

<https://doi.org/10.1038/s43856-024-00653-1>

At-home wearable-based monitoring predicts clinical measures and biological biomarkers of disease severity in Friedreich's Ataxia

Check for updates

Ram Kinker Mishra^{1,4}, Adonay S. Nunes^{1,4}, Ana Enriquez¹, Victoria R. Profeta^{2,3}, McKenzie Wells^{2,3}, David R. Lynch^{2,3,5} ✉ & Ashkan Vaziri^{1,5} ✉

Abstract

Background Friedreich ataxia (FRDA) results in progressive impairment in gait, upper extremity coordination, and speech. Currently, these symptoms are assessed through expert examination at clinical visits. Such in-clinic assessments are time-consuming, subjective, of limited sensitivity, and provide only a limited perspective of the daily disability of patients.

Methods In this study, we recruited 39 FRDA patients and remotely monitored their physical activity and upper extremity function using a set of wearable sensors for 7 consecutive days. We compared the sensor-derived metrics of lower and upper extremity function as measured during activities of daily living with FRDA clinical measures (e.g., mFARS and FA-ADL) and biological biomarkers of disease severity (guanine-adenine-adenine (GAA) and frataxin (FXN) levels), using Spearman correlation analyses.

Results The results show significant correlations with moderate to high effect sizes between multiple sensor-derived metrics and the FRDA clinical and biological outcomes. In addition, we develop multiple machine learning-based models to predict disease severity in FRDA using demographic, biological, and sensor-derived metrics. When sensor-derived metrics are included, the model performance enhances 1.5-fold and 2-fold in terms of explained variance, R^2 , for predicting FRDA clinical measures and biological biomarkers of disease severity, respectively.

Conclusions Our results establish the initial clinical validity of using wearable sensors in assessing disease severity and monitoring motor dysfunction in FRDA.

Plain language summary

Friedreich ataxia (FRDA) is a condition that impairs movement and coordination. Current clinical assessments are subjective, highlighting the need for better ways to monitor disease severity. By using wearable devices to track symptoms in everyday life, we can gain better insights into how patients function outside the clinical environment, offering a more comprehensive understanding of the disease's impact. In this study, 39 patients were observed using wearable sensors for a week to track their physical activity and arm movements. The data collected was compared with traditional clinical tests and biological markers of the disease. The findings demonstrate that wearable sensors can accurately predict disease severity, offering continuous real-world monitoring that could enhance patient care and treatment outcomes.

Friedreich ataxia (FRDA) is an autosomal recessive neurodegenerative disease caused by deficiency of the protein frataxin (FXN). The primary clinical manifestations of FRDA include progressive ataxia, cardiomyopathy, scoliosis and, in some individuals, diabetes¹⁻⁴. Ninety-six percent of causative mutations in the FXN gene consist of expanded guanine-adenine-adenine (GAA) repeats; such repeats partially silence the FXN gene and markedly decrease levels of frataxin protein². The remaining mutations include point mutations and deletions, all of which decrease levels of functional FXN². The

FXN is essential for the proper function of mitochondria and the regulation of iron-sulfur clusters, which are essential for the proper function of enzymes involved in cellular respiration and ATP production. FXN deficiency ultimately results in progressive muscle weakness, coordination problems, and difficulty with balance and movement.

The prevalence of FRDA is estimated to be about 5000 individuals in the US, and about 22,000 patients worldwide⁵. There is currently no cure for FRDA, and therapeutic focus continues to be directed at slowing disease

