Prediction of Gold Standard Gait Data from Inertial Data: A Machine Learning Approach

William Fröhlich[1], **Diego Pinheiro**[2], **Sandro Rigo**[3], **Rafael Baptista**[1], **César Marcon**[1]

[1] Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) – Porto Alegre – RS – Brazil

[2]Universidade Feevale – Novo Hamburgo – RS – Brazil

[3]Universidade do Vale do Rio dos Sinos (UNISINOS) – São Leopoldo – RS – Brazil

william_r_f@hotmail.com, diegopinheiro@feevale.br, rigo@unisinos.br

rafael.baptista@pucrs.br, cesar.marcon@pucrs.br

Abstract. *Biomechanics laboratories typically use kinematic cameras and force platforms as the gold standard for gait assessment. However, these systems are expensive and have limited availability. Wearable devices, equipped with sensors—primarily inertial sensors—can capture movement data, enabling the inference of human gait behavior. To enhance the quality of measurements obtained from wearables, this study investigates the feasibility of predicting kinematic parameters from inertial data collected by wearable sensors. Machine learning techniques, including Random Forest, XGBoost, and Gradient Boosting, were used to correlate inertial measurements with data from traditional motion capture systems. Feature importance analysis and SHAP highlighted the significance of velocity and acceleration in predicting kinematic parameters. Experimental results indicate that tree-based models, particularly Gradient Boosting and XGBoost, achieved the best performance, with coefficient of determination values close to 0.989, demonstrating the feasibility of the proposed approach.*

Resumo. *Laboratórios de biomecânica normalmente utilizam câmeras cinemáticas e plataformas de força como padrão-ouro na avaliação da marcha. Contudo, tais sistemas são custosos e têm disponibilidade limitada. Dispositivos vestíveis podem conter sensores, principalmente inerciais, que capturam movimentos permitindo inferir o comportamento humano em uma marcha. Com o objetivo aumentar a qualidade das medições obtidas por vestíveis, este estudo investiga a viabilidade de prever parâmetros cinemáticos a partir de dados inerciais coletados por sensores vestíveis. O estudo utiliza técnicas de aprendizado de máquina, incluindo Random Forest, XGBoost e Gradient Boosting, que correlacionam medições inerciais com dados obtidos por sistemas tradicionais de captura de movimento. A análise de importância de características e SHAP destacou a relevância da velocidade e aceleração na predição dos parâmetros cinemáticos. Os resultados experimentais indicam que modelos baseados em árvores, especialmente Gradient Boosting e XGBoost, apresentaram os melhores desempenhos, com coeficientes de determinação próximos a 0,989, mostrando a viabilidade da abordagem proposta.*

1. Introduction

Biomechanical gait assessments are essential for identifying locomotion issues, enabling personalized rehabilitation, and enhancing athletic performance [Benson et al. 2018] [Akhtaruzzaman et al. 2016]. Traditional motion capture systems utilizing kinematic cameras and plantar pressure measurements are considered the gold standard due to their precision and ability to capture detailed biomechanical data [Zhang et al. 2017] [Jakob et al. 2021]. Inertial Measurement Units (IMUs) have emerged as a promising alternative, offering portability, cost-effectiveness, and versatility for gait analysis outside laboratory settings [Akhtaruzzaman et al. 2016] [Kotiadis et al. 2010], in contrast to biomechanics laboratories, which are associated with high costs, the need for specialized equipment, and limitations to controlled laboratory environments [Benson et al. 2018].

While traditional motion capture systems provide high precision in measuring joint angles, stride length, and speed, wearable sensors offer the advantage of continuous real-time monitoring, despite their lower accuracy [Kotiadis et al. 2010]. The primary challenge lies in establishing a direct relationship between inertial and kinematic data [Silva and Stergiou 2020]. Our study employs data processing and correlation analysis to address this issue and identify significant relationships between the datasets. Based on these correlations, we developed a comprehensive analysis using various machine learning algorithms, including Linear Regression, Random Forest, XGBoost, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Gradient Boosting, to predict kinematic parameters from inertial data.

Our approach incorporates advanced techniques such as hyperparameter optimization, cross-validation, and feature importance analysis (including Feature Importance and SHAP values). We aim to establish model accuracy and understand which aspects of inertial data are most relevant for predicting kinematic parameters. Through this comprehensive methodology, we strive to develop wearable IMUs as reliable tools for gait analysis.

