

Article

Prediction of Peripheral Artery Disease Treatment Decisions and Outcomes Using Machine Learning and Gait-Derived Digital Biomarkers

Ali Al Ramini ¹, Farahnaz Fallahtafti ², Mohammad Ali Takallou ³, Iraklis I. Pipinos ^{4,5}, Sara Myers ^{2,5} and Fadi Alsaleem ^{3,*}

¹ Mechanical Engineering Department, University of Nebraska-Lincoln, Lincoln, NE 68588, USA; ali.hazem.ramini@gmail.com

² Department of Biomechanics, University of Nebraska at Omaha, Omaha, NE 68182, USA; ffallahtafti@unomaha.edu (F.F.); samyers@unomaha.edu (S.M.)

³ Architectural Engineering Department, University of Nebraska–Lincoln, Omaha, NE 68182, USA; mtakallou@unomaha.edu

⁴ Department of Surgery, Division of Vascular Surgery, University of Nebraska Medical Center, Omaha, NE 68198, USA; ipipinos@unmc.edu

⁵ Department of Surgery and VA Research Service, VA Nebraska–Western Iowa Health Care System, Omaha, NE 68105, USA

* Correspondence: falsaleem2@unl.edu

Abstract

Background: Peripheral artery disease (PAD) is clinically heterogeneous, complicating identification, treatment selection, and prediction of therapeutic benefit. Gait biomechanics provide high-resolution, objective digital biomarkers that can be used in routine care. **Methods:** We built a computational framework for three tasks: (1) distinguish controls versus PAD; (2) predict clinician-selected treatment (nonoperative vs. revascularization); and (3) forecast post-treatment ground-reaction force (GRF) from baseline GRF features. The dataset included 55 unique patients with PAD and 42 controls. Standard classifiers and regressors were evaluated, and performance was measured by accuracy, balanced accuracy, Matthew’s correlation coefficient (MCC), and mean absolute error. **Results:** A single GRF feature (propulsive peak) achieved 90.9% accuracy, 86.7% balanced accuracy, and 0.868 MCC for distinguishing PAD from controls and predicting treatment, outperforming patient-reported outcomes and kinematics. For outcome forecasting, a Random Forest model predicted changes in the propulsive peak with 80% accuracy and low error. **Conclusions:** GRF-based models accurately identify PAD, anticipate this clinical team’s treatment selection, and forecast functional recovery within this single-center cohort; generalization to other practice patterns and settings remains untested. The approach relies on minimal features and modest computation and is readily generalized to other medical conditions that affect gait. Because it is compatible with wearable sensors, it is well suited to real-time, remote assessment and can advance personalized, data-driven decision support in vascular care.

Keywords: Gait; PAD; digital biomarkers; peripheral artery disease



Academic Editor: Silvestro Roatta

Received: 30 June 2026

Revised: 16 July 2026

Accepted: 22 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Peripheral artery disease (PAD) affects more than 8 million individuals in the United States and remains a leading cause of mobility impairment, disability, and cardiovascular

morbidity [1–3]. The disease is driven by progressive atherosclerotic narrowing of the arteries supplying the lower limbs, leading to a heterogeneous clinical presentation and a complex landscape of therapeutic options, including pharmacologic therapy [4], supervised exercise [5], ankle-foot orthoses (AFOs) [6,7], endovascular interventions, and open surgical revascularization [8].

Despite decades of clinical research, there remains a critical lack of reliable, individualized tools to guide treatment selection in PAD. One of the central challenges is the marked variability in treatment outcomes [4], coupled with a limited understanding of the factors that predict success or failure [1,9]. AFOs, for example, show potential to enhance walking performance, but evidence on long-term efficacy and adherence is sparse [10]. Pharmacologic approaches offer limited benefit in reversing underlying muscle pathology or improving gait function [11,12]. Surgical interventions, though more definitive, carry considerable risk and yield variable durability, with no clear superiority between endovascular and open procedures as demonstrated in comparative trials [13–16].

An additional challenge in PAD treatment is that there are no standard, agreed-upon measures to confirm improvement after treatment. Existing outcome measures vary, and the most common measures are not inclusive, as they only assess blood flow improvements or are very subjective, such as patient questionnaires. These limitations, when combined with PAD's complex nature, have limited physicians' ability to understand which course of treatment for a given patient with PAD would lead to the best outcomes [17].

This gap underscores the need for new methods to assist with selection of treatment and predict therapeutic benefit. Traditional PAD metrics, such as the ankle-brachial index and patient-reported outcomes, offer only partial insight and fail to capture the degree of walking impairment and the full biomechanical consequences of impaired perfusion and muscular dysfunction. In contrast, gait analysis is emerging as a powerful, non-invasive modality for functional assessment, with growing recognition of gait as “the sixth vital sign.” Gait dynamics not only reflect the severity of PAD but also offer promising prognostic value in predicting how patients will respond to specific interventions. Leveraging these insights may enable a more personalized, data-driven approach to PAD care.

Building on this rationale and the recent literature in using machine learning for PAD research [18], we propose a data-driven framework that leverages gait biomechanics and machine learning to enhance PAD treatment decision-making. Predictive models offer the potential to support clinicians by identifying optimal interventions, tailoring care to individual patients, and avoiding unnecessary or high-risk procedures, especially when the predicted probability of benefit is low [19]. To this end, we present a proof of concept using a comprehensive dataset that includes clinical assessments and gait biomechanics from both healthy individuals (evaluated once) and patients with PAD evaluated before and after treatment. The PAD cohort comprised two subgroups: (i) PAD patients with intermittent claudication selected to undergo revascularization operations and (ii) patients selected for nonoperative management who elected to participate in a research trial evaluating treatment with an AFO. In contemporary clinical practice, revascularization is generally reserved for patients with lifestyle-limiting symptoms that persist despite optimized conservative therapy, provided that anatomic and technical factors favor procedural success and the patient's operative risk is acceptable. Treatment assignment was non-randomized and reflected standard clinical care, emerging from shared decision-making that weighed symptom burden, expected durability, and perioperative risk. Accordingly, revascularization constituted a more invasive, higher-risk intervention, whereas AFO therapy represented a conservative approach. Both groups received guideline-directed medical therapy, including lipid-lowering and antiplatelet agents and counseling for smoking cessation.

Our previous work has demonstrated that machine learning models trained on gait data can accurately classify individuals with PAD (Figure 1a) [20,21]. Expanding on these findings, the current study explores whether the same biomechanical features can be used to predict treatment type selection and to forecast post-treatment functional outcomes (Figure 1b).

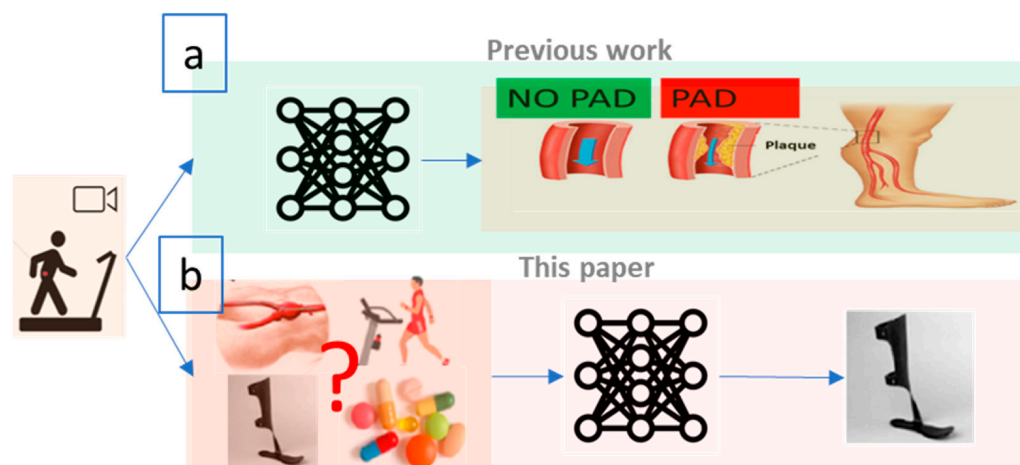


Figure 1. (a) Our previous work showed that biomechanics data could be used to classify individuals as having or not having PAD [20,21]. (b) This paper developed models to predict the treatment a physician ended up using to treat a patient with PAD based on their gait biomechanics signature before treatment.

2. Methods

2.1. Data Sources and Description

Biomechanics [22] data for this study were sourced from research approved by the Institutional Review Boards at the University of Nebraska Medical Center and the Nebraska-Western Iowa Veterans Affairs Medical Center, involving 55 unique individuals with PAD (35 undergoing revascularization, 20 treated with AFO) and 42 healthy controls. Because some AFO patients were evaluated on multiple visits prior to treatment, the pre-intervention PAD dataset comprises 65 records (35 pre-surgery, 30 pre-AFO) rather than 55; Table 1 summarizes this cohort accounting in full. Not all patient data were collected after the intervention; 30 revascularization patients' data were collected after revascularization, and eight AFO-treated patients' data were collected after three months of treatment.

