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 2.5 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 (1). 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 (2).

Data was collected from 85 adults without Parkinson's disease, approximately three-quarters of whom were classified as older adults ($\geq$ 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). 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.

Although many of these engineered features are highly correlated, model accuracy remains the primary priority. 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 (3). 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

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 (4). 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 (5). 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 (5). 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.

The configuration where dominant and nondominant foot features were merged into a single model did not yield improved predictive performance, with RMSE values consistently worse than the individual-foot models. As a result, analyses in this report focus solely on the models trained separately for each 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 (6). 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 (>0.8 for Gait Time, >0.75 for TUG Time), 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 (7). The optimal lambda was chosen, corresponding to the minimum cross-validated mean squared error. All predictors were automatically standardized by **glmnet** prior to fitting (7).

Ridge

Ridge regression was applied to the full set of predictors using the **glmnet** package, setting $\alpha = 0$ to implement pure Ridge regularization (7). Predictors were standardized by **glmnet** prior to fitting (7). 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 (8).

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 (7). Train 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 (9).

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 (10). The number of candidate variables considered at each split was set to the default m = p/3 where p = 136 represents the number of predictor set, and 500 trees were grown to ensure the out-of-bag (OOB) error estimate stabilized. 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 (10). 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 (11). 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 (12). 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 were implemented using the **e1071** and **kernlab** packages (13) (14). 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 (3).

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 (15). 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 gait and TUG time 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 (16). 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 (17).

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 (9). 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 bias $\pm$ 1.96 $\times$ SD(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.

Table 1: Dominant Foot Gait Time Model Summaries

Model	Number of Features Selected	$R^2_{\text{Train,Adj}}$	$\text{RMSE}_{\text{Train}}$	$\text{RMSE}_{\text{Test}}$
Lasso	5	0.446	0.802	0.882
Lasso Fitted	5	0.492	0.735	0.980
Ridge	136	0.366	0.858	0.792
Elastic Net, $\alpha = 0.3$	11	0.421	0.820	0.855
Forward Selection	8	0.618	0.619	1.269
Mixed Stepwise	8	0.618	0.619	1.269
Random Forest	136	0.206	0.960	0.967
Bagging	136	0.199	0.964	0.969
Boosting	136	0.636	0.650	0.818
Support Vector Machine	136	0.99	0.107	0.841
PCR	136	0.222	0.848	0.842
PLSR	136	–	0.429	1.230

Gait Results

Results for the gait time models are presented in Table 1 (dominant foot) and Table 2 (non-dominant foot).

Table 2: Nondominant Foot Gait Time Model Summaries

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

Stepwise

After the correlation filter, 75 features remained for the dominant foot and 68 for the nondominant foot. Forward and mixed stepwise produced highly similar models. For the dominant foot, both methods selected the same 8 predictors; for the nondominant foot, mixed stepwise selected 20 and forward selected 17. Mixed stepwise slightly reduced training RMSE and improved adjusted R-squared ($R^2_{Train,Adj}$), while forward selection had the lower test RMSE.

Lasso

Lasso performed very effectively in reducing the dimensionality of the dataset, selecting only 5 features for both feet out of the original 136 predictors at optimal $\lambda = 0.1944252$ (dominant) and $\lambda = 0.2026074$ (nondominant). 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 minimally (less than a 0.1 change for the dominant and only 0.022 for the nondominant foot). A deeper examination of the summary table for the nondominant linear model (Table 3) shows that only one feature, the standard deviation of the intertap interval from the accelerometer (acctime_ITIsd), was statistically significant along with the intercept ($\alpha = 0.05$). The overall model p-value was extremely small, indicating that the regression model as a whole was significant.

Table 3: 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_ITIsd	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 among all gait models, using $\lambda = 9.904288$ (dominant) and $\lambda = 14.97415$ (nondominant). Ridge retained all predictors and produced stable predictions across both feet.

Elastic Net

The average RMSE for all alpha values were very similar. However it was identified that 0.3 and 0 were the optimal values for dominant and nondominant. For the fixed seed retrain, I decided to use 0.1 instead of 0 for nondominant 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. Both elastic net simulations did not perform better than their respective ridge regression in terms of test set RMSE.

