixed Stepwise	20	0.692	0.450	1.384
Random Forest	136	0.076	0.971	0.755
Bagging	136	0.038	0.991	0.767
Boosting	136	0.365	0.805	0.764
Support Vector Machine	136	0.359	0.808	0.755
PCR	136	0.271	0.764	0.853
PLSR	136	–	0.434	1.129

Lasso

Lasso performed very effectively in reducing the dimensionality of the dataset, selecting only 5 features for nondominant foot out of the original 136 predictors at optimal $\lambda = 0.2026074$. Predictive performance was strong, with test RMSE among the best observed across all models. To further assess the impact of variable selection, the selected features from lasso were fitted in a standard linear model. The training fit improved slightly, while the test RMSE increased only marginally (+0.022 seconds). A deeper examination of the summary table for the linear model (Table 2) shows that only one feature, the standard deviation of the intertap interval from the accelerometer (`acctime_ITIstd`), was statistically significant at the $\alpha = 0.05$ level, along with the intercept. Four of the five selected predictors were derived from accelerometer features, with the remaining one being participant age. The overall model p -value was extremely small (8.996e-05).

Table 2: Summary of the Nondominant LASSO-Selected Linear Model for Gait Time

Term	Estimate	Std. Error	t value	p value
(Intercept)	3.13941	1.21968	2.57396	0.01300
acctime_ITIstd	4.38596	1.81723	2.41354	0.01943
Age.numerical	0.00635	0.00622	1.02143	0.31187
acctime_Powerlog_3	-0.00003	0.00003	-1.17394	0.24587
accTimeFrequency_SDSumenergy_2	-5.62929	5.31784	-1.05857	0.29479
accTimeFrequency_MeanSumenergy_3	-0.05117	4.25786	-0.01202	0.99046

Ridge

Ridge regression produced the lowest test RMSE for gait time, using optimal $\lambda = 14.97415$. Ridge retained all 136 predictors and produced stable predictions.

Elastic Net

The average RMSE for all α values were very similar. However it was identified that 0 was the optimal value. For the fixed seed retrain, I decided to use 0.1 instead of 0 since ridge regression is already represented in our results table, as this was the alpha with the second lowest RMSE and a better representative of elastic net regularization. The elastic net simulation performed similarly to lasso but did not perform better than ridge regression in terms of test set RMSE.

Random Forest

Random forest produced a weak training fit but a competitive test RMSE, only slightly behind ridge. A small range of m values was examined, but results remained consistent and did not surpass ridge. Therefore, the default number of predictors considered at each split was retained. In the variable importance plot for the nondominant foot (Figure 3), the top three predictors in terms of percent increase in mean squared error were the minimum signal and variance of the z-axis accelerometer (`acctime_Minsignal_3`, `acctime_Varsignal_3`) and the standard deviation of summed foot-tapping energy in the y-axis (`accTimeFrequency_SDSumenergy_2`). Feature importance values decreased notably after these top three predictors. All of the top ten predictors (%IncMSE) were derived from accelerometer features.

Bagging

The model's performance on the train and test set was very similar to that of the random forest, showing limited improvement and overall slightly weaker predictive accuracy compared to ridge regression and random forest. The variable importance plot also revealed some of the same influential features as those identified by the random forest model.

Boosting

Cross-validation selected 118 trees for the nondominant foot, which was then used for predictions. Boosting achieved the second lowest test error among the ensemble methods, performing slightly worse than random forest but better than bagging.