Table 1. Cohort accounting used consistently throughout Methods and Results.

Quantity	Value	Notes
Unique subjects, all groups	97	42 healthy, 55 PAD
Unique patients with PAD	55	35 revascularization, 20 AFO
Pre-intervention PAD records	65	35 pre-surgery, 30 pre-AFO (repeated AFO visits)
Outcome-forecasting subset (paired pre/post)	38	30 revascularization, 8 AFO
Outcome-forecasting training set	28	23 revascularization, 5 AFO
Outcome-forecasting test set	10	7 revascularization, 3 AFO

In addition to biomechanical data, qualitative and quantitative assessments were conducted to understand the PAD condition better. The WIQ was administered to gauge the self-perceived walking ability of patients with PAD [23–25]. This questionnaire evaluates

various aspects of walking, including pain, distance, speed, and stair climbing, providing insights into patients' daily challenges with PAD. SF-36 was also employed to assess the participant's health status and quality of life [12,26]. This comprehensive questionnaire evaluates physical and mental health domains, offering a broader perspective on the impact of PAD on patients' daily lives.

Furthermore, the absolute claudication distance [24] test was conducted. This is a standardized treadmill protocol with constant speed and incrementally increasing the incline that quantifies the maximum distance a patient can walk before claudication pain necessitates cessation, providing a direct, objective measure of PAD symptom severity. Collectively, these assessments, combined with the biomechanical data, aim to offer a multi-dimensional perspective on PAD's impact on patients. Furthermore, this multi-dimensional dataset offers an opportunity to compare various measurement sources to identify the most precise metric for assessing PAD severity.

For this analysis, multiple measurement sources were considered:

1. Ankle, Hip, and Knee Kinematics: Each anatomical region was characterized by seven distinct features, capturing the nuances of movement and biomechanical alterations [27].
2. Ground Reaction Forces (GRF): Nine features were extracted to understand the forces exerted during walking, providing insights into gait alterations [22].
3. Walking Impairment Questionnaire (WIQ): Four features were derived from this self-administered questionnaire, offering a patient-centric perspective on walking ability [24].
4. Medical Outcomes Study Short Form 36 (SF-36): Seven features were considered from this tool, gauging the participants' health status and quality of life.
5. Absolute Claudication Distance (ACD): This singular feature was exclusively available for patients with PAD, measuring the distance they could walk before the claudication pain.

2.2. Data Transformation

In the preliminary stages of our analysis, we identified the need to standardize and transform our biomechanical data to ensure comparability across various biomechanical features and conditions. We employed the Yeo–Johnson transformation [28], a method optimized for varying data distributions, including zero and negative values. Initially, this transformation was determined and performed based on the pre-intervention dataset. Subsequently, the same transformation parameters derived from the pre-intervention data were applied to both the post-intervention and healthy datasets, ensuring consistency across all conditions. After the transformation, we implemented feature scaling to standardize the data distribution further, which is crucial for the performance of certain machine learning models and the validity of statistical tests that assume data uniformity. This standardization facilitates more accurate analyses and helps in achieving reliable results. This scaling was configured using the transformed pre-intervention data, ensuring that all biomechanical features ranged between -1 and 1 . The same scaling parameters were then applied to the post-intervention and healthy data. This systematic data transformation and scaling approach enhanced data clarity, separation, and comparability across all conditions and features.

2.3. Statistical Comparison with Control Group Using the Mann–Whitney U-Test

We employ the non-parametric Mann–Whitney U-test [29] to distinguish the differences in measurements between the three independent groups—healthy, AFO, and revascularization. The healthy group comprised individuals without PAD, serving as the control, while the AFO (low severity) and revascularization (high severity) groups represented patients with PAD varying in degrees of disease severity.

Prior to pairwise comparisons, we ran an omnibus Kruskal–Wallis test for each of the 21 individually tested features across the three groups (healthy, AFO, revascularization); all 21 features reached statistical significance ($p < 0.0024$, Bonferroni-corrected across features), justifying the pairwise comparisons that follow. We chose the Mann–Whitney U-test for these pairwise comparisons due to its robustness in comparing non-normally distributed data, which is often the case in medical research; many features in our data have shown non-normality. Both Bonferroni- and Benjamini–Hochberg-corrected p -values were computed per feature; we summarize these as the median Bonferroni p -value for each measurement source (the median of the per-feature Bonferroni-corrected p -values), which provides a compact indicator of typical corrected significance strength within that source, with lower values indicating stronger and more consistent significance across the source’s constituent features [30]. It is worth noting that the ACD data were only available for patients with PAD. Thus, comparisons involving the healthy group did not apply to this measure.

This statistical approach aims to identify the measurement source that best differentiates between the three groups. Ideally, the most informative measurement source would be the one that exhibits significant differences across all three pairwise comparisons: healthy vs. AFO, healthy vs. revascularization, and AFO vs. revascularization. For instance, if the GRF features consistently show significant disparities among these three comparisons, it would suggest that GRF features are the most potent identifiers of PAD severity.

2.4. Machine Learning Approach to Evaluate Features’ Ability to Quantify PAD Severity

In this section, we take a more granular approach to PAD treatment classification by focusing on individual features within each data source—GRF, WIQ, and SF-36. Each feature is modeled and tested separately. This allows us to isolate the predictive power of each feature, providing a more subtle understanding of its role in PAD treatment classification (Figure 2).

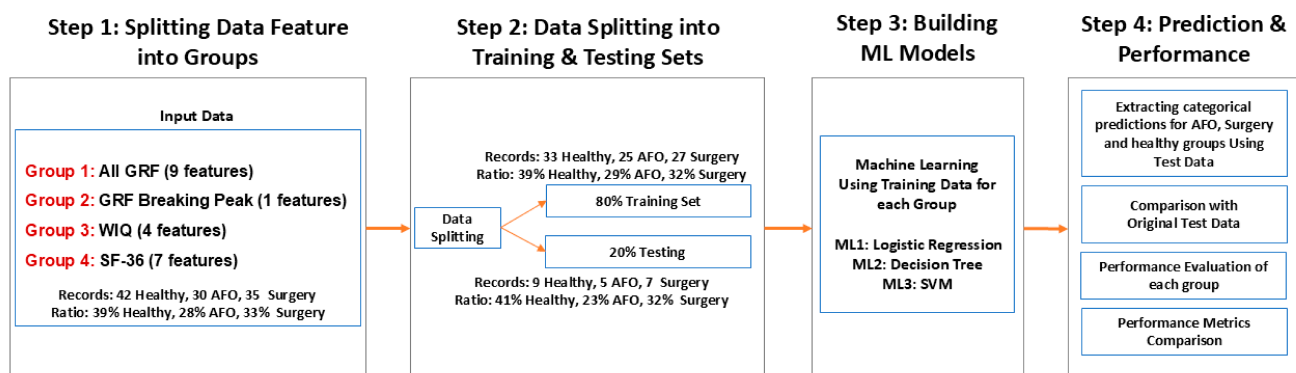


Figure 2. Flowchart overview of feature-level assessment for PAD treatment decision: A comparative analysis of individual features using Logistic Regression, Decision Tree, and SVM models.

The data is initially segmented into individual features and then subjected to a machine learning pipeline involving data splitting, data preprocessing, model training using default hyperparameters, and performance evaluation. Given the limited sample size (107 rows from 55 unique patients), we did not perform a formal hyperparameter search: nested cross-validation and tuning on a dataset this small risks overfitting the search itself rather than improving genuinely generalizable performance and is noted as future work once a larger cohort is available. For the same reason, we did not evaluate higher-capacity ensemble or deep-learning models (e.g., XGBoost, LightGBM) in this feature-level analysis. We continue to employ Logistic Regression [31,32], Decision Tree [33], and Support Vector Machine (SVM) [34] models for this analysis. The goal is to identify the most informative

individual features for PAD treatment quantification, offering insights into the potential for streamlined, feature-centric diagnostic approaches. This feature-level assessment aims to refine our interpretation of which aspects of GRF, WIQ, and SF-36 most relate to PAD treatment decisions. It could thus be a focal point in future diagnostic assessment tools.