Random Forest

Random forest produced moderate training fit but competitive test RMSE, slightly behind Ridge for the nondominant foot. Some noticeably influential variables seen in Figure 1 (in terms of percent increase in Mean Squared Error) for dominant foot in the importance plot were coefficient of variation and standard deviation of the intertap interval features (acctime_ITIcv, acctime_ITIsd), Age , Standard deviation of summed foot-tapping energy over time in the z-axis (accTimeFrequency_SDSumenergy_3), and Kurtosis signal in the x-axis (acctime_Kurtosissignal_1). Three stand out important features for the nondominant foot (shown in Figure 2) is the minimum and variance signal in the z-axis (acctime_Minsignal_3, acctime_Varsignal_3) as well as standard deviation of summed foot-tapping energy over time in the y-axis (accTimeFrequency_SDSumenergy_2). All top ten features (in terms of %IncMSE) in Figure 1 and Figure 2 were extracted from the accelerometer.
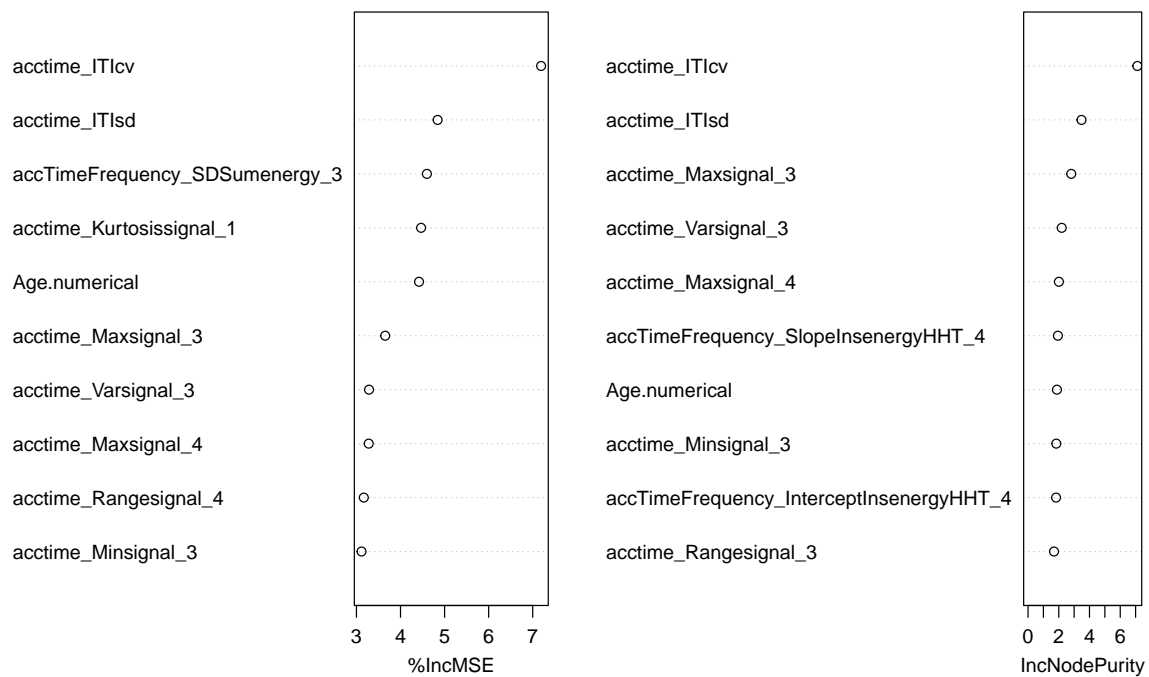
Figure 1: Top 10 Influential Variables in Dominant Gait Time Prediction Using Random Forest
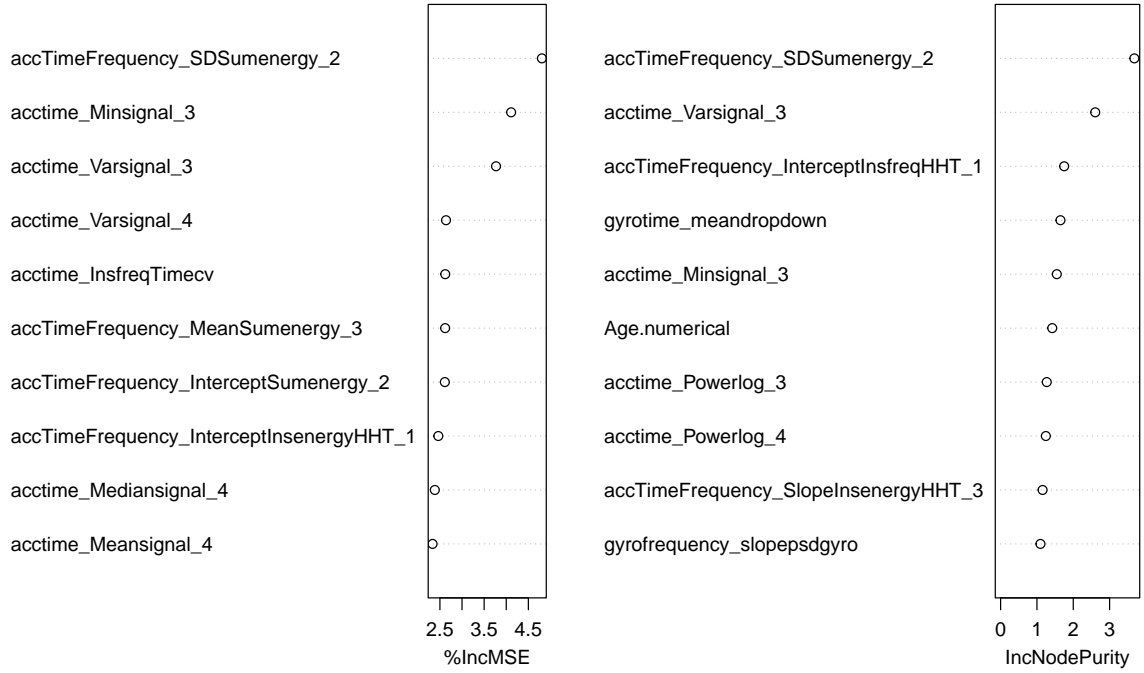
Figure 2: Top 10 Influential Variables in Nondominant Gait Time Prediction Using Random Forest