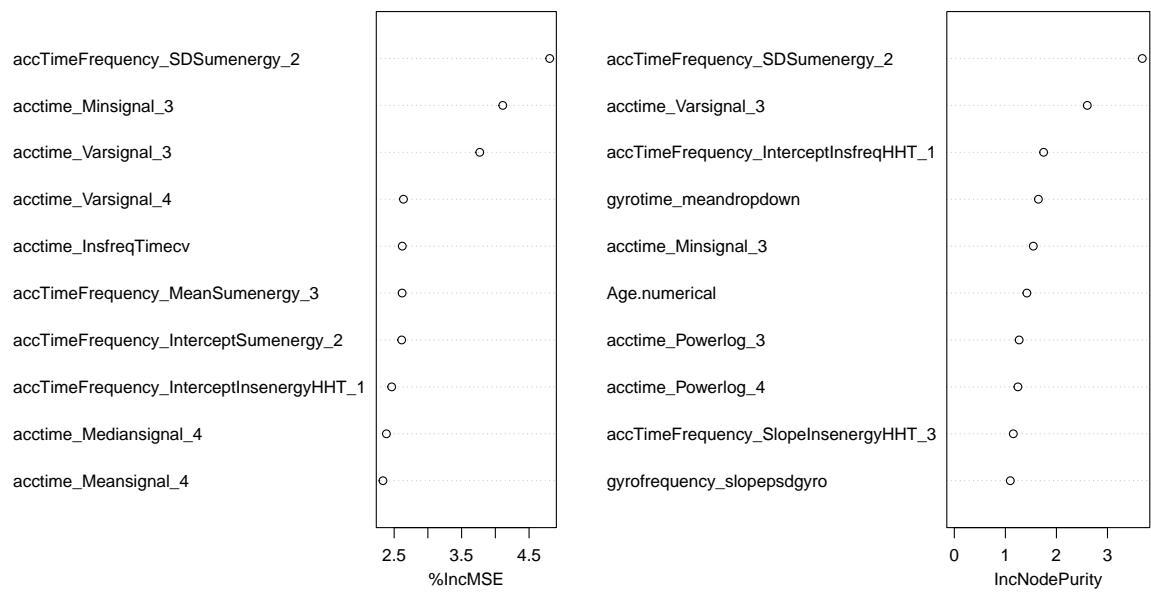


Figure 3: Top 10 Influential Variables in Nondominant Gait Time Prediction Using Random Forest

Support Vector Machines

From the grid search, $\gamma = 0.001$ and $C = 1$ produced the lowest test RMSE among all combinations evaluated. The resulting test RMSE was identical to that of random forest, the ensemble method with the lowest test error.

Principal Component Analysis

The Kaiser criterion retained 22 components, while the scree plot showed no clear elbow point but suggested roughly 2–3 components. To capture a substantial portion of the variance, 12 components were selected, which explained approximately 80% of the variance in the training data. These 12 principal components were fitted in a linear model predicting gait time. Only two components were statistically significant, and predictions were noticeably less accurate than other tested models.

Partial Least Squares Regression

Seven components were required to capture roughly 80% of the variance. The PLSR models exhibited clear overfitting, with one of the lowest training RMSE values among all methods but one of the highest test RMSE values.

Bland-Altman

For gait time, the Bland–Altman plot for one of the best-performing models (the LASSO-fitted model) is shown in Figure 4. The mean bias is slightly below zero, and all observations fall within the 95% limits of agreement. The spread of the differences increases as the average gait time increases, with points appearing more dispersed at higher values. The scale-location plot in Figure 5 similarly shows a mild upward trend in the LOESS curve (approximately 0.5 to 1), which aligns with the Bland–Altman indication of non-constant variance; however, this increase is small and not practically substantial.