2.5. Validation Design

Because AFO patients were evaluated on multiple visits prior to treatment (Table 1), a naive row-level train/test split can place the same patient's repeated visits on both sides of the split. We verified the impact of this by re-running every feature/model combination in Figure 3c (23 feature sets) with 100 repeated patient-grouped splits (grouped by subject number) alongside the original row-level method. The mean accuracy difference (grouped – row-level) across all 23 features was -0.6 percentage points (SD 1.6pp; range -3.2 pp to $+2.4$ pp), and the mean MCC difference was -0.3 pp (SD 2.1pp)—a real but small effect. An order of magnitude smaller than the 7–11 percentage-point split-to-split standard deviation was observed at this sample size. We conclude that single-split reporting variance, not leakage per se, is the dominant source of uncertainty in these results, and we report repeated-split mean $\pm$ SD alongside the original single-split values throughout Figure 3c accordingly. No formal hyperparameter search was performed, for the reasons stated above; results should be interpreted as representative of default-hyperparameter model performance rather than an upper bound achievable with tuning.

(a) Pre-intervention feature distributions by group (representative feature per source)

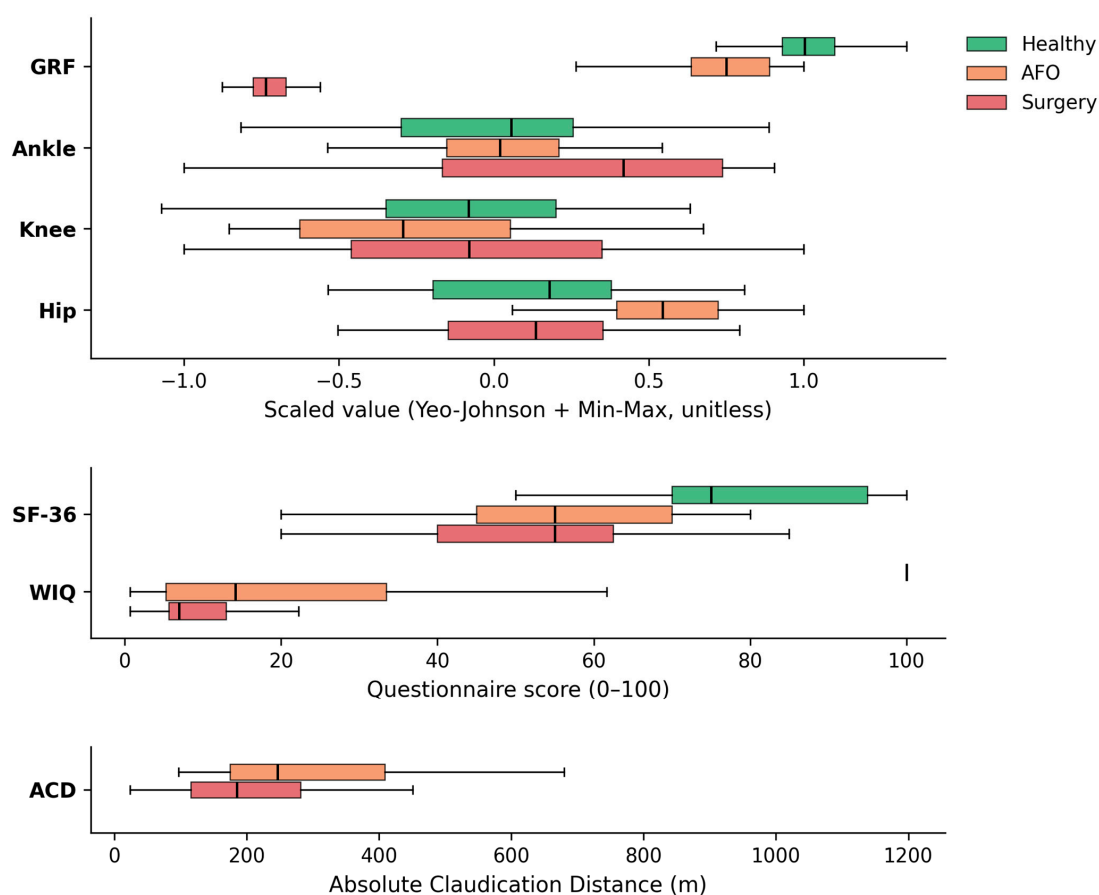
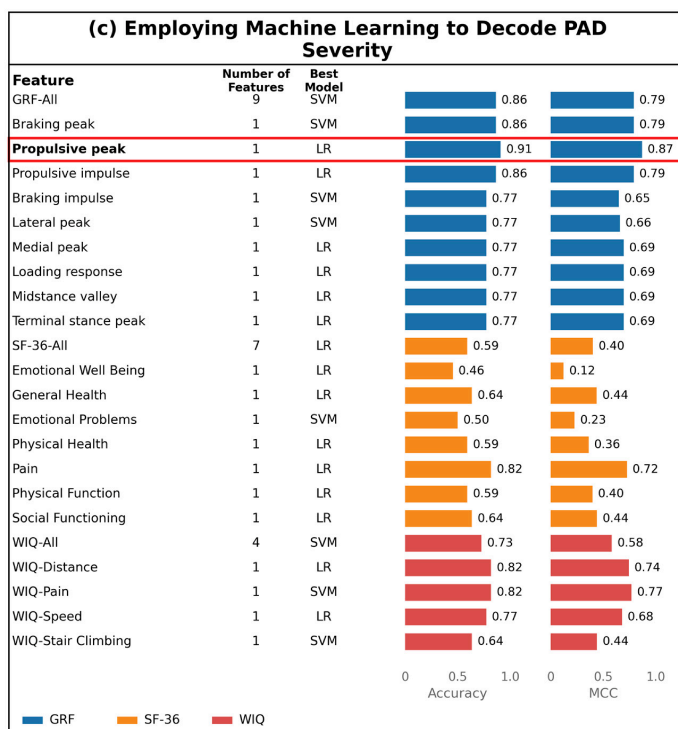


Figure 3. Cont.

(b). Statistical Testing: Median Bonferroni-corrected Mann–Whitney U p -value per measurement source (* = significant at $p < 0.05$)			
Measurement Source	# of Features	Pair	p -value
GRF	9	Healthy vs. pre-AFO	0.003 *
		Healthy vs. pre-surgery	0.000 *
		AFO vs. pre-surgery	0.000 *
Ankle	7	Healthy vs. pre-AFO	0.018 *
		Healthy vs. pre-surgery	1.000
		AFO vs. pre-surgery	0.559
Hip	7	Healthy vs. pre-AFO	0.000 *
		Healthy vs. pre-surgery	0.206
		AFO vs. pre-surgery	0.000 *
Knee	7	Healthy vs. pre-AFO	1.000
		Healthy vs. pre-surgery	0.075
		AFO vs. pre-surgery	1.000
WIQ	4	Healthy vs. pre-AFO	0.000 *
		Healthy vs. pre-surgery	0.000 *
		AFO vs. pre-surgery	1.000
SF-36	7	Healthy vs. pre-AFO	0.000 *
		Healthy vs. pre-surgery	0.000 *
		AFO vs. pre-surgery	1.000
ACD	1	AFO vs. pre-surgery	0.030 *



Feature	Model Type	Data Split	Accuracy	MCC
Propulsive peak	Logistic Regression	test	0.91	0.87
		train	0.85	0.79

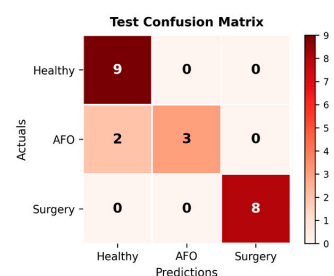
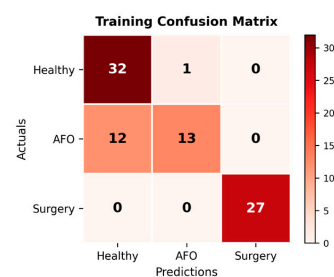
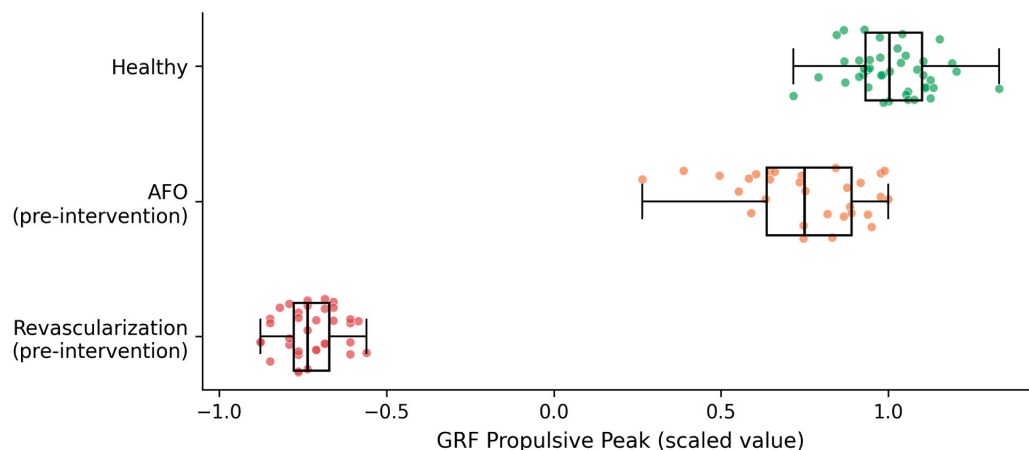


Figure 3. Cont.