Bagging

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

Boosting

Cross-validation selected 338 trees for the dominant foot and 118 for the nondominant foot, which was then used for predictions. Boosting achieved the lowest test error among most ensemble methods, outperforming both random forest and Bagging for the dominant dataset.

Support Vector Machines

For the dominant foot, $\gamma = 0.01$ and C = 10 produced the best performance. For the nondominant foot, $\gamma = 0.001$ and C = 1 produced the lowest test RMSE of all possible combinations from the grid search.

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, selected 12 components, 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, although the overall model provided one of the better predictive performances for the dominant foot, while predictions for the nondominant foot were less accurate.

Partial Least Squares Regression

Seven components were required to capture roughly 80% variance. PLSR models showed a large difference between training and test RMSE.

Bland-Altman

For gait time, the Bland–Altman plot for one of the best-performing models (the nondominant LASSO-fitted model) is shown in Figure 3. 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 4 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.
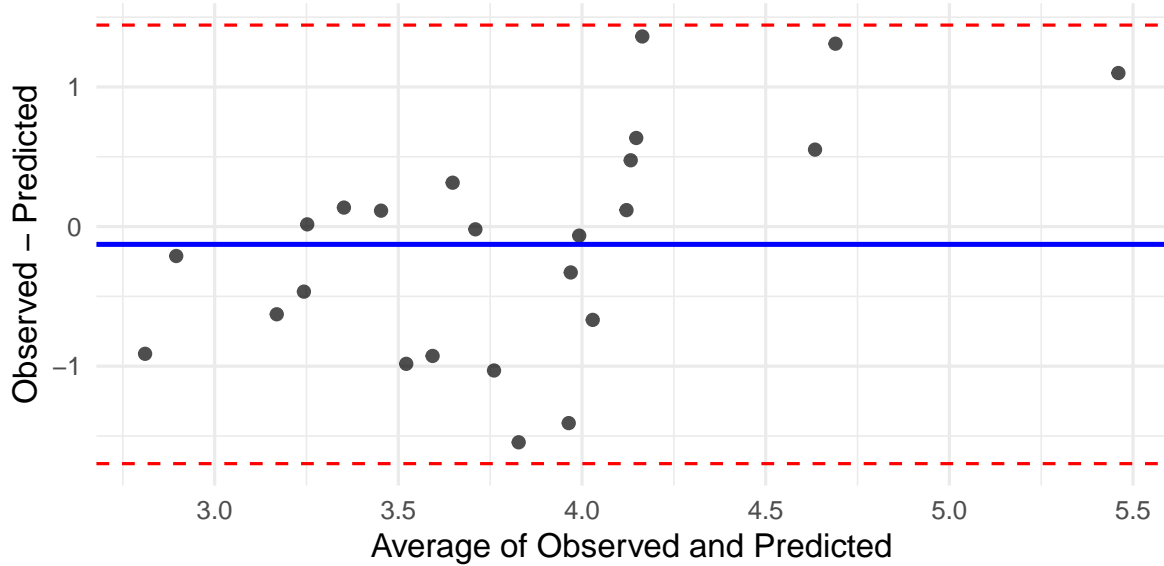