2. Related Work

Understanding human biomechanics is crucial for optimizing health outcomes, enhancing athletic performance, and accelerating recovery processes [Silva and Stergiou 2020, Benson et al. 2018, Akhtaruzzaman et al. 2016]. In this context, wearable devices for gait analysis have gained prominence due to their portability and practicality, enabling studies beyond traditional laboratory settings [Benson et al. 2018].

While biomechanics laboratories rely on sophisticated equipment such as high-speed cameras, force platforms, and electromyographs to capture precise movement data during gait [Akhtaruzzaman et al. 2016], wearable sensors have emerged as a more versatile alternative, albeit with some limitations in precision and accuracy [Akhtaruzzaman et al. 2016]. These devices, primarily composed of strategically placed inertial sensors on the body, allow continuous motion monitoring in various environments [Kotiadis et al. 2010].

Current scientific literature reveals a significant gap in the correlation between data obtained from inertial sensors and those derived from laboratory-based optical camera systems [Tsakanikas et al. 2023, Silva and Stergiou 2020]. Existing research has predominantly focused on specific applications, such as Parkinson's disease diagnosis, rather than direct comparative analyses between the two methodologies [da Rosa Tavares et al. 2023].

Investigating the correlation between datasets is essential to validate the accuracy and reliability of inertial sensors in replicating kinematic measurements traditionally obtained in laboratory settings [Desai et al. 2024, He et al. 2024, Kvist et al. 2024, Ripic et al. 2023, Rousanoglou et al. 2024]. This study examines the distinctions between laboratory-based motion capture systems and wearable sensors; it seeks to understand how inertial data correlates with kinematic parameters and how it can contribute to precise biomechanical analyses.

3. Methodology

Figure 1 presents the proposed methodology for analyzing and training gait kinematic data obtained from a motion capture system (gold standard) and inertial data collected from wearable sensors. The goal is to develop a model capable of accurately relating inertial and kinematic data using a dataset of inertial measurements highly correlated with kinematic points. The first step involves data collection (Subsection 3.1), conducted in a biomechanics laboratory equipped with high-speed cameras and a wearable IMU system, both from BTS Bioengineering. Next, data preprocessing (Subsection 3.2) was performed to make the data comparable and apply necessary filters.

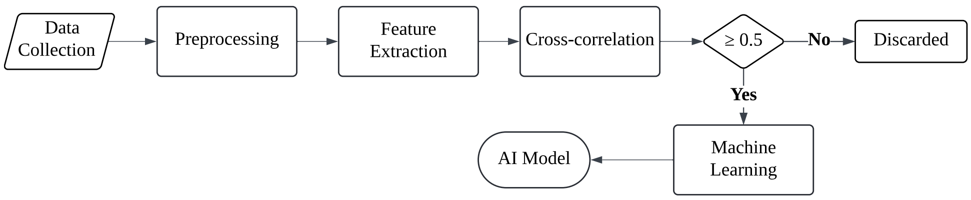

Figure 1. Flowchart of the proposed methodology for analyzing kinematic and inertial data.

After preprocessing, new features were extracted from the inertial data (Subsection 3.3) using mathematical and physical techniques, generating information such as velocity, angular acceleration, magnitude, jerk (a derivative of acceleration), and position from the raw data. Subsequently, cross-correlation was applied to temporally align the data and identify the highest correlations (Subsection 3.4). This analysis identified the time lag between the kinematic point and the inertial data with the highest correlation. Finally, in Subsection 3.5, the data grouped by highest correlation was used to train AI models, varying the algorithms, to evaluate which performs best with the lowest error.

3.1. Data Collection

The data collection experiments were conducted following a standardized gait analysis protocol after approval of the ethics committee. Participants performed a linear course that included walking in a straight line, stepping on a force platform, and returning to the starting point. Data was collected at the GaitLab biomechanics laboratory [BTS Bioengineering 2024b], using a motion capture camera system and force platform. For inertial data acquisition, we employed the GWalk wearable sensor [BTS Bioengineering 2024a], positioned in the participants' lumbar region. This device records accelerometer and gyroscope data, including acceleration (acc) and rotational movement ($gyro$) in three axes, as well as roll ($roll$), pitch ($pitch$), and yaw (yaw) orientation angles. Figure 2 shows the positioning of wearable IMUs on participants during collection, demonstrating the orientation of X, Y, and Z axes and rotation directions. Raw data were extracted through device-specific software.