(d) GRF Propulsive Peak separates groups by actual treatment received (individual subjects; real data)



(e) GRF Propulsive Peak decision scale
Individual subjects (real data); zone boundaries are this cohort's fitted logistic-regression crossover points, not independently validated diagnostic cutoffs.

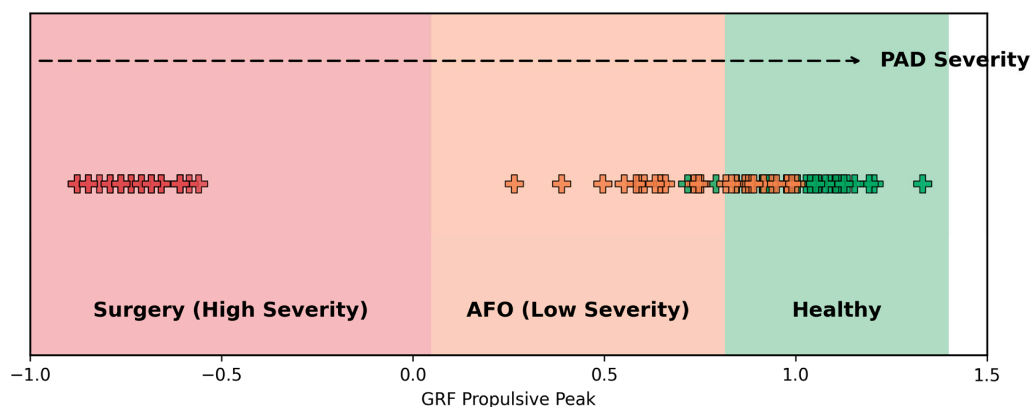


Figure 3. (a) Pre-intervention distribution of all features categorized based on diagnosis and treatment groups (healthy, AFO, and revascularization). (b) Statistical comparison across measurement sources (Mann–Whitney U test, median Bonferroni-corrected p -value per source; see Methods). (c) Comparative performance of machine learning models in feature-level PAD treatment quantification across all features as independent inputs. (d) GRF propulsive peak separates groups by actual treatment received (individual subjects; real data). (e) Illustrative single-feature decision-scale schematic (conceptual): boundaries are this cohort's fitted logistic-regression crossover points, not independently validated diagnostic cutoffs, shown to illustrate the framework's potential rather than as a clinical tool. Acronyms: GRF = ground reaction force; AFO = ankle-foot orthosis; SF-36 = 36-Item Short Form Health Survey; WIQ = Walking Impairment Questionnaire; ACD = absolute claudication distance; SVM = Support Vector Machine; LR = Logistic Regression; PAD = peripheral artery disease.

2.6. Stratification of Modeled Treatment Selection Using Optimal Features

This section aims to provide a more granular understanding of treatment-selection prediction by leveraging the best-performing feature identified in the previous analysis. Figure 3d shows that GRF propulsive peak separates the three groups using real, individual-subject pre-intervention values; no synthetic or hypothetical data are involved. To illustrate how the underlying Logistic Regression model (Figure 3c) uses this single feature, Figure 3e presents an illustrative decision-scale schematic: we swept GRF propulsive peak values across their observed range and recorded the model's predicted class at each point, pro-

ducing three contiguous zones separated by the model's fitted decision boundaries. These boundaries are this cohort's fitted crossover points, not independently validated diagnostic cutoffs, and Figure 3e should be read as an illustration of the model's decision logic rather than a clinical tool.

2.7. Quantifying Intervention Impact Using Effect Size

Effect size [35] is a quantitative representation of the strength of the relationship between variables. Unlike statistical tests that merely confirm the existence of an effect or relationship, the effect size elucidates its magnitude, independent of sample size. In our context, with a control group and two patient groups (revascularization and AFO), both pre- and post-intervention, the effect size aids in quantifying the magnitude of change due to intervention in each patient group and offers a comparative measure against the control.

We employ Cohen's *d*, a measure calculated as the difference between two means divided by the pooled standard deviation, to quantify the differences between our groups. For instance, comparisons such as healthy vs. pre-revascularization provide insights into the deviation of the revascularization group from the control before the intervention. At the same time, healthy vs. post-revascularization reveals the deviation post-intervention. Analogously, comparisons for the AFO group, healthy vs. pre-AFO and healthy vs. post-AFO, provide similar insights. Cohen's guidelines suggest that a *d* value of 0.2 indicates a 'small' effect size, 0.5 a 'medium' effect size, and 0.8 a 'large' effect size. The sign of Cohen's *d* further provides directional information: a positive sign indicates the post-intervention group has higher scores than the pre-intervention or control group. In contrast, a negative sign indicates the opposite. This approach aims to provide a nuanced, quantitative understanding of PAD severity and its response to interventions.

2.8. Machine Learning Approach Estimating Post-Treatment Outcomes

The section utilizes pre-intervention data, including gait kinematics and kinetics measurements, WIQ scores, and ACD as predictive features. As shown in Figure 4, the initial step involves gathering pre-intervention data from a cohort of 38 patients diagnosed with PAD. This data is sourced from various measurements, including gait kinematics covering ankle, hip, and knee features and gait kinetics focusing on ground reaction forces (GRF), WIQ, and ACD. Additionally, patients are categorized based on their treatment group: AFO for those with low severity and revascularization for those with high severity. Out of the total, 30 patients underwent revascularization due to their high severity, while eight were treated with AFO due to their lower severity.

The data is split into training and testing sets. The training set comprises 28 patients: five from the AFO group and 23 from the revascularization group. The remaining 10 patients (three AFO, seven revascularization) are reserved for the test set. This division allows the model to be trained on a diverse dataset and subsequently validated on unseen data to gauge its predictive accuracy.

Before feeding the data into machine learning models, it undergoes a feature selection process to enhance its predictive capability. This method employs principal component analysis (PCA) [36] for dimensionality reduction. PCA transforms the original features into a set of linearly uncorrelated variables, capturing the most significant patterns in the data while reducing its complexity (90% of the total variance is maintained). We applied this feature selection method at the training data source level to ensure that every data source contributed to the study while avoiding data leakage between the train and test sets.

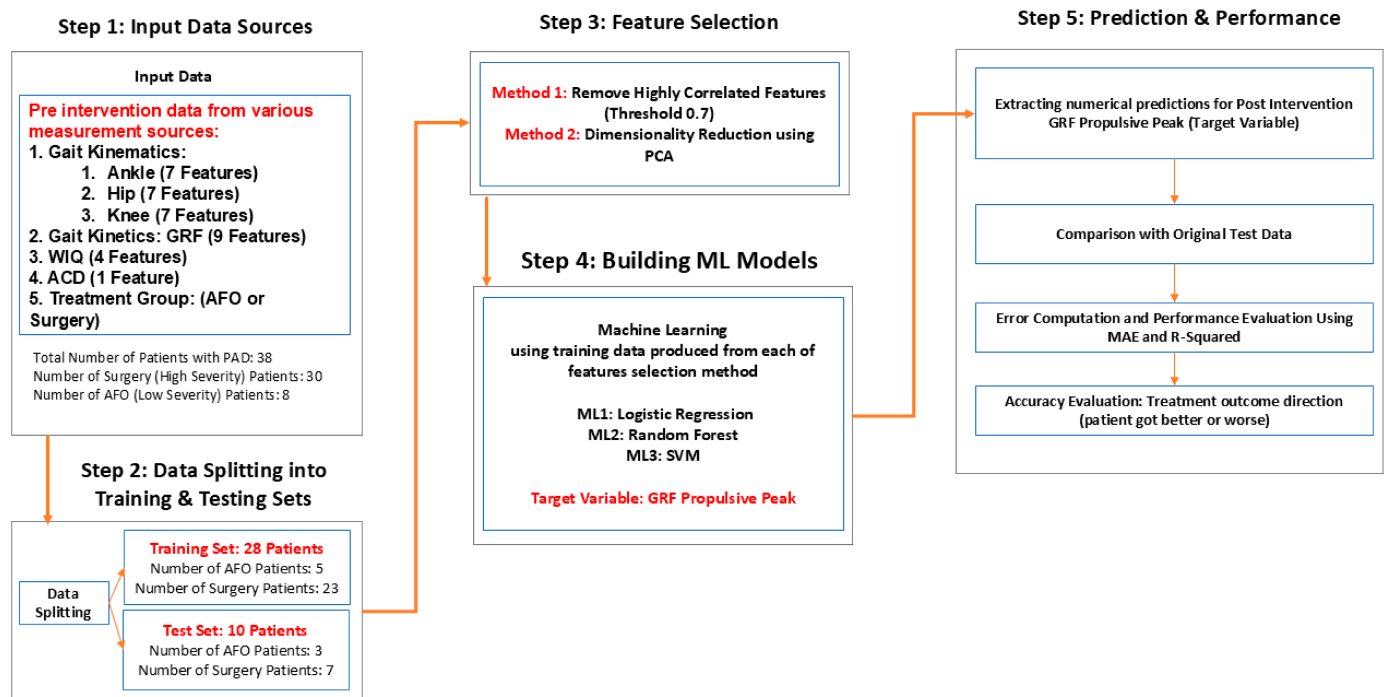


Figure 4. Flowchart depicting the machine learning approach for predicting PAD treatment outcomes using pre-intervention data and evaluating model performance.