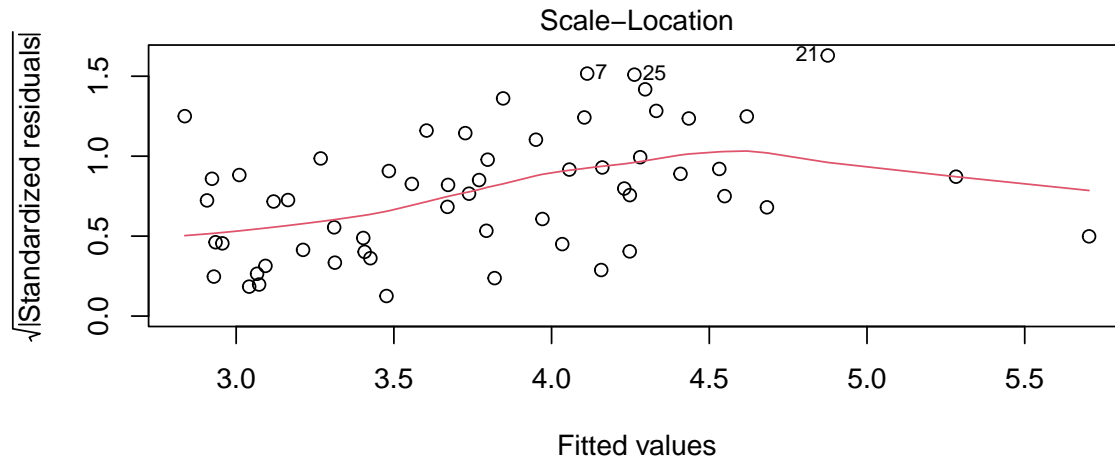Figure 3: Bland–Altman Plot for Nondominant Foot Gait Time (LASSO-Selected Linear Model)

Figure 4: Scale–Location Plot for the LASSO-Selected Linear Model of Nondominant Foot Gait Time

TUG Results

Results for the TUG time models are presented in Table 4 (dominant foot) and Table 5 (non-dominant foot).

Table 4: Dominant Foot TUG Time Model Summaries

Model	Number of Features Selected	$R^2_{\text{Train,Adj}}$	$\text{RMSE}_{\text{Train}}$	$\text{RMSE}_{\text{Test}}$
Lasso	19	0.622	2.536	3.151
Lasso Fitted	19	0.597	2.157	4.215
Ridge	136	0.504	2.904	3.242
Elastic Net, $\alpha = 0.1$	39	0.61	2.575	3.142
Forward Selection	12	0.658	2.154	4.459
Mixed Stepwise	11	0.656	2.183	4.395
Random Forest	136	0.163	3.774	2.729
Bagging	136	0.153	3.796	2.793
Boosting	136	0.786	1.909	3.064
Support Vector Machine	136	0.348	3.331	3.382
PCR	136	0.386	2.885	4.044
PLSR	136	–	1.664	4.294

Table 5: Nondominant Foot TUG Time Model Summaries

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	0.259	3.306	2.500
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

Stepwise

After correlation filtering, 61 features were retained for dominant and 57 for nondominant foot. Both forward and mixed stepwise approaches produced nearly identical models, with the forward method selecting 12 features, just one more than mixed for dominant feet. While the two stepwise methods obtained the same 13 features for nondominant.