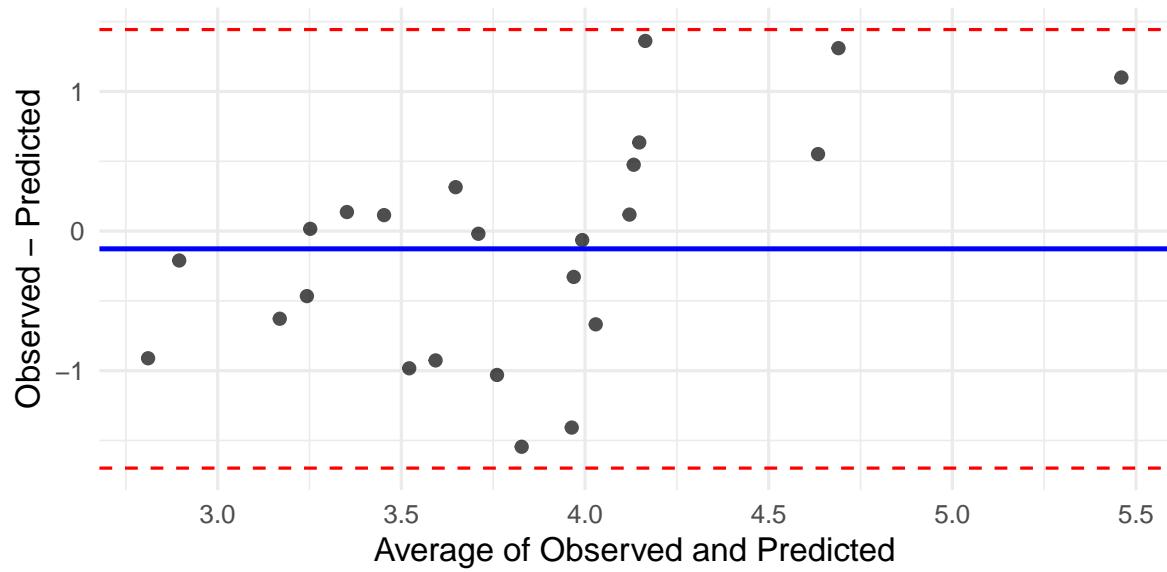


Figure 4: Bland–Altman Plot for Nondominant Foot Gait Time (LASSO-Selected Linear Model)

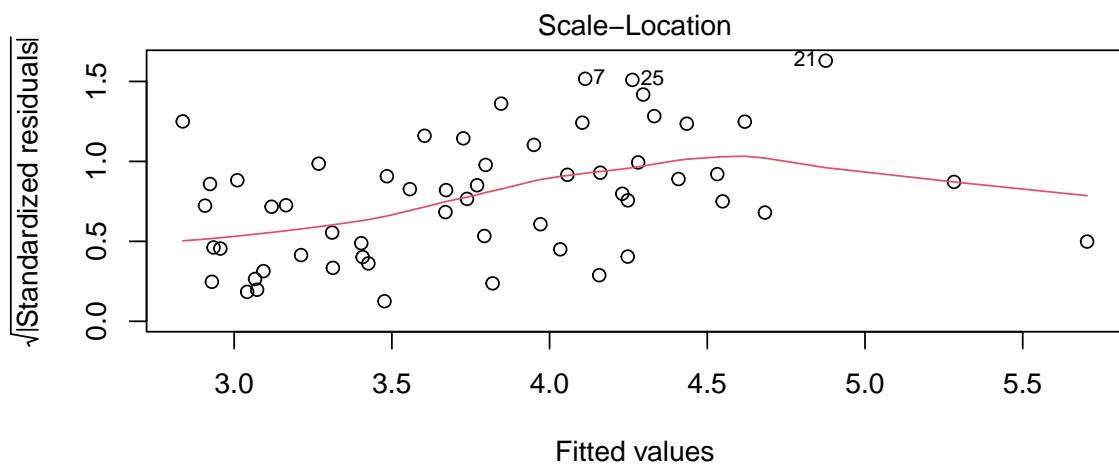


Figure 5: Scale–Location Plot for the LASSO-Selected Linear Model of Nondominant Foot Gait Time

To summarize the gait results, ridge regression achieved the lowest test RMSE, making it the best-performing predictive model. However, the lasso-selected linear model produced a very similar RMSE while using far fewer predictors, offering a substantially simpler alternative. Further interpretation of these findings is provided in the Discussion section.

TUG Results

Results for the TUG time models are presented in Table 3.

Stepwise

After correlation filtering, 61 features were retained. Both forward and mixed stepwise approaches selected identical models, with 13 features. The selected features included a mixture of accelerometer and gyroscope variables, along with age. Despite this reduction in dimensionality, the stepwise models produced the highest test RMSE among all methods.