Three machine learning models were trained on the features of Linear Regression, Support Vector Regression (SVR), and Random Forest [37]. These models were chosen for their versatility and capability to handle complex nonlinear relationships in the data. The trained models were then used to predict the GRF propulsive peak post-intervention, which, as established in previous chapters, serves as a measure of PAD severity and a reference for treatment outcome.

Three machine learning models are trained with optimized features: Logistic Regression, Random Forest, and Support Vector Machine (SVM). Each model is trained using the data produced from both feature selection methods. Post-training, the models predict the post-intervention GRF propulsive peak (established in the previous chapter, which is the most distinctive measure for PAD severity), serving as the target variable. These predictions are then compared with the original test data. The model's performance is evaluated using MAE for average prediction error magnitude and R-squared for the proportion of variance explained by the model. Additionally, the accuracy of the treatment outcome direction, indicating whether a patient's condition improves or worsens, is assessed.

3. Results

Figure 3a presents a comprehensive distribution of all features in our study for healthy control patients with PAD before two different treatments (AFO and revascularization).

The features are represented after the post-application of the Yeo–Johnson transformation. It is evident from the figure that kinematics features (ankle, hip, and knee) failed to differentiate between healthy individuals and those with PAD. While qualitative questionnaire evaluations, such as SF-36 and WIQ, effectively distinguished between healthy controls and patients with PAD, they failed to discern PAD patients who were treated by wearing an AFO or undergoing a revascularization operation. The GRF features distinguished all three categories and presented an ordered patient distribution. This observation highlights the potential of GRF features as a potent instrument for helping to predict the treatment type a physician may end up selecting for a patient with PAD.

Figure 3b presents the outcomes of the Mann–Whitney U test aiming to recognize statistically significant differences among the healthy, AFO, and revascularization groups. The results confirm the observations that GRF’s predictive capacity consistently outperforms conventional clinical metrics.

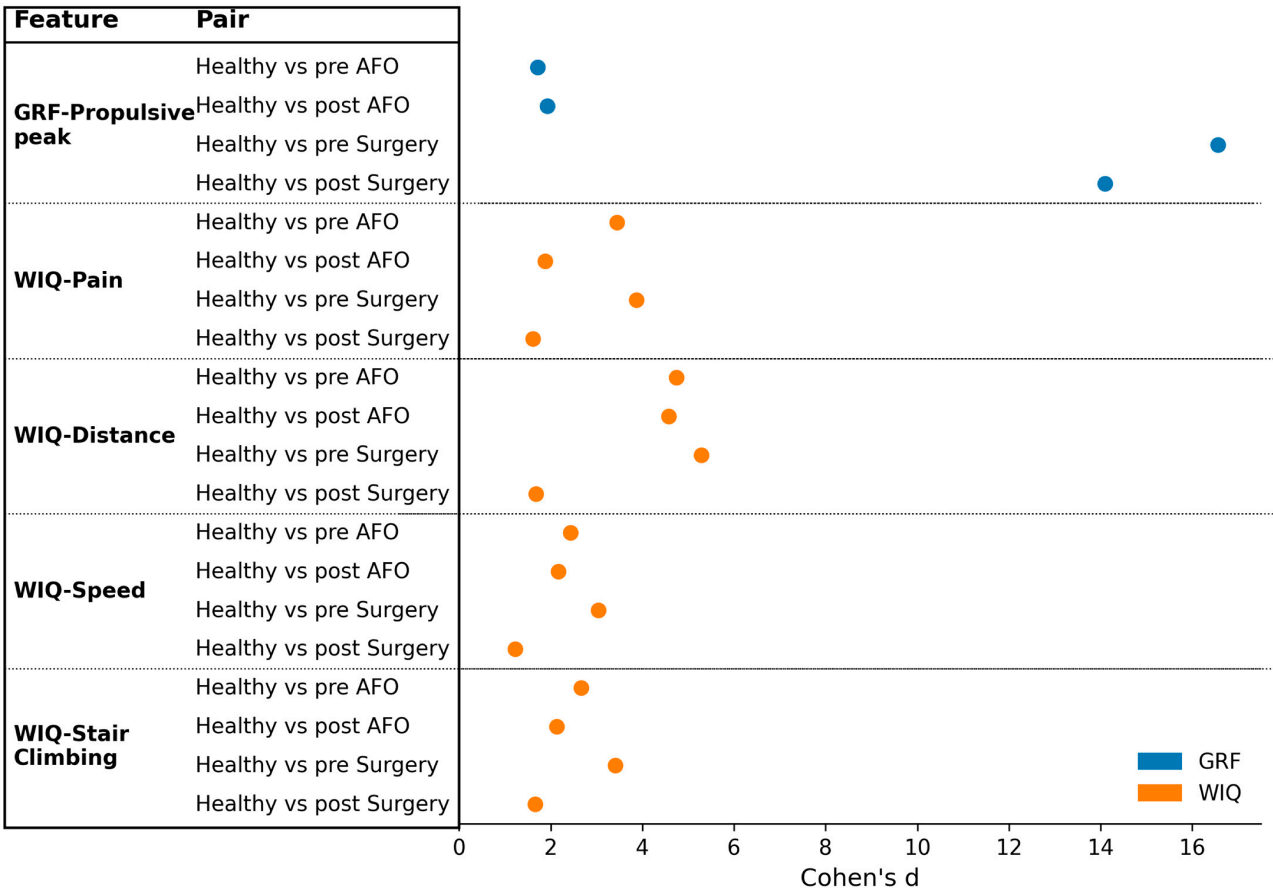
Next, the results of using machine learning models to predict the PAD treatment type using the different features are shown in Figure 3c. The GRF, using Logistic Regression with a single feature, “Propulsive peak,” achieved an accuracy [38] of 90.9%, a balanced accuracy [38] of 86.7%, and a Matthew’s correlation coefficient (MCC) [38] of 0.868. The GRF’s high performance across these metrics, particularly the MCC that accounts for unbalanced data, highlights its effectiveness in accurately predicting the PAD treatment the physician ended up choosing. Additionally, the consistent results between the model’s training and testing performance suggest that the GRF signature effectively captures the essential patterns in the data without being overly fitted to the training set or missing the signature’s complexity. This confirms the robustness of GRF features for reliable treatment choice prediction in PAD.

To assess the robustness of this result, we additionally re-ran the classifier using 100 repeated patient-grouped train/test splits (grouping by subject number so that a given patient’s repeated visits never span both sets; see Methods, Validation Design). Across these repeats, mean performance had an 82.3% accuracy (SD 8.6%), an 81.0% balanced accuracy (SD 8.6%), and a MCC of 0.758 (SD 0.108)—consistent with, though somewhat lower than, the single-split result above, and confirming that the original random state split was a favorable but plausible draw rather than an outlier requiring leakage to explain.

Next, interest is shifted to understanding the response of patients with PAD to a treatment using effect size analysis [35]. This analysis quantitatively represents the strength of the relationship between two variables using Cohen’s *d* measure. Cohen’s guidelines suggest that a large value of *d* indicates a larger effect size (there is a significant difference between the two variables). In our context, with a control group and two patient groups (revascularization and AFO), both pre- and post-intervention, the effect size aids in quantifying the magnitude of differences between patients with AFO and patients with revascularization, and the change due to intervention in each patient group, and offers a comparative measure against the control. For instance, comparisons such as healthy vs. pre-revascularization provide insights into the deviation of the revascularization group from the control before the intervention. At the same time, healthy vs. post-revascularization reveals the deviation post-intervention.