Lasso

Lasso effectively reduced the dimensionality of the dataset, selecting 19 features for the dominant foot at $\lambda = 0.340021$ and 24 features at $\lambda = 0.2707325$ for the nondominant foot out of the original 136 predictors. While the model demonstrated reasonable predictive performance, its test RMSE was higher than that of several ensemble methods, suggesting somewhat lower generalization capability. When the nonzero coefficients from Lasso were refitted into a standard linear model, the test RMSE for dominant foot increased substantially, indicating that the shrinkage imposed by Lasso played a critical role in maintaining predictive stability.

Ridge

The Ridge model showed comparable results to Lasso, delivering solid predictive performance across both training and test sets, specifically for the nondominant dataset. However, its test RMSE remained slightly higher than those achieved by the ensemble approaches. These models were obtained at $\lambda = 21.8568$ for dominant and $\lambda = 23.00559$ for nondominant.

Elastic Net

The dominant foot achieved its best performance at $\alpha = 0.1$, while the nondominant foot performed best at $\alpha = 0$, though all configurations for the dominant foot produced test RMSE values within 0.1 of each other. Once again, for the fixed seed retrain, I decided to use 0.1 instead of 0 for better representation. Both elastic net models outperformed lasso, but were not as accurate as the ensemble methods.

Random Forest

Random forest achieved the lowest test RMSE across all models, making it one of the strongest predictive approaches for TUG Time. Presented in Figure 5 and Figure 6, the most influential features for the dominant foot (based on percent increase in Mean Squared Error) were age, the mean intertap interval (acctime_ITImean), and the minimum and variance of the accelerometer's z-axis