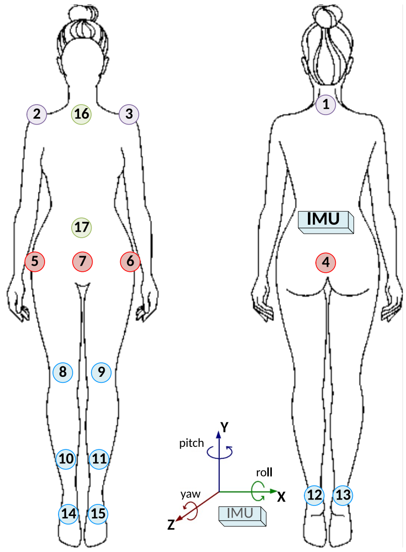

Figure 2. Positioning of kinematic points and wearable IMUs for data collection experiments.

Among the various kinematic points available in the biomechanics laboratory, we selected the most relevant ones based on literature [Delval et al. 2021]. These points were organized into anatomical groups, as illustrated in Figure 2, according to Upper Trunk - Cervical vertebra C7 ($c7$ - 1), Right shoulder (r_should - 2), Left shoulder (l_should - 3); Lower Trunk - Sacrum ($sacrum\_s$ - 4), Right anterior superior iliac spine (r_asis - 5), Left anterior superior iliac spine (l_asis - 6), Midpoint between iliac crests ($MIDASIS$ - 7); for Lower Limbs - Right and left knees (r_knee_1 - 8, l_knee_1 - 9), Right and left ankles (r_mall - 10, l_mall - 11), Right and left heels (r_heel - 12, l_heel - 13), Right and left metatarsals (r_met - 14, l_met - 15); and Calculated Points - Average shoulder position (PO - 16), and Center of mass (SHO - 17)

In the collected kinematic data, as illustrated in Figure 2, the X-axis corresponds to lateral movement, the Y-axis to vertical movement, and the Z-axis to gait progression. Complementary to the inertial data, the force platforms provide force measurements for both feet (r_force and l_force) in all three axes. The complete dataset[1] is publicly available for experiment reproduction.

3.2. Data Preprocessing

In gait analysis, abrupt changes in X, Y, and Z axes derived from acceleration data represent significant variations in forces acting on the body, potentially indicating specific events or gait irregularities.

[1] www.kaggle.com/dataset/wrfrohlich/artemis-dataset

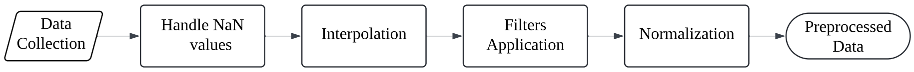

Figure 3. Flowchart for evaluating similarities and correlation between kinematic and inertial data.

Data processing and analysis were conducted using Python 3.10, employing NumPy, Pandas, Scipy, and Scikit-Learn libraries. Data preprocessing followed established methodologies in the literature [Millecamps et al. 2015, Parashar et al. 2023], proving crucial for ensuring data quality and consistency for advanced analyses.

Initial treatment focused on missing data (NaN), which can arise for various reasons and potentially distort results. We removed NaNs from the beginning and end of files, corresponding to periods outside recorded movement. For technical failures, we performed imputation using the mean of neighboring values, while for data not detected by force platforms, we substituted zeros, indicating absence of contact. We also applied linear interpolation to fill additional gaps, maintaining temporal continuity of gait data.

We implemented a low-pass Butterworth filter for noise reduction, which proved more effective in preserving crucial data patterns. We applied a 3.0 Hz cutoff frequency with a 5th-order filter for GaitLab (250 Hz) and GWalk (100 Hz) data. This configuration allowed retention of essential low-frequency components critical for gait analysis.

Data normalization was performed using Min-Max scaling, which proved superior to standardization by better preserving relative importance between features, especially in a dataset with different scales. Finally, we performed data fusion to temporally align different equipment sampling rates based on timestamps from each system to ensure synchronized analysis of gait cycles captured by various sensors.

3.3. Features Extraction

We implemented a feature extraction process from the inertial data to ensure that all kinematic data exhibit at least a 0.5 correlation. This process is fundamental in gait analysis for establishing significant correlations between inertial and kinematic data. Inertial sensors, comprising accelerometers and gyroscopes, capture forces and rotations acting on the body during movement.

The extracted features include velocities (vel) in X, Y, and Z axes, obtained through temporal integration of acceleration data in each axis. Angular acceleration (ang_acc_gyro) in all three axes was calculated from the derivative of gyroscope data, which measures rotation rate in each direction. Acceleration magnitude (mag_acc) was determined by the square root of the sum of squares of accelerations in all three axes, representing the total intensity of the resultant force on the body.