Figure 5a illustrates the effect sizes, measured with Cohen’s *d*, for one GRF measure and four WIQ measures, compared to a healthy control group. The top section of Figure 5a displays the effect size magnitude based on the GRF propulsive peak. The figure highlights a significant difference in Cohen’s *d* between PAD patients who underwent revascularization and those who received an AFO intervention. This difference supports the distinct GRF differences among PAD patients treated with AFO compared to revascularization. Additionally, the pronounced effect size difference between post-revascularization and pre-revascularization, as opposed to the smaller difference between pre- and post-AFO treatment, may indicate that revascularization has a more substantial impact as a treatment than AFO, which is expected since the arterial blockage (primary problem for claudication) is treated. Similar conclusions are not as evident in Figure 5a for the various WIQ scores. This again confirms the importance of GRF as a single important parameter in PAD treatment decision-making.

(a) Effect size (Cohen’s d) comparisons for GRF and WIQ



(b) ML Regression Comparison

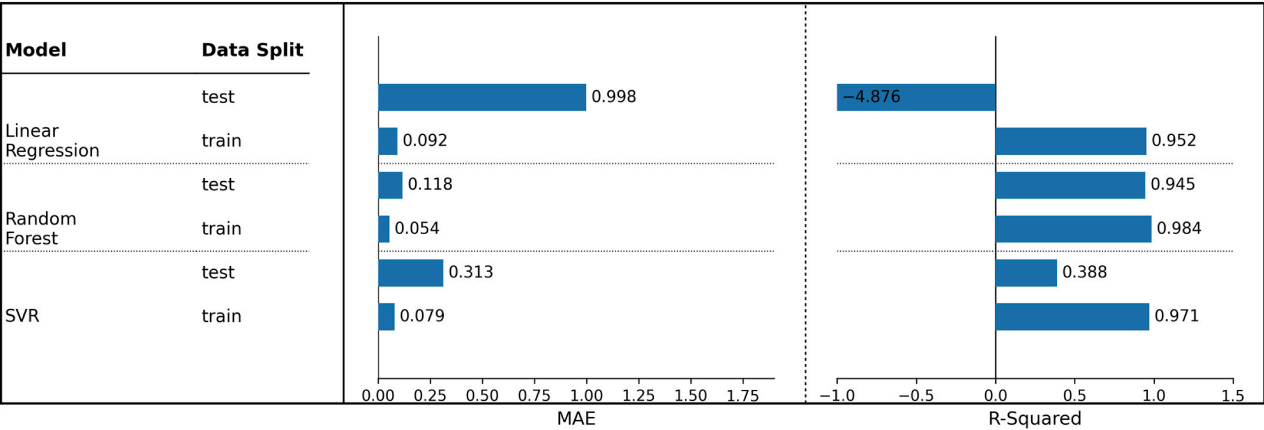
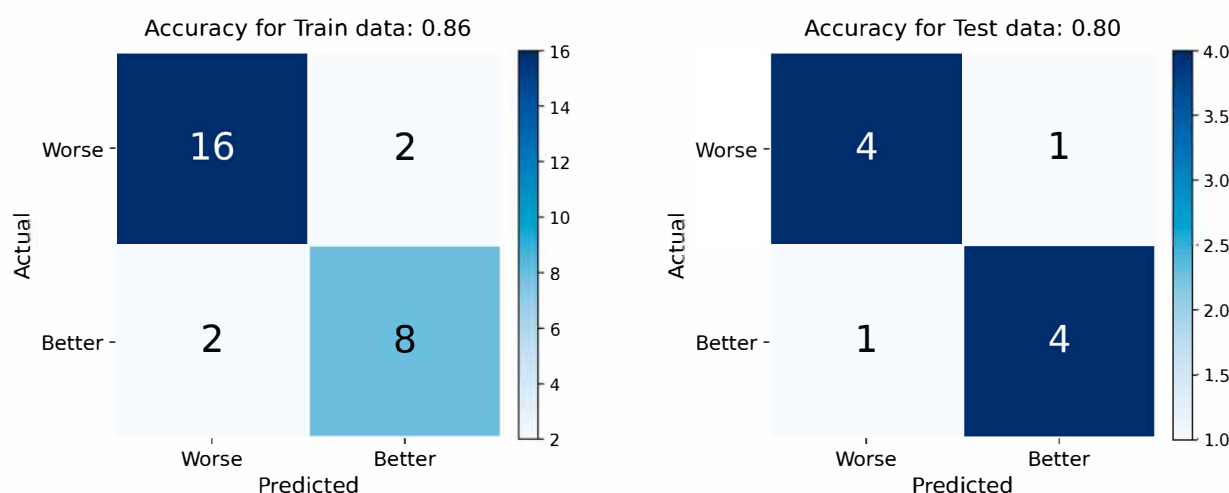


Figure 5. Cont.

(c)



(d) GRF Prediction after treatment

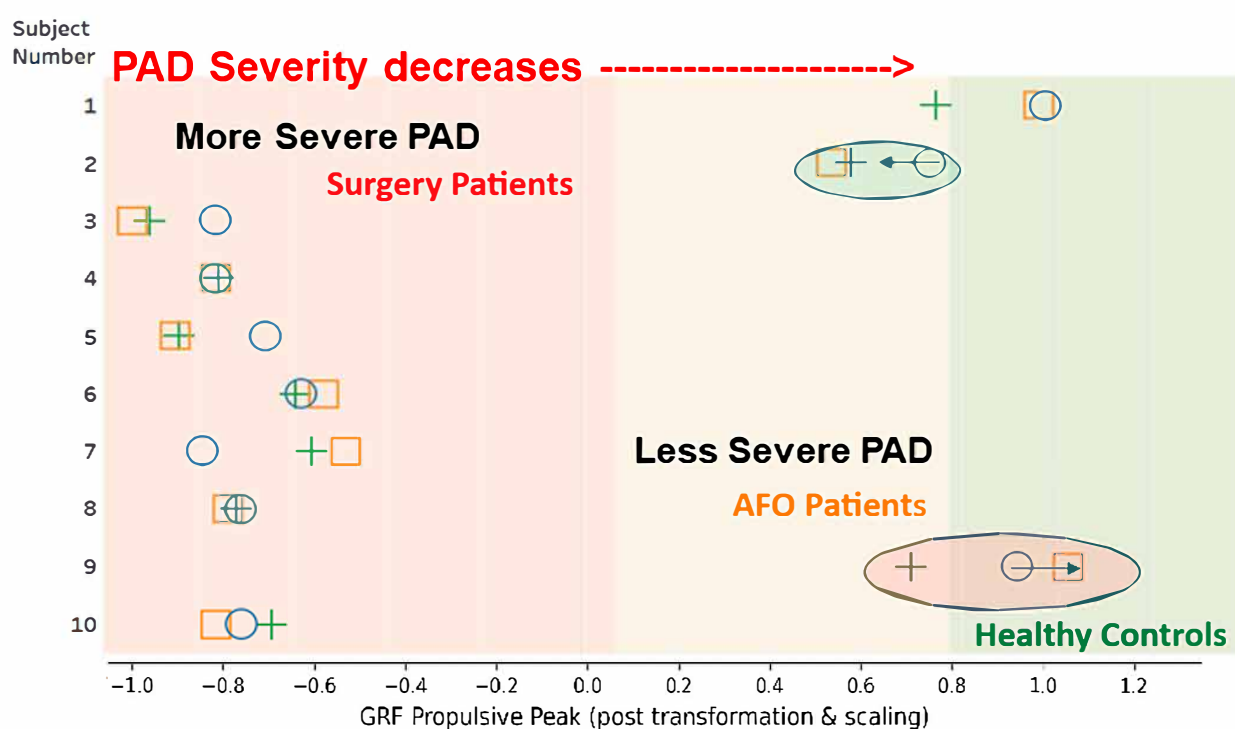


Figure 5. (a) Effect size (Cohen's d) comparisons for GRF and WIQ scores between healthy individuals and patients with PAD, both pre- and post-intervention. Higher Cohen's d values indicate larger differences between the groups, suggesting more pronounced effects of interventions or more distinct group characteristics. Blue dots represent comparisons involving the 'pre' status, and orange dots signify the 'post' status. (b) Performance metrics comparison of machine learning models including Linear Regression, Random Forest, and SVR. Metrics shown are MAE, R-Squared, and adjusted R-Squared, assessing model accuracy and predictive capability with both correlation and PCA feature selection methods on training and test datasets. (c) Confusion matrices for GRF prediction after treatment using the Random Forest model on training and test datasets, showing the counts of correct and incorrect predictions. (d) Individual predictions of GRF propulsive peak for all subjects post-intervention. The graph plots the actual vs. predicted GRF propulsive peak values, illustrating prediction accuracy.