signal (acctime_Minsignal_3, acctime_Varsignal_3). For the nondominant foot, Age and the minimum and variance of the z-axis signal were again among the top predictors, along with the log-transformed power in the z-axis and magnitude (acctime_Powerlog_3, acctime_Powerlog_4). Gyroscope-derived features were also more prominent in the TUG models compared to the gait random forest models.
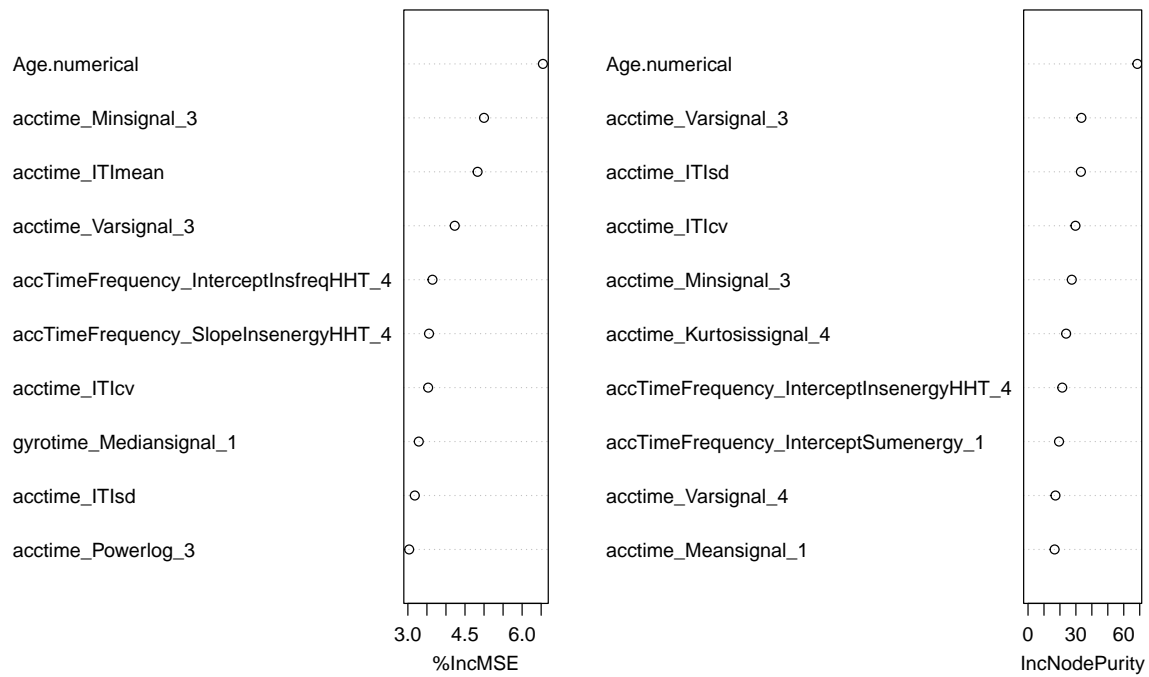
Figure 5: Top 10 Influential Variables in Dominant TUG Time Prediction Using Random Forest
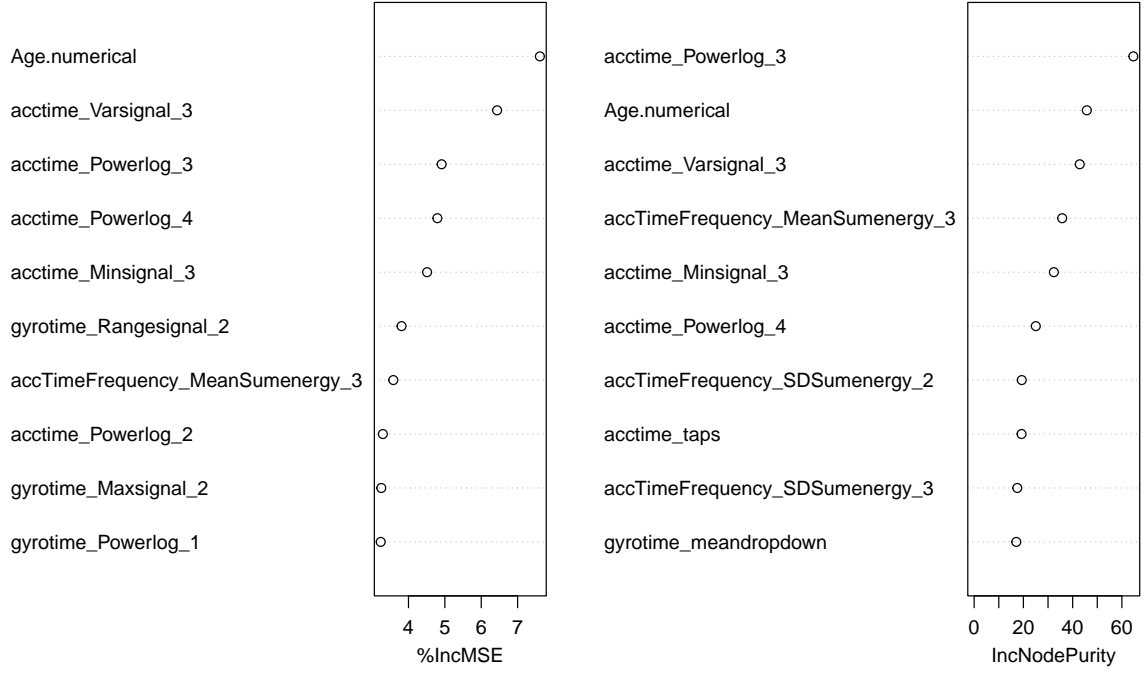
Figure 6: Top 10 Influential Variables in Nondominant TUG Time Prediction Using Random Forest

Bagging

Bagging achieved the second-lowest test RMSE behind random forest, while still exhibiting low training variance $(R^2_{Train,Adj})$.

Boosting

Cross-validation identified the optimal number of trees as 660 for the dominant foot and 417 for the nondominant foot. Boosting produced predictive results slightly behind the other ensemble methods but still among the top-performing models overall.

Support Vector Machines

The dominant foot achieved its best performance with $\gamma = 0.001$ and C = 1, while the nondominant foot performed best with $\gamma = 0.01$ and C = 10. However, the nondominant model displayed classic overfitting. In contrast, the dominant foot model demonstrated a more balanced generalization.

Principal Component Regression

The Kaiser criterion suggested retaining 22–23 components for the dominant and nondominant feet, while the scree plot indicated an elbow point around 2–3 components for both. To capture a substantial portion of the variance, 12 components were selected to explain 80% of the total variance in the training data. The 12 components were then used in a linear model predicting TUG Time, and 4–5 of them were found to be statistically significant based on their p-values.