Lasso

Lasso effectively reduced the dimensionality of the dataset, selecting 24 features at best $\lambda = 0.2707325$ out of the original 136 predictors. While the model demonstrated a reasonable fit to the training set, its test RMSE was higher than several ensemble methods. Refitting the nonzero coefficients in a standard linear model did not improve performance; the test RMSE increased relative to the lasso model. As the refitted linear model did not provide performance gains, no further examination of its coefficient estimates was pursued.

Ridge

The ridge model showed improvement in results compared to lasso, delivering solid predictive performance on the test set. However, its test RMSE remained slightly higher than those achieved by the ensemble approaches. This model was obtained at $\lambda = 23.00559$.

Elastic Net

The nondominant foot achieved its best performance at $\alpha = 0$. For consistency with the fixed-seed retraining procedure used in the gait analysis, the model was retrained with $\alpha = 0.1$, yielding similar performance. The elastic net model achieved a lower test RMSE than lasso but remained less accurate than the ensemble methods.

Table 3: Nondominant Foot TUG Time Model Summaries (Bold Indicates Lowest Test RMSE)

Model	Number of Features Selected	$R^2_{\text{Train,Adj}}$	$\text{RMSE}_{\text{Train}}$	$\text{RMSE}_{\text{Test}}$
Lasso	24	0.691	2.135	3.261
Lasso Fitted	24	0.644	1.732	3.502
Ridge	136	0.551	2.575	2.654
Elastic Net, $\alpha = 0.1$	51	0.646	2.285	2.953
Forward Selection	13	0.672	1.928	3.811
Mixed Stepwise	13	0.672	1.928	3.811
Random Forest	136	0.268	3.287	2.391
Bagging	136	0.261	3.301	2.570
Boosting	136	0.715	2.049	2.529
Support Vector Machine	136	0.99	0.380	2.700
PCR	136	0.467	2.487	2.943
PLSR	136	–	1.659	2.866

Random Forest

Random forest achieved the lowest test RMSE across all models, making it one of the strongest predictive approaches for TUG time. Across the range of m values tested, $m = 50$ produced the lowest test RMSE and was retained for interpretation. As shown in Figure 6, Age was by far the most influential predictor based on percent increase in mean squared error, followed by the minimum and variance of the z-axis accelerometer signal (`acctime_Minsignal_3`, `acctime_Varsignal_3`). Log-transformed power features from the z-axis and magnitude (`acctime_Powerlog_3`, `acctime_Powerlog_4`) also ranked highly based on percent increase in Mean Squared Error. Gyroscope-derived features were also more prominent in the TUG models compared to the gait random forest models.