Due to its importance, our last analysis is to develop a machine learning model to predict GRF after treatment. We anticipate that this prediction is the first step toward predicting post-intervention effectiveness. To this end, the model is trained using pre-intervention data, including gait kinematics and kinetics measurements, and WIQ scores to predict GRF propulsive peak after the treatment. Figure 5b results provide a comprehensive overview of the performance metrics for various machine learning models (Linear Regression, Random Forest, and SVM). Figure 5c shows that the Random Forest model consistently outperforms the other models in mean absolute error (MAE) and R-squared. Furthermore, we added a binary representation for the model outcomes by assessing whether the model can predict an increase or decrease in the GRF value. The confusion matrices for both training and test datasets, as depicted in Figure 5c, provide a clear visual representation of the model's predictions against the actual outcomes. For the training data, the model achieves an accuracy of 86%, while for the test data, the accuracy stands at 80%. These high accuracy scores indicate the model's ability to predict changes in the GRF after performing treatment. As with the treatment-classification model above, this is drawn from a single train/test split without a repeated-split uncertainty estimate; extending the repeated-split validation used for Figure 3c to this regression pipeline is recommended for future work.

The individual predictions of GRF propulsive peak post-intervention for each subject, as shown in Figure 5d, offer a granular view of the model's performance. Each subject's data point is plotted against their pre-intervention GRF propulsive peak, with distinct markers indicating the actual post-intervention values and the model's predictions. A consistent alignment between the actual and predicted values is evident across most subjects. This individual-level analysis complements the accuracy metrics, reinforcing the model's ability to predict GRF after treatment. The colored bars visually represent the model's confidence in its predictions, with the proximity of the actual and predicted markers indicating the precision of the model's forecasts.

4. Discussion

The anterior–posterior ground-reaction force (GRF) waveform captures braking and propulsion during stance. The propulsive peak, defined as the maximum positive A–P GRF normalized to body weight, reflects late-stance ankle plantarflexor output. Physiologically, GRF acts as a powerful summary signal, combining the downstream effects of reduced perfusion in PAD, including loss of muscle power and endurance, altered neuromotor control, compensatory joint mechanics, and pain-avoidant strategies, all within a single waveform. It therefore serves as a single measure of cumulative tissue injury and functional reserve. Using a simple logistic model with one input, the pre-treatment propulsive peak, we created an easy-to-understand, single-number decision scale with three zones (pipeline-scaled units): Healthy (−1 to −0.8), nonoperative management (−0.8 to 0.25), and revascularization (0.25 to 1.0).

Because this feature directly reflects distal power, post-treatment increases provide a physiologically meaningful response marker. We built models that use a patient's baseline GRF (and other features) together with planned treatment, nonoperative management (AFO) or revascularization, to predict post-treatment GRF. This allows clinicians to anticipate recovery of push-off power with each intervention. Forecasts aligned with clinical expectations, with larger gains projected for surgical patients. GRF-based predictions outperformed traditional kinematic measures and patient-reported outcomes, indicating that gait mechanics capture clinically important information often missed by routine assessments. Clinically, GRF offers a compact, objective proxy for overall clinical judgment that can support screening, severity grading, and monitoring responses to revascularization, supervised exercise, and pharmacologic care.

Important limitations remain. Treatment decisions in this study reflect the practice of a single surgical team, and the classification model predicts what this team selected rather than an independently adjudicated or objectively optimal treatment; treatment labels reflect both disease severity and this team's local decision-making pattern. Because both the treatment-selection labels and the GRF decision-scale thresholds in Figure 3d were derived entirely from this one practice's decision patterns, the reported accuracy and the specific thresholds should be treated as internal, hypothesis-generating estimates rather than externally validated diagnostic thresholds. External, multicenter validation is needed to confirm generalizability and calibrate decision boundaries at other sites.

The GRF propulsive peak and its direction of change after treatment are single biomechanical surrogates. It has not been validated in this dataset against treatment durability, adverse events, walking capacity outside the laboratory, or patient-reported benefits, and should not be interpreted as a composite measure of overall treatment success.

In practice, we envision the GRF decision scale in Figure 3e as one input alongside standard clinical workup rather than a standalone trigger for intervention. Predictions near a zone boundary should be interpreted cautiously given the small sample underlying these thresholds and should prompt additional clinical evaluation rather than a default course of action. Because the model reflects this single team's historical decision pattern, disagreement between a model prediction and clinical judgment should not be assumed to indicate model error; it may instead reflect patient-specific factors outside the model's feature set, or a genuinely different practice pattern at another site. Future work should quantify GRF responsiveness across therapies, relate key GRF features to muscle oxygenation during walking, and test whether GRF-based indices improve triage, treatment selection, and follow-up. As wearable sensors mature, estimating GRF from accelerometers and instrumented insoles could enable passive, real-world monitoring and continuous risk stratification, extending assessment beyond specialized gait laboratories.

5. Conclusions

GRF, particularly the propulsive peak, provides an easy-to-understand and mechanistically grounded biomarker that predicts clinician treatment selection and functional recovery in PAD. With appropriate validation, GRF-informed models could become a scalable triage tool in vascular care and help direct patients with locomotor complaints to the right specialty, accelerating timely referral and individualized management.