Partial Least Square Regression

For PLSR, the optimal number of components was identified as six for the dominant foot and five for the nondominant foot; this amount was required to capture approximately 80% of the variance in the training data. Similar to the gait time models, the PLSR models for TUG exhibited classic overfitting, with substantially lower training RMSE compared to the test RMSE.

Bland-Altman

For TUG time, the Bland–Altman plot for the best-performing model (the nondominant random forest model) 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.
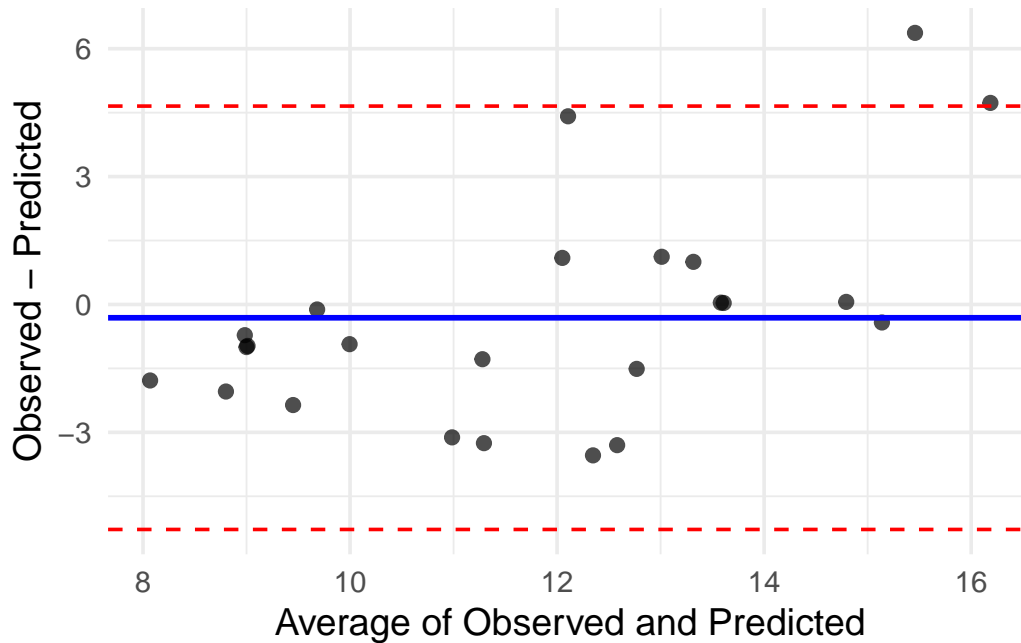

Figure 7: Bland–Altman Plot for Nondominant Foot TUG Time (Random Forest Model)

Discussion

For further interpretation, we will mainly focus on the nondominant foot results because they achieved lower test RMSE and therefore provided stronger predictive performance.

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 (18). 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 (19). Although the difference was not significant, the dominant foot showed a slightly higher average number of taps within the fixed time interval compared to the nondominant foot.

Although not tested in this analysis, this subtle difference in foot-tapping performance may help explain why the dominant-foot models did not achieve the same predictive accuracy as the nondominant-foot models. Greater variability in dominant-foot response patterns, such as slightly higher tap counts and additional fluctuations in other features can make prediction more difficult, contributing to the dominant-foot models' 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 offered a balance between interpretability and predictive accuracy in this high-dimensional setting. Both approaches showed signs of overfitting, achieving low training RMSE but noticeably higher test RMSE, suggesting that stepwise procedures, while useful for variable reduction, may have limited generalizability for predicting gait and TUG time.

For the gait model, Lasso regression performed well: the nearly identical test RMSE between the Lasso model and the refitted linear model suggests that the simplified linear version preserved predictive accuracy while drastically reducing the number of features. 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 results of the Lasso model for gait. For TUG and gait, the nondominant Elastic Net models did not outperform Ridge in terms of test RMSE, though Elastic Net offered a modest improvement over Lasso for TUG time. Both approaches provided a balance between sparsity and 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 for the nondominant foot showed that the gait random forest model produced a balanced fit with strong predictive performance, second only to Ridge regression. In contrast, the TUG random forest 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, PLSR offered a compact representation of the data but generalized less effectively than PCR.