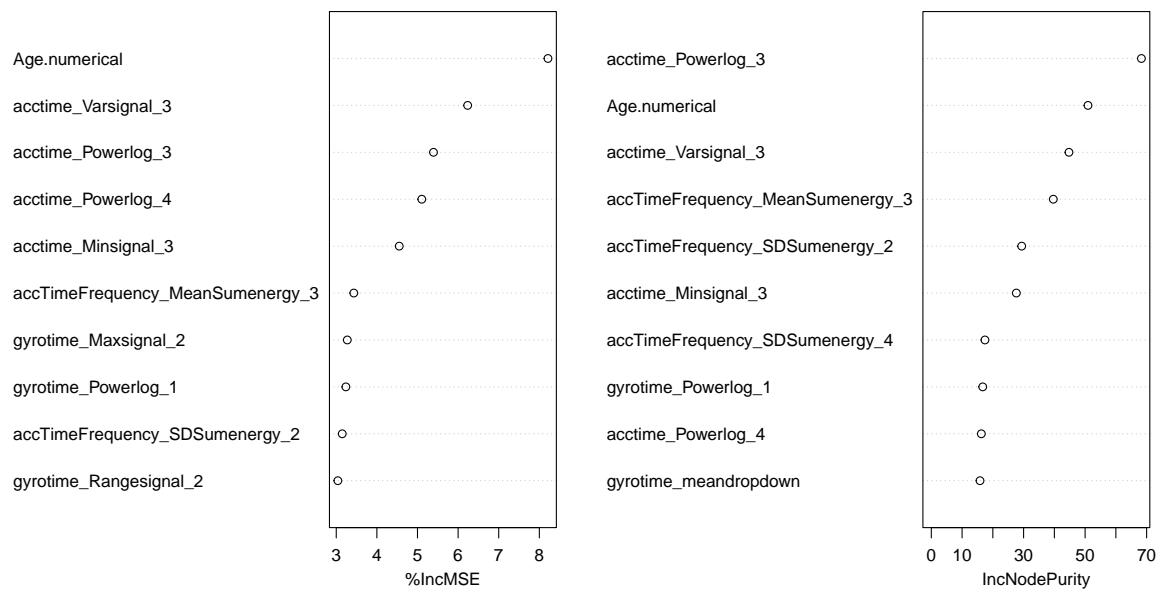


Figure 6: Top 10 Influential Variables in Nondominant TUG Time Prediction Using Random Forest

Bagging

Bagging achieved the third-lowest test RMSE just behind the other two ensemble methods involved, while still exhibiting low training variance ($R^2_{Train,Adj}$).

Boosting

Cross-validation identified 417 trees as the optimal number for this dataset. The boosting model produced test RMSE just slightly higher than random forest, ranking among the better-performing models overall.

Support Vector Machines

This model achieved its lowest test RMSE with $\gamma = 0.01$ and $C = 10$. SVM exhibited very low training RMSE, while the test RMSE was substantially higher, indicating a large discrepancy between training and test performance.

Principal Component Regression

Coincidentally, mirroring the gait analysis, the Kaiser criterion again suggested retaining 22 components, while the scree plot indicated an elbow point around 2–3 components. To capture a substantial portion of the variance, 12 components were selected to explain 80% of the total variance in the training data. These components were used in a linear model to predict TUG time, and 4 of them were statistically significant based on their p -values. The principal component regression model produced higher test RMSE than the ensemble methods.

Partial Least Square Regression

For PLSR, the optimal number of components was identified as five; this amount was required to capture approximately 80% of the variance in the training data. The PLSR model exhibited overfitting, with training RMSE substantially lower than test RMSE, consistent with the pattern observed in the gait time model.

Bland-Altman

For TUG time, the Bland–Altman plot for the best-performing model (random forest) is shown in Figure 7. The mean bias is slightly below zero, and the differences remain relatively stable across the range of average TUG values. Two observations fall outside the 95% limits of agreement, while all remaining points lie within the expected range.