Author Contributions: A.A.R. conducted the analysis, F.F. Provided the data and edited the paper, M.A.T. Reviewed and edited the paper, I.I.P. and S.M. provided the funding, clinical feedback, and edited the paper, F.A. managed the team and drafted with A.A.R. the initial version of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institutes of Health (R01AG034995, R01HD090333, R01AG049868), United States Department of Veterans Affairs Rehabilitation Research and Development Service (I01RX000604, I01RX003266), and the University of Nebraska Collaboration Initiative.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Boards. The study was approved by the Institutional Review Boards at the University of Nebraska Medical Center and the Nebraska-Western Iowa Veteran Affairs Medical Center (IRB #1576199, Approval date is 7 February 2017; And IRB #1576204, Approval date is 24 August 2010).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and analyzed during the current study are not publicly available due to restrictions related to human subject privacy and ethical considerations, but are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Soyoye, D.O.; Abiodun, O.O.; Ikem, P.T.; Kolawole, B.A.; Akintomide, A.O. Diabetes and peripheral artery disease: A review. *World J. Diabetes* **2021**, *12*, 827–838. [CrossRef] [PubMed]
2. Lin, J.; Chen, Y.; Jiang, N.; Li, Z.; Xu, S. Burden of peripheral artery disease and its attributable risk factors in 204 countries and territories from 1990 to 2019. *Front. Cardiovasc. Med.* **2022**, *9*, 868370. [CrossRef] [PubMed]
3. Golledge, J. Update on the pathophysiology and medical treatment of peripheral artery disease. *Nat. Rev. Cardiol.* **2022**, *19*, 456–474. [CrossRef] [PubMed]
4. Gerhard-Herman, M.D.; Gornik, H.L.; Barrett, C.; Barshes, N.R.; Corriere, M.A.; Drachman, D.E.; Fleisher, L.A.; Fowkes, F.G.R.; Hamburg, N.M.; Kinlay, S.; et al. 2016 AHA/ACC Guideline on the Management of Patients with Lower Extremity Peripheral Artery Disease. *Circulation* **2017**, *135*, e726–e779. [CrossRef] [PubMed]
5. McDermott, M.M.; Guralnik, J.M.; Criqui, M.H.; Liu, K.; Kibbe, M.R.; Ferrucci, L. Six-minute walk is a better outcome measure than treadmill walking tests in therapeutic trials of patients with peripheral artery disease. *Circulation* **2014**, *130*, 61–68. [CrossRef] [PubMed]
6. Dinkel, D.; Hassan, M.; Rech, J.P.; DeSpiegelaere, H.; Johanning, J.; Pipinos, I.; Myers, S. Assessing wear time and perceptions of wearing an ankle foot orthosis in patients with peripheral artery disease. *PM&R* **2023**, *15*, 493–500.
7. Myers, S.; Dinkel, D.; Hassan, M.; Despiegelaere, H.; Johanning, J.; Pipinos, I. Examining ankle foot orthosis wear time in patients with peripheral artery disease. *Innov. Aging* **2020**, *4*, 211. [CrossRef]
8. Scatena, A.; Apicella, M.; Mantuano, M.; Liistro, F.; Ventoruzzo, G.; Petrucci, P.; Miranda, C.; Monge, L.; Raghianti, B.; Silverii, A.; et al. Bypass surgery versus endovascular revascularization for occlusive infrainguinal peripheral artery disease: A meta-analysis of randomized controlled trials for the development of the Italian guidelines for the treatment of diabetic foot syndrome. *Acta Diabetol.* **2023**, *61*, 19–28. [CrossRef] [PubMed]
9. Rontoyanni, V.G.; Lopez, O.N.; Fankhauser, G.T.; Cheema, Z.F.; Rasmussen, B.B.; Porter, C. Mitochondrial bioenergetics in the metabolic myopathy accompanying peripheral artery disease. *Front. Physiol.* **2017**, *8*, 141. [CrossRef] [PubMed]
10. Klukowska, A.M.; Schröder, M.L.; Stienen, M.N.; Staartjes, V.E. Objective functional impairment in lumbar degenerative disease: Concurrent validity of the baseline severity stratification for the five-repetition sit-to-stand test. *J. Neurosurg. Spine* **2020**, *33*, 4–11. [CrossRef] [PubMed]
11. Aboyans, V.; Ricco, J.-B.; Bartelink, M.-L.E.L.; Björck, M.; Brodmann, M.; Cohnert, T.; Collet, J.-P.; Czerny, M.; De Carlo, M.; Debus, S.; et al. 2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS). *Eur. Heart J.* **2018**, *39*, 763–816. [CrossRef] [PubMed]
12. Ware, J.E., Jr.; Kosinski, M.; Gandek, B. *The SF-36 Health Survey: Manual and Interpretation Guide*; Quality Metric Inc.: Lincoln, RI, USA, 2000.
13. Eghbalieh, S.D.D.; Dardik, A. Basic data related to surgical infrainguinal revascularization procedures: A twenty year update. *Ann. Vasc. Surg.* **2011**, *25*, 413–422. [CrossRef]
14. Lundgren, F.; Dahllöf, A.G.; Lundholm, K.; Scherstén, T.; Volkmann, R. Intermittent claudication—Surgical reconstruction or physical training? A prospective randomized trial of treatment efficiency. *Ann. Surg.* **1989**, *209*, 346–355. [CrossRef] [PubMed]
15. Murphy, T.P.; Cutlip, D.E.; Regensteiner, J.G.; Mohler, E.R.; Cohen, D.J.; Reynolds, M.R.; Massaro, J.M.; Lewis, B.A.; Cerezo, J.; Oldenburg, N.C.; et al. Supervised exercise versus primary stenting for claudication resulting from aortoiliac peripheral artery disease: Six-month outcomes from the claudication: Exercise versus endoluminal revascularization (CLEVER) study. *Circulation* **2012**, *125*, 130–139. [CrossRef] [PubMed]
16. Whyman, M.R.; Fowkes, F.; Kerracher, E.; Gillespie, I.; Lee, A.; Housley, E.; Ruckley, C. Is intermittent claudication improved by percutaneous transluminal angioplasty? A randomized controlled trial. *J. Vasc. Surg.* **1997**, *26*, 551–557. [CrossRef] [PubMed]
17. RAND. 36-Item Short Form Survey (SF-36) Scoring Instructions. *Med. Outcomes Study* **2016**, 2–6. Available online: <https://www.rand.org/health/surveys/mos/36-item-short-form/scoring.html> (accessed on 29 June 2026).
18. Aant, N.; Arabbeiki, M.; Niroomand, M.R. Applications of machine learning for peripheral artery disease diagnosis and management: A systematic review. *Comput. Biol. Med.* **2025**, *196*, 110744. [CrossRef] [PubMed]
19. Chekroud, A.M.; Bondar, J.; Delgadillo, J.; Doherty, G.; Wasil, A.; Fokkema, M.; Cohen, Z.; Belgrave, D.; DeRubeis, R.; Iniesta, R.; et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry* **2021**, *20*, 154–170. [CrossRef] [PubMed]
20. Al-Ramini, A.; Hassan, M.; Fallahtafi, F.; Takallou, M.A.; Rahman, H.; Qolomany, B.; Pipinos, I.I.; Alsaleem, F.; Myers, S.A. Machine learning-based peripheral artery disease identification using laboratory-based gait data. *Sensors* **2022**, *22*, 7432. [CrossRef] [PubMed]

21. Takallou, M.A.; Fallahtafti, F.; Hassan, M.; Al-Ramini, A.; Qolomany, B.; Pipinos, I.; Myers, S.; Alsaleem, F. Diagnosis of disease affecting gait with a body acceleration-based model using reflected marker data for training and a wearable accelerometer for implementation. *Sci. Rep.* **2024**, *14*, 1075. [[CrossRef](#)] [[PubMed](#)]
22. Chen, S.-J.; Pipinos, I.; Johanning, J.; Radovic, M.; Huisinga, J.M.; Myers, S.A.; Stergiou, N. Bilateral claudication results in alterations in the gait biomechanics at the hip and ankle joints. *J. Biomech.* **2008**, *41*, 2506–2514. [[CrossRef](#)] [[PubMed](#)]
23. McDermott, M.M.; Liu, K.; Guralnik, J.M.; Martin, G.J.; Criqui, M.H.; Greenland, P. Measurement of walking endurance and walking velocity with questionnaire: Validation of the walking impairment questionnaire in men and women with peripheral arterial disease. *J. Vasc. Surg.* **1998**, *28*, 1072–1081. [[CrossRef](#)] [[PubMed](#)]
24. Myers, S.A.; Johanning, J.M.; Stergiou, N.; Lynch, T.G.; Longo, G.M.; Pipinos, I.I. Claudication distances and the Walking Impairment Questionnaire best describe the ambulatory limitations in patients with symptomatic peripheral arterial disease. *J. Vasc. Surg.* **2008**, *47*, 550–555. [[CrossRef](#)] [[PubMed](#)]
25. Regensteiner, J.G.; Steiner, J.F.; Panzer, R.J.; Hiatt, W.R. Evaluation of walking impairment by questionnaire in patients with peripheral arterial disease. *J. Vasc. Med. Biol.* **1990**, *2*, 142–152.
26. Brazier, J.E.; Harper, R.; Jones, N.M.; O’Cathain, A.; Thomas, K.J.; Usherwood, T.; Westlake, L. Validating the SF-36 health survey questionnaire: New outcome measure for primary care. *BMJ* **1992**, *305*, 160–164. [[CrossRef](#)] [[PubMed](#)]
27. Celis, R.; Pipinos, I.I.; Scott-Pandorf, M.M.; Myers, S.A.; Stergiou, N.; Johanning, J.M. Peripheral arterial disease affects kinematics during walking. *J. Vasc. Surg.* **2009**, *49*, 127–132. [[CrossRef](#)] [[PubMed](#)]
28. Atkinson, A.C.; Riani, M.; Corbellini, A. The Box–Cox transformation: Review and extensions. *Stat. Sci.* **2021**, *36*, 239–255. [[CrossRef](#)]
29. Ruxton, G.D. The unequal variance *t*-test is an underused alternative to Student’s *t*-test and the Mann–Whitney U test. *Behav. Ecol.* **2006**, *17*, 688–690. [[CrossRef](#)]
30. Hazra, A.; Gogtay, N. Biostatistics series module 3: Comparing groups: Numerical variables. *Indian J. Dermatol.* **2016**, *61*, 251. [[CrossRef](#)] [[PubMed](#)]
31. Menard, S. *Applied Logistic Regression Analysis*; Sage: Thousand Oaks, CA, USA, 2002.
32. McDermott, M.M.; Kerwin, D.R.; Liu, K.; Martin, G.J.; O’Brien, E.; Kaplan, H.; Greenland, P. Prevalence and significance of unrecognized lower extremity peripheral arterial disease in general medicine practice. *J. Gen. Intern. Med.* **2001**, *16*, 384–390. [[CrossRef](#)] [[PubMed](#)]
33. Podgorelec, V.; Kokol, P.; Stiglic, B.; Rozman, I. Decision trees: An overview and their use in medicine. *J. Med. Syst.* **2002**, *26*, 445–463. [[CrossRef](#)] [[PubMed](#)]
34. Battineni, G.; Chintalapudi, N.; Amenta, F. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Inform. Med. Unlocked* **2019**, *16*, 100200. [[CrossRef](#)]
35. Goulet-Pelletier, J.-C.; Cousineau, D. A review of effect sizes and their confidence intervals, Part I: The Cohen’s *d* family. *Quant. Methods Psychol.* **2018**, *14*, 242–265. [[CrossRef](#)]
36. Horst, F.; Lapuschkin, S.; Samek, W.; Müller, K.-R.; Schöllhorn, W.I. Explaining the unique nature of individual gait patterns with deep learning. *Sci. Rep.* **2019**, *9*, 2391. [[CrossRef](#)] [[PubMed](#)]
37. Qutrio Baloch, Z.; Raza, S.A.; Pathak, R.; Marone, L.; Ali, A. Machine learning confirms nonlinear relationship between severity of peripheral arterial disease, functional limitation and symptom severity. *Diagnostics* **2020**, *10*, 515. [[CrossRef](#)] [[PubMed](#)]
38. Chicco, D.; Tötsch, N.; Jurman, G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min.* **2021**, *14*, 13. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.