Similarly, we calculated angular velocity magnitude (mag_gyro) from gyroscope-recorded rotations. Jerk ($jerk$), obtained from the derivative of acceleration in X, Y, and Z axes, quantifies the acceleration change rate over time, revealing sudden changes in acting forces. Finally, position (pos) in each axis was determined through temporal integration of velocity data, allowing estimation of spatial location of body segments and approximating a kinematic measure derived from velocity.

3.4. Cross-Correlation Analysis

After preprocessing and feature extraction, we applied a cross-correlation technique to quantify relationships between kinematic points and inertial data. This technique is particularly suitable for time series analysis, as it evaluates similarity between two signals considering different time lags. This analysis identified patterns of highest correlation, and the data showed no significant relationships in their temporal behavior.

Cross-correlation serves two essential functions in our analysis: besides quantifying the degree of correlation between signals, it determines the optimal temporal synchronization between data. The method systematically shifts one signal relative to another and calculates correlation for each time lag, thus identifying the displacement that maximizes correlation between signals. Although data were initially synchronized during collection based on their timestamps, cross-correlation provided additional refinement of this alignment, precisely determining the optimal time lag between inertial and kinematic signals.

Cross-correlation analysis was systematically applied to all possible kinematic points and inertial data combinations. We established a 0.5 correlation threshold as a criterion for selecting the most relevant data pairs, thus ensuring that only statistically significant correlations were considered in subsequent analyses. This threshold was defined based on previous literature studies indicating that correlations above 0.5 represent moderate to strong associations between biomechanical variables.

3.5. Machine Learning Model Development

The core development of this work focuses on applying machine learning algorithms based on methods from previous stages. We aim to evaluate how well inertial data correlate with kinematic points and predict their behavior. We began this phase by implementing ML modeling using regression models to predict gold-standard data from inertial data.

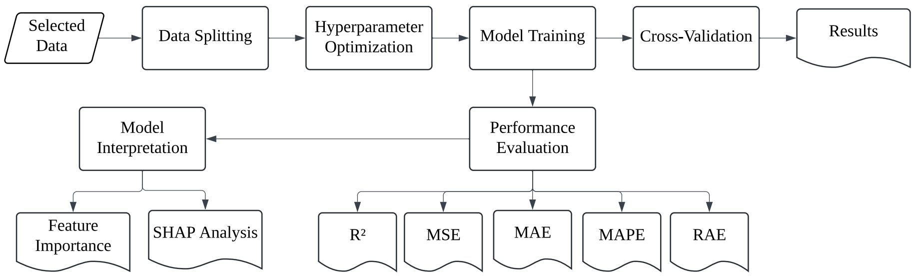

Figure 4. Flowchart for artificial intelligence model training stages.

Experiments were conducted using six algorithms: linear regression, Random Forest, XGBoost, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Gradient Boosting. Each algorithm was used for all subsequent steps. Data was split into training (80%) and test (20%) sets for model performance evaluation.

We employed Randomized Search with cross-validation to optimize model performance and identify the best-performing algorithm (k=3). This approach efficiently explored each algorithm's hyperparameter space, identifying combinations that maximize performance. Linear Regression required no hyperparameter tuning due to its fundamental nature. For Random Forest, we adjusted the number of trees ($n\_estimators$: 100-500), maximum tree depth (max_depth: None, 10, 20, 30, 50), minimum samples for node splitting ($min\_samples\_split$: 2-20), minimum samples per leaf ($min\_samples\_leaf$: 1-10), and bootstrap sampling options ($bootstrap$).

For XGBoost, we tuned the number of trees ($n\_estimators$: 100-500), maximum depth (max_depth: 3-10), learning rate ($learning\_rate$: 0.01, 0.05, 0.1, 0.2), sample proportion per tree ($subsample$: 0.8, 0.9, 1.0), and feature proportion per tree ($colsample\_bytree$: 0.8, 0.9, 1.0). The MLP algorithm was optimized by adjusting hidden layer neurons ($hidden\_layer\_sizes$: 50-200), activation functions ($activation$: "relu"/"tanh"), L2 regularization term ($alpha$: 0.0001-0.01), and learning rate strategy ($learning\_rate$: "constant"/"adaptive").

For SVM, we configured the regularization parameter (C: 0.1-100), kernel coefficient ($gamma$: "scale", "auto", 0.01-1), and kernel type ($kernel$: "rbf", "linear", "poly"). Gradient Boosting parameters included number of trees ($n\_estimators$: 100-500), maximum depth (max_depth: 3-10), learning rate ($learning\_rate$: 0.01-0.2), and sample proportion ($subsample$: 0.8-1.0).