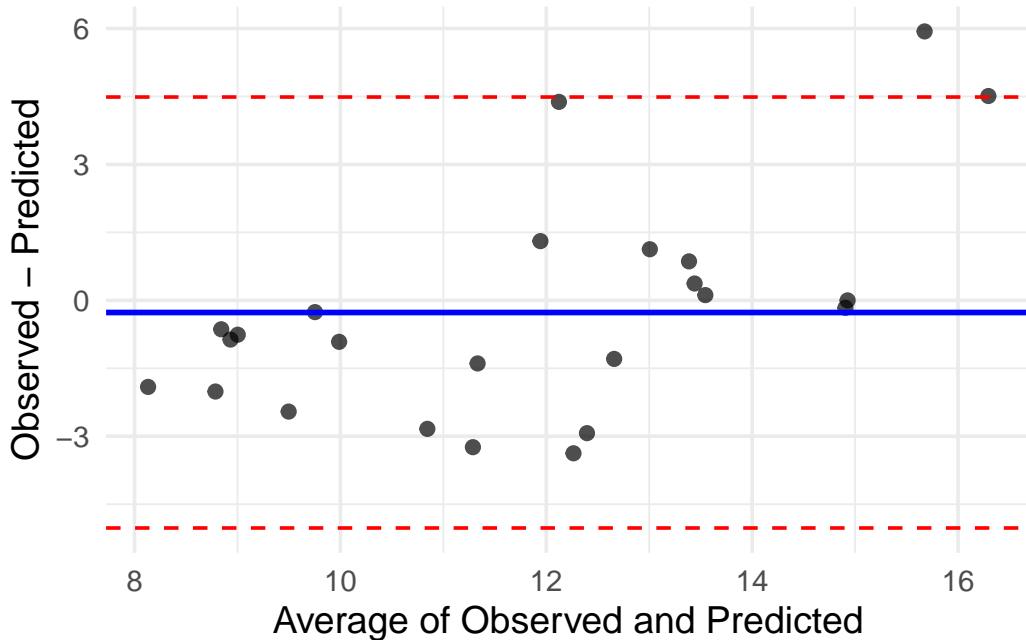


Figure 7: Bland–Altman Plot for Nondominant Foot TUG Time (Random Forest Model)

In summary of the TUG results, the ensemble methods (random forest, bagging, and boosting) achieved the lowest test RMSE values overall. Among them, random forest produced the strongest predictive performance and is considered the best-performing model for TUG time.

Discussion

Although Batula et al. did not reach a clear conclusion regarding dominant versus nondominant feet, their findings on hand movements still offer relevant insight. All participants in their study were right-handed, meaning their left-right hand comparisons effectively reflected dominant versus nondominant limb use. They found that the dominant hand showed more efficient motor activation patterns during both motor execution and motor imagery. This aligns with broader evidence that dominant limbs generally possess more practiced neural pathways and greater cortical representation (Batula et al., 2017). These mechanisms may help explain why the dominant foot in our tapping test demonstrated slightly different performance compared to the nondominant foot.

Our results were consistent with this pattern and with findings from similar foot-tapping studies (Hinman, 2019). Although not statistically significant, the dominant foot showed a slightly higher average number of taps within the fixed time interval compared to the nondominant foot.

This small performance difference may also relate to the predictive accuracy gap observed between the dominant and nondominant models. While not tested directly in this analysis, greater variability in dominant foot response patterns, such as slightly higher tap counts and additional fluctuations in other features could make the dominant foot more difficult to model, contributing to its slightly higher test RMSE.

Focusing now on the nondominant foot findings, the additional features in the mixed stepwise model slightly improved variance explained, reducing training RMSE and increasing $R^2_{Train,Adj}$. However, forward selection produced a lower test RMSE, indicating better generalization to unseen data. Overall, stepwise regression provided a useful balance between interpretability and predictive performance in this high-dimensional setting. Both variants still showed evidence of overfitting, with low training RMSE but higher test RMSE, suggesting that stepwise procedures may offer limited generalizability for predicting gait and TUG time.

For the gait model, lasso regression performed well. The nearly identical test RMSE between the original lasso model and the refitted linear model indicates that the simplified linear version preserved predictive accuracy while enabling full statistical inference. Notably, all selected features came from the accelerometer (plus age), suggesting that gait time can be predicted effectively without gyroscope-derived inputs. The very small overall model *p*-value confirms that the fitted linear model is statistically significant. Although lasso did not produce the lowest overall RMSE, the difference was small enough that the gain in model simplicity outweighed the slight loss in predictive performance. In contrast, the TUG lasso model showed weaker predictive performance. When the nonzero coefficients were refitted into a standard linear model, test RMSE increased more noticeably, indicating that the shrinkage imposed by lasso was essential for stabilizing predictions.

Ridge regression did not reduce the number of predictors and explained less variance in the training set than some alternatives, but it provided stable and accurate predictions. Ridge achieved the lowest test RMSE for the gait model, though it underperformed relative to ensemble methods for TUG. This suggests that ridge captured much of the underlying signal but may not generalize as effectively as more flexible models in certain settings.