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 displayed slight curvature, the variance of the residuals increased slightly across fitted values, but the change was small

and not large enough to suggest problematic heteroskedasticity. Overall, the assumption of constant variance was considered acceptable.

In contrast, the Bland–Altman plot for TUG time using the nondominant 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 Lasso-selected linear model provided the best balance between interpretability and predictive performance. This model achieved an average prediction error of 0.795 seconds, indicating strong accuracy while using only 5 predictors compared to the 136 required by Ridge regression. Although Ridge achieved the lowest test RMSE, the small improvement in accuracy did not outweigh the substantial gain in model simplicity and interpretability offered by Lasso. Consequently, the Lasso-based linear model is more practical and well-suited for clinical or applied settings.

For TUG time, the nondominant-foot random forest model delivered the strongest predictive performance on unseen data, with an average prediction error of approximately 2.5 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 measures identified the most influential features, offering meaningful insight into the movement characteristics associated with TUG performance.

Together, these findings illustrate the complementary strengths of sparse linear models and ensemble methods for predicting functional mobility outcomes from high-dimensional wearable-sensor data.

References

1. Bohannon RW, Wang YC. Four-meter gait speed: Normative values and reliability determined for adults participating in the NIH toolbox study. Archives of physical medicine and rehabilitation. 2019;100(3):509–13.

2. Podsiadlo D, Richardson S. The timed "up & go": A test of basic functional mobility for frail elderly persons. Journal of the American geriatrics Society. 1991;39(2):142–8.

3. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: With applications in r. Vol. 103. Springer; 2013.

4. Wickham H, François R, Henry L, Müller K, Vaughan D. Dplyr: A grammar of data manipulation [Internet]. 2023. Available from: https://CRAN.R-project.org/package=dplyr

5. Kuhn, Max. Building predictive models in r using the caret package. Journal of Statistical Software [Internet]. 2008;28(5):1–26. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v028i05

6. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available from: https://www.R-project.org/

7. Tay JK, Narasimhan B, Hastie T. Elastic net regularization paths for all generalized linear models. Journal of Statistical Software. 2023;106(1):1–31.

8. Chan JYL, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong ZW, et al. Mitigating the multicollinearity problem and its machine learning approach: A review. Mathematics. 2022;10(8):1283.

9. Choi J, Parker SM, Knarr BA, Gwon Y, Youn JH. Wearable sensor-based prediction model of timed up and go test in older adults. Sensors. 2021;21(20):6831.

10. Liaw A, Wiener M. Classification and regression by randomForest. R News [Internet]. 2002;2(3):18–22. Available from: https://CRAN.R-project.org/doc/Rnews/

11. Ridgeway G, Developers G. Gbm: Generalized boosted regression models [Internet]. 2024. Available from: https://CRAN.R-project.org/package=gbm

12. Konstantinov AV, Utkin LV. Interpretable machine learning with an ensemble of gradient boosting machines. Knowledge-Based Systems. 2021;222:106993.

13. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), TU wien [Internet]. 2024. Available from: https://CRAN.R-project.org/package=e1071

14. Karatzoglou A, Smola A, Hornik K, Zeileis A. Kernlab – an S4 package for kernel methods in R. Journal of Statistical Software. 2004;11(9):1–20.

15. Rahmat F, Zulkafli Z, Ishak AJ, Abdul Rahman RZ, Stercke SD, Buytaert W, et al. Supervised feature selection using principal component analysis. Knowledge and Information Systems. 2024;66(3):1955–95.

16. Liland KH, Mevik BH, Wehrens R. Pls: Partial least squares and principal component regression [Internet]. 2024. Available from: https://CRAN.R-project.org/package=pls

17. Boulesteix AL, Strimmer K. Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. Briefings in bioinformatics. 2007;8(1):32–44.

18. Batula AM, Mark JA, Kim YE, Ayaz H. Comparison of brain activation during motor imagery and motor movement using fNIRS. Computational intelligence and neuroscience. 2017;2017(1):5491296.

19. Hinman M. Validity and reliability of a 10-second foot-tap test in older adults. MOJ Gerontol Geriatr. 2019;4:42–6.