Model evaluation employed k-fold cross-validation (k=5) to assess generalization and prevent overfitting. Performance metrics included R-squared (Coefficient of Determination), Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Relative Absolute Error (RAE). Additionally, we conducted residual analysis to identify prediction patterns or outliers.

Furthermore, we utilized SHapley Additive exPlanations (SHAP) analysis to interpret model predictions, generating feature importance and dependency plots as additional evaluation metrics.

4. Results and Discussion

This section presents and discusses the results obtained through the experiments performed. Initially, as detailed in Subsection 4.1, the hyperparameters that demonstrated the best performance for each algorithm are presented. Next, the results related to the errors (Subsection 4.2) obtained by each algorithm are presented. Subsequently, we evaluate the results regarding the importance of variables (Subsection 4.3) and, finally, we discuss the results of the SHAP analysis (Subsection 4.4).

4.1. Hyperparameters

Regarding hyperparameters, we found that the best-performing models were Gradient Boosting and XGBoost, with average scores close to 0.989. Random Forest also performed well, with an average score of 0.959. On the other hand, SVM and MLP showed inferior results, with average scores of 0.835 and 0.682, respectively.

We observed the best performance for Gradient Boosting, with an average test score of 0.989 using a learning rate of 0.05, a maximum depth of 4, 291 estimators, and a subsampling rate of 0.8. Thus, we can conclude that slightly smaller trees, a moderate number of estimators, and a reduced learning rate effectively prevented overfitting and generalization.

For XGBoost, we achieved an average test score of 0.989 using a learning rate of 0.1, a maximum depth of 5, 269 estimators, a $colsample\_bytree$ of 0.8, and a subsampling rate of 1.0. Like Gradient Boosting, we found that moderate-sized trees, a reasonable number of estimators, and a learning rate of 0.1 strike a good balance between bias and variance. Gradient Boosting and XGBoost produced better results using a lower learning rate and smaller maximum depth.

The Random Forest model showed an average test score of 0.959 using a maximum depth of 10, a relatively small number of estimators (154), and $bootstrap$ enabled. It also used a $min_samples_leaf$ of 3 and $min_samples_split$ of 13. Based on smaller trees, $bootstrap$ and careful tuning of leaf and split constraints benefit this dataset. The greater diversity among the trees generated by the bootstrap benefited the final result.

The model generated by the SVM algorithm with the best performance used an RBF kernel with a gamma of 'scale' and a large C value of 100, achieving an average test score of 0.8346. This suggests that a non-linear kernel with appropriate scaling and a high penalty for misclassification is essential for good performance on this data type. As for the MLP models, even with hyperparameter tuning, they achieved lower average test scores compared to the other algorithms (0.6817), obtained with a constant learning rate, hidden layer sizes of (100, 50), an alpha of 0.0001, and the $ReLU$ activation function.

4.2. Error Analysis

Regarding the error analysis for the algorithms, Random Forest and Gradient Boosting show the best results in terms of error (MSE, MAE) and coefficient of determination (R^2). These models have MSE values close to zero and R^2 values close to 1, indicating an almost perfect fit to the data. XGBoost also stands out, with very similar metrics, suggesting it is a robust approach for this data, as observed in the hyperparameter analysis.

MLP (Multi-Layer Perceptron) and SVM (Support Vector Machine) perform worse compared to tree-based models, with higher MSE values and lower R^2, especially for some variables, making further experiments unnecessary. Linear Regression showed the worst performance, with significantly higher MSE values and lower R^2, indicating that the data may not be linearly separable or that there are nonlinear relationships.

Gradient Boosting achieved very good performance, reaching low MSE values and relatively high R^2. The minimal performance difference between GBM and XGBoost suggests that either would be an excellent choice. XGBoost achieved very good error values, though it is more computationally demanding. Random Forest also showed strong performance, with high R^2 values, mostly above 0.99 for position variables, and low MSE values indicating its effectiveness in modeling complex relationships.

Variables with the Best Performance were $c7\_z$, r_should_z, l_should_z, $sacrum\_s\_z$, and r_asis_z, which have the lowest errors and highest R^2 values across all models, with MSE in the order of 10^{-7} to 10^{-6}, suggesting these variables are easier to predict. On the other hand, l_force_x, r_met_x, and l_met_y have the highest errors and lowest R^2 values.

4.3. Feature Importance

The feature importance analysis is relevant to identifying which inertial points influence the model most and which could introduce noise or bias. All three main models prioritize velocity for positions. Random Forest tends to have the highest dependence on velocity features, often with a single dominant velocity component. It is followed by Gradient Boosting and then XGBoost.