Elastic Net closely mirrored the lasso model for gait. For TUG and gait measures, the elastic net models did not surpass ridge regression in terms of test RMSE, as simulations indicated that $\alpha = 0$ (equivalent to ridge) produced the most accurate model. Nonetheless, elastic net offered a modest improvement over lasso in predicting TUG time, highlighting its ability to balance sparsity with predictive consistency.

Across the three ensemble methods tested, performance was relatively similar, with random forest yielding the strongest generalization to unseen data. These findings reinforce that ensemble approaches consistently achieved the best predictive accuracy for TUG time compared to all other modeling techniques. Notably, most influential features from random forest originated from the accelerometer, while gyroscope-derived variables were less impactful than expected. Time-domain features also appeared more influential than frequency- or time-frequency-based

variables. This indicates that basic movement magnitude and time-based characteristics provide stronger predictive information for TUG and gait time than more complex frequency or time-frequency features. Despite strong predictive performance, the random forest model explained relatively little variance in the training set, indicating a model that generalized well but did not capture as much within-sample structure as other methods.

Grid search results showed that the gait SVM model produced a balanced fit with strong predictive performance, achieving results similar to random forest and second only to Ridge regression. In contrast, the TUG SVM model showed extreme overfitting, with near-perfect $R^2_{Train, Adj}$ and substantially high test RMSE, and would not be considered a reliable predictive model.

For dimensionality reduction approaches, PCR and PLSR achieved some of the lowest training RMSE values. PLSR in particular showed exceptionally strong training performance with less components, making this method more efficient than PCR; however, the large gap between train and test RMSE indicated substantial overfitting. PCR provided reasonable predictive performance while reducing dimensionality, but it did not surpass the stronger regularization or ensemble approaches. Overall, while PLSR produced a more compact representation of the predictors, PCR ultimately generalized more effectively on unseen data.

The Bland–Altman plot for gait time showed that the best linear model (the non-zero-coefficients fitted model) slightly overpredicted on average, though the overall bias was minimal. The spread of differences increased at higher gait times, suggesting some degree of heteroskedasticity and indicating that the model performed less consistently for participants with longer gait times. A similar pattern was suggested in the residual diagnostic plots, which showed mild increases in variance at larger fitted values.

However, this behaviour was not substantial. The Scale–Location plot did not provide strong evidence of heteroskedasticity. Although the LOESS curve showed slight curvature and a mild increase in residual spread across fitted values, the change was small and unlikely to affect model validity. Overall, the assumption of constant variance was considered acceptable.

In contrast, the Bland–Altman plot for TUG time using the random forest model showed minimal overprediction, with relatively stable differences across the range of averages. Although two observations fell outside the limits of agreement, the overall error structure appeared more uniform than in the gait time model.

Conclusion

For gait time, the nondominant foot lasso-selected linear model provided the best balance between interpretability and predictive performance. It achieved a test RMSE of approximately 0.795 seconds, indicating strong accuracy while using only 5 predictors, all accelerometer-derived features plus age, compared to the 136 required by ridge regression. Although ridge achieved the lowest test RMSE overall, its marginal improvement in accuracy did not outweigh

the major gains in simplicity and interpretability provided by the lasso model. Consequently, the lasso-based linear model is more practical and better suited for clinical or applied settings.

For TUG time, the nondominant foot random forest model delivered the strongest predictive performance on unseen data, with a test RMSE of roughly 2.391 seconds. While this value appears larger than the gait time RMSE, it is proportional to the substantially longer duration of the TUG test, making the model’s accuracy reasonable in context. Although random forest does not provide interpretable coefficients or p -values, its variable importance rankings identified the most influential features, providing meaningful insight into movement characteristics associated with TUG performance. Despite its relatively low $R^2_{Train,Adj}$, our primary objective was prediction rather than inference, making test RMSE the appropriate criterion; under this metric, random forest was clearly the best-performing model. These results indicate that accelerometer-derived features, particularly those capturing time-domain dynamics, are more informative for predicting TUG performance than gyroscope or frequency-based measures.