¹BioSensics LLC, Newton, MA, USA. ²Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ³Departments of Pediatrics and Neurology, The Children's Hospital of Philadelphia, 502F Abramson Research Center, Philadelphia, PA, USA. ⁴These authors contributed equally: Ram Kinker Mishra, Adonay S. Nunes. ⁵These authors Jointly supervised this work: David R. Lynch, Ashkan Vaziri. ✉e-mail: lynchd@mail.med.upenn.edu; ashkan.vaziri@biosensics.com

progression, managing symptoms, and optimizing quality of life. While a variety of agents have reached clinical trials, meaningful results remain difficult partially due to insensitivity of clinical outcome measures of physical function in FRDA. The standard remains the modified Friedreich's Ataxia Rating Scale (mFARS) exam, a quantified neurologic exam directed toward crucial components of disease progression in FRDA⁵⁻¹². While sufficient in some situations, the mFARS has subjective components, ceiling effects, significant day-to-day variability, and a modest sensitivity to change. Based on longitudinal natural history studies, clinical trials using mFARS as a primary outcome may require greater than 100 subjects for adequate power¹⁰. Similarly, the presently used FRDA-specific ADL instrument (Friedreich Ataxia Rating Scale Activity of Daily Living or FA-ADL) is insensitive to change and highly subjective¹³. Thus, while mFARS and other exam-based scales are adequate for establishing differences in large cohorts, novel measures might prove more sensitive as well as more relevant to daily activities.

The limited sensitivity of the mFARS has a significant impact on the outcomes of clinical trials. In several studies, therapeutic agents, which demonstrated acceptable side effect profiles, positive subjective responses from patients, and substantial mechanistic data supporting their use, fell short of achieving their mFARS-based endpoints¹⁴. Outcome measure insensitivity played a crucial role in their failure to reach potential approval from the regulatory agencies. More sensitive outcome measures are thus critical for developing novel therapies in FRDA.

Inherent lack of sensitivity is not the only weakness of the mFARS. Exam-based measures are performed on an intermittent schedule in supervised conditions in the clinic. Thus, they do not necessarily reflect real-world situations, and data is collected at a single point in time, making day-to-day fluctuation a significant issue. A previous novel approach was the use of performance measures: single quantitative tasks (timed 25-foot walk test (25FWT), the 9-hole peg test (9 HPT)) closer to events of daily living^{4,15}. Such measures capture dysfunction in FRDA, and composites from such measures have roughly equal sensitivity to change as the mFARS. Several reasons have kept such measures from being widely adopted. In longitudinal analysis, the 9 HPT has substantial trial-to-trial variability; this could be avoided by collection of a greater volume of data, a condition not readily achievable in clinic-based measures. In addition, the timed 25FWT does not directly reflect the ambulatory issue in FRDA, as FRDA patients do not change in walking speed initially, but rather in accuracy and stability.

In contrast, using home-based wearable sensors to continuously monitor motor functions in FRDA during activities of daily living can address these shortcomings. In this study, we used such measures to capture both lower and upper extremities motor dysfunction in FRDA and compared results to those measures with FRDA clinical measures (e.g., mFARS, and FA-ADL) and biological biomarkers of disease severity (GAA and FXN levels). The study aims to leverage the capabilities of wearable sensors, which allow for the collection of granular, real-time data on various aspects of physical activity and upper limb goal-directed movements (GDMs). This is critical to understanding the daily functional capacities of individuals with FRDA and minimizing reporting bias. By comparing these sensor-derived metrics with established clinical measures and biological biomarkers of disease severity, this study aims to establish the initial clinical validity of using wearable sensors in tracking disease symptoms related to FRDA. In addition, we developed machine learning models that utilize sensor-derived measures to predict clinical and biological outcomes related to the disease severity of FRDA.

In this study, we demonstrate that wearable sensor-derived metrics significantly correlate with key clinical measures, such as mFARS and FA-ADL, as well as biological markers, including GAA and FXN levels. These correlations are moderate to high in effect size, emphasizing the clinical relevance of sensor-derived physical activity and upper limb function data in FRDA. Moreover, we show that integrating these sensor-derived metrics into machine learning models improves the prediction of disease severity. These findings establish the clinical utility of wearables for real-world monitoring of FRDA symptoms and underscore their potential to refine disease assessments and support therapeutic interventions.

Methods

Study population and clinical assessments

39 