XGBoost utilizes a broader range of inertial features, particularly acceleration, gyroscope, and IMU-derived position, suggesting it can capture additional nuances in gait data. While the overall trends are similar, there are differences in specific feature importance for certain anatomical landmarks among the three models. Random Forest relies more on vel_v for l_asis_y, whereas XGBoost uses less vel_v compared to Random Forest and more vel_x and pos_x, in addition to significantly increasing pos_z and acc_z.

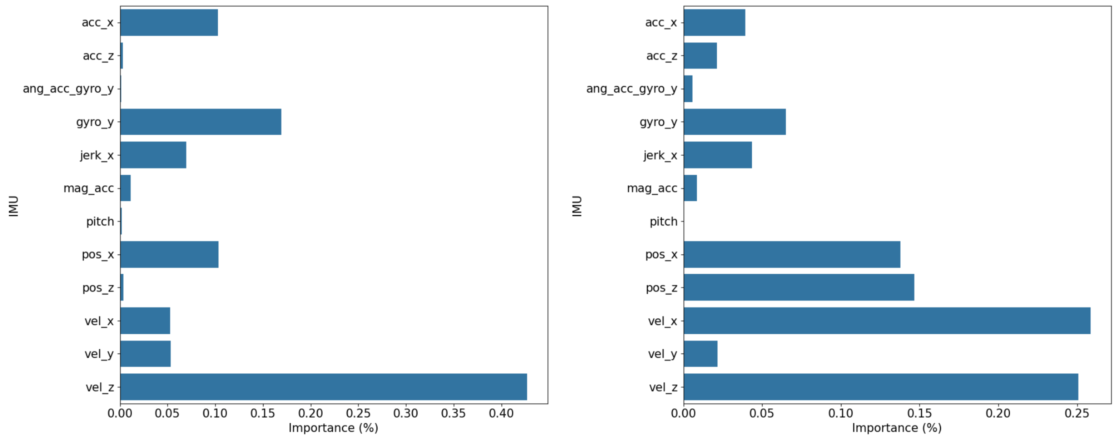

Figure 5. Feature Importance in l_asis_y for the machine learning models. a) Random Forest model. b) XGBoost model

The feature importance from XGBoost highlights its focus on a broad set of features and its tendency to find clear patterns in some cases, such as $sacrum\_s\_z$. All three models heavily rely on velocity for predicting the position of anatomical landmarks, emphasizing the fundamental relationship between movement and position. The relatively lower importance of explicit orientation features ($pitch$, $roll$, yaw) for position prediction suggests that their influence is indirectly captured through velocity and acceleration.

4.4. SHAP

The SHAP method provides an in-depth understanding of each feature's impact on machine learning models, allowing for the evaluation of how each variable contributes to prediction outcomes. This interpretability is crucial for identifying the features that most influence model performance.

Analyzing the SHAP plot results, velocity (vel_x, vel_y, vel_z) is the dominant factor in kinematic predictions, with vel_x and vel_z standing out. High values of these variables (represented by red dots in the plots) tend to increase predictions, while lower values (blue dots) decrease them. Additionally, some variables, such as r_mall_x and $sacrum\_s\_z$, strongly depend on velocities for their predictions.

Specifically, vel_x and vel_z exhibit positive and negative impacts depending on their magnitudes, whereas vel_y has less relevance in the model. Regarding sensors that capture rotational movements, variables like $gyro\_y$ and yaw stand out, directly affecting sensors such as r_should_y and r_asis_y. Accelerations (acc_x, acc_z) also prove essential in specific sensors like $c7\_y$ and r_should_y, though their relevance varies depending on the model. For example, in XGBoost, acc_x and acc_z have greater importance than Random Forest.

When comparing algorithms, XGBoost assigns more weight to accelerations (acc_x, acc_z) and angles (yaw, $pitch$), while Random Forest places more value on magnitudes (mag_acc) and velocities (vel_x, vel_z). In XGBoost, the most relevant features were vel_x and acc_x, which had high SHAP values. In contrast, vel_z and $gyro\_y$ dominated in Random Forest. For Gradient Boosting, velocities vel_x and vel_z had the most significant impact on predictions.