Together, these findings highlight the complementary value of a sparse linear model derived from lasso-selected features and an ensemble-based approach for predicting functional mobility outcomes from high-dimensional wearable-sensor data.

References

- Batula, A. M., Mark, J. A., Kim, Y. E., & Ayaz, H. (2017). Comparison of brain activation during motor imagery and motor movement using fNIRS. *Computational Intelligence and Neuroscience*, 2017(1), 5491296.
- Bobić, V., urić-Jovičić, M., Radovanović, S., Dragašević, N., Kostić, V., & Popović, M. (2018). Classification of parkinsonism based on foot tapping test. *5th International Conference on Electrical, Electronic and Computing Engineering*.
- Bohannon, R. W., & Wang, Y.-C. (2019). Four-meter gait speed: Normative values and reliability determined for adults participating in the NIH toolbox study. *Archives of Physical Medicine and Rehabilitation*, 100(3), 509–513.
- Boulesteix, A.-L., & Strimmer, K. (2007). Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, 8(1), 32–44.
- Chan, J. Y.-L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z.-W., & Chen, Y.-L. (2022). Mitigating the multicollinearity problem and its machine learning approach: A review. *Mathematics*, 10(8), 1283.
- Choi, J., Parker, S. M., Knarr, B. A., Gwon, Y., & Youn, J.-H. (2021). Wearable sensor-based prediction model of timed up and go test in older adults. *Sensors*, 21(20), 6831.
- Delsys Incorporated. (2023). In *Trigno Wireless Biofeedback System User's Guide*. <https://delsys.com/downloads/USERSGUIDE/MAN-031-1-7> Trigno Wireless Biofeedback System.pdf
- Hinman, M. (2019). Validity and reliability of a 10-second foot-tap test in older adults. *MOJ Gerontol. Geriatr.*, 4, 42–46.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in r* (Vol. 103). Springer.
- Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab – an S4 package for kernel methods in R. *Journal of Statistical Software*, 11(9), 1–20. <https://doi.org/10.18637/jss.v011.i09>
- Konstantinov, A. V., & Utkin, L. V. (2021). Interpretable machine learning with an ensemble of gradient boosting machines. *Knowledge-Based Systems*, 222, 106993.
- Kuhn, & Max. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22. <https://CRAN.R-project.org/doc/Rnews/>
- Liland, K. H., Mevik, B.-H., & Wehrens, R. (2024). Pls: Partial least squares and principal component regression. <https://CRAN.R-project.org/package=pls>
- Liu, Y., An, H., & Bian, S. (2020). Hilbert-huang transform and the application. *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)*, 534–539.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2024). e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), TU wien. <https://CRAN.R-project.org/package=e1071>

- Noroozi, S., Smith, E. J., Sagert, M. M., Jakobi, J. M., & Weyand, S. (2025). Detecting tap events from a foot-tapping test with an accelerometer: Validation of teager-kaiser energy operator and discrete wavelet transform. *Available at SSRN 5297926*.
- Podsiadlo, D., & Richardson, S. (1991). The timed “up & go”: A test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, 39(2), 142–148.
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rahmat, F., Zulkaffli, Z., Ishak, A. J., Abdul Rahman, R. Z., Stercke, S. D., Buytaert, W., Tahir, W., Ab Rahman, J., Ibrahim, S., & Ismail, M. (2024). Supervised feature selection using principal component analysis. *Knowledge and Information Systems*, 66(3), 1955–1995.
- Ridgeway, G., & Developers, G. (2024). *Gbm: Generalized boosted regression models*. <https://CRAN.R-project.org/package=gbm>
- Tay, J. K., Narasimhan, B., & Hastie, T. (2023). Elastic net regularization paths for all generalized linear models. *Journal of Statistical Software*, 106(1), 1–31. <https://doi.org/10.18637/jss.v106.i01>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. <https://CRAN.R-project.org/package=dplyr>