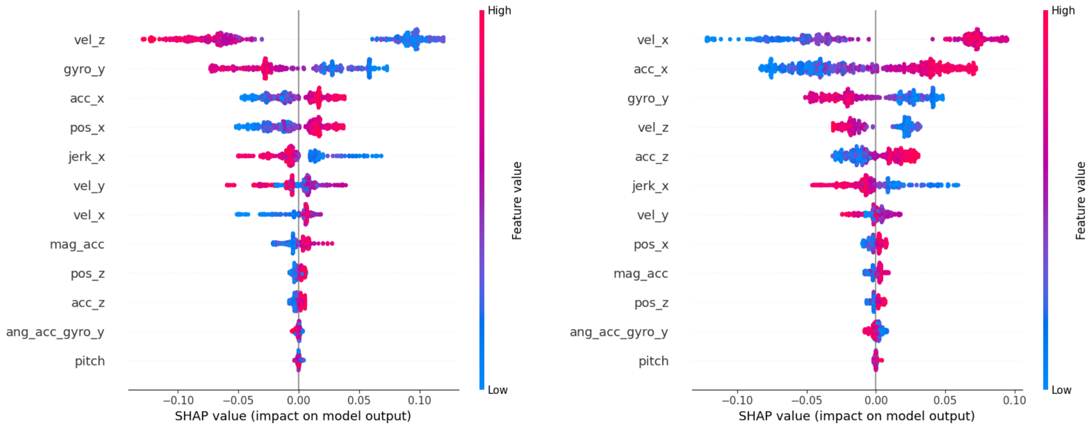

Figure 6. SHAP summary plots for l_asis_y in the machine learning models. a) Random Forest model. b) XGBoost model

Analyzing the SHAP summary plots for l_asis_y reveals distinct patterns between the models. In XGBoost, vel_x shows a mixed impact, with high values increasing predictions and low values decreasing them. In contrast, Random Forest displays a dominant positive impact from vel_z. In both models, $gyro\_y$ shows a mixed impact, indicating that its influence depends on the specific context. The variable acc_x stands out in XGBoost and Gradient Boosting but holds less relevance in Random Forest.

Overall, vel_x and vel_z emerge as the most consistent and influential features across all analyzed algorithms, positively impacting the predictions of l_asis_y. These results highlight the importance of velocity components in kinematic modeling using inertial data.

5. Conclusion

The biomechanical analysis of gait is essential for identifying locomotion issues. Two main methods are used for this purpose: biomechanics laboratories, considered the gold standard, and wearable sensors. This study aims to evaluate how to approximate the results from wearable sensors to those obtained in biomechanics laboratories. The primary challenge lies in establishing a direct relationship between inertial and kinematic data. Our study employs data processing and correlation analysis to identify significant relationships between the dataset. Based on these correlations, we developed a comprehensive analysis using several machine learning algorithms, including Linear Regression, Random Forest, XGBoost, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Gradient Boosting, to predict kinematic parameters from inertial data.

Our approach incorporates advanced techniques such as hyperparameter optimization, cross-validation, Feature Importance analysis, and SHAP values, aiming not only to establish model accuracy but also to understand which aspects of inertial data are most relevant for predicting kinematic parameters. The results demonstrate the superiority of tree-based models, with Gradient Boosting and XGBoost achieving the best scores, around 0.989, based on error metrics like MSE and R^2. Random Forest also showed competitive performance, scoring 0.959, while SVM and MLP yielded lower results.

The hyperparameter analysis revealed that Gradient Boosting and XGBoost benefited from moderate learning rates, controlled maximum tree depth, and an appropriate number of estimators, avoiding overfitting and promoting generalization. On the other hand, Random Forest performed well with smaller trees, bootstrap usage, and precise tuning of leaf and split parameters. The error analysis confirmed the effectiveness of tree-based models, especially for variables such as $c7\_z$, r_should_z, l_should_z, $sacrum\_s\_z$, and r_asis_z.

The feature importance analysis highlighted the significance of velocity variables for predicting the positions of anatomical landmarks across all three main models. XGBoost demonstrated greater sensitivity to a broader set of inertial features, including acceleration, gyroscope, and IMU-derived position. The SHAP analysis corroborated these findings, revealing the predominant influence of velocities vel_x and vel_z on predictions, with variable impacts from other features, such as accelerations and angles, depending on the algorithm.

The comparison between Random Forest and XGBoost, for instance, showed differences in the relative importance of acc_x, $gyro\_y$, and vel_z. In summary, this study demonstrates the potential of machine learning algorithms, particularly Gradient Boosting and XGBoost, for modeling kinematics from inertial data. Future work could explore incorporating additional features, such as data from more sensors or contextual information, to enhance prediction accuracy and robustness further.

References

Akhtaruzzaman, M., Shafie, A., and Khan, M. (2016). Gait analysis: Systems, technologies, and importance. *Journal of Mechanics in Medicine and Biology*, 16(07):1630003.

Benson, L., Clermont, C., Bošnjak, E., and Ferber, R. (2018). The use of wearable devices for walking and running gait analysis outside of the lab: A systematic review. *Gait and Posture*, 63:124–138.

BTS Bioengineering (2024a). BTS G-WALK - Wireless Sensor for Gait Analysis. https://www.btsbioengineering.com/products/g-walk/. Accessed: 2024-08-31.

BTS Bioengineering (2024b). BTS GAITLAB - 3D Motion Analysis System. https://www.btsbioengineering.com/products/bts-gaitlab/. Accessed: 2024-08-31.

da Rosa Tavares, J., Ullrich, M., Roth, N., Kluge, F., Eskofier, B., Gaßner, H., Klucken, J., Gladow, T., Marxreiter, F., da Costa, C., da Rosa Righi, R., and Victória Barbosa, J. (2023). utug: An unsupervised timed up and go test for parkinson's disease. *Biomedical Signal Processing and Control*, 81:104394.

Delval, A., Betrouni, N., Tard, C., Devos, D., Dujardin, K., Defebvre, L., Labidi, J., and Moreau, C. (2021). Do kinematic gait parameters help to discriminate between fallers and non-fallers with parkinson's disease? *Clinical Neurophysiology*, 132(2):536–541.

Desai, R., Martelli, D., Alomar, J., Agrawal, S., Quinn, L., and Bishop, L. (2024). Validity and reliability of inertial measurement units for gait assessment within a post stroke population. *Topics in Stroke Rehabilitation*, 31(3):235–243. PMID: 37545107.

He, Y., Chen, Y., Tang, L., Chen, J., Tang, J., Yang, X., Su, S., Zhao, C., and Xiao, N. (2024). Accuracy validation of a wearable imu-based gait analysis in healthy female. *BMC Sports Science, Medicine and Rehabilitation*, 16(1):2.

Jakob, V., Küderle, A., Kluge, F., Klucken, J., Eskofier, B., Winkler, J., Winterholler, M., and Gassner, H. (2021). Validation of a sensor-based gait analysis system with a gold-standard motion capture system in patients with parkinson's disease. *Sensors*, 21(22).

Kotiadis, D., Hermens, H., and Veltink, P. (2010). Inertial gait phase detection for control of a drop foot stimulator: Inertial sensing for gait phase detection. *Medical Engineering and Physics*, 32(4):287–297.

Kvist, A., Tinmark, F., Bezuidenhout, L., Reimeringer, M., Conradsson, D., and Franzén, E. (2024). Validation of algorithms for calculating spatiotemporal gait parameters during continuous turning using lumbar and foot mounted inertial measurement units. *Journal of Biomechanics*, 162:111907.

Millecamps, A., Lowry, K., Brach, J., Perera, S., Redfern, M., and Sejdić, E. (2015). Understanding the effects of pre-processing on extracted signal features from gait accelerometry signals. *Computers in Biology and Medicine*, 62:164–174.

Parashar, A., Parashar, A., Ding, W., Shabaz, M., and Rida, I. (2023). Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches. *Pattern Recognition Letters*, 172:65–73.

Ripic, Z., Nienhuis, M., Signorile, J., Best, T., Jacobs, K., and Eltoukhy, M. (2023). A comparison of three-dimensional kinematics between markerless and marker-based motion capture in overground gait. *Journal of Biomechanics*, 159:111793.

Rousanoglou, E., Foskolou, A., Emmanouil, A., and Boudolos, K. (2024). Inertial sensing of the abdominal wall kinematics during diaphragmatic breathing in head standing. *Biomechanics*, 4(1):63–83.

Silva, L. and Stergiou, N. (2020). Chapter 7 - the basics of gait analysis. In Stergiou, N., editor, *Biomechanics and Gait Analysis*, pages 225–250. Academic Press.

Tsakanikas, V., Ntanis, A., Rigas, G., Androutsos, C., Boucharas, D., Tachos, N., Skaramagkas, V., Chatzaki, C., Kefalopoulou, Z., Tsiknakis, M., and Fotiadis, D. (2023). Evaluating gait impairment in parkinson's disease from instrumented insole and imu sensor data. *Sensors*, 23(8).

Zhang, Y., Wang, M., Awrejcewicz, J., Fekete, G., Ren, F., and Gu, Y. (2017). Using gold-standard gait analysis methods to assess experience effects on lower-limb mechanics during moderate high-heeled jogging and running. *Journal of visualized experiments : JoVE*, 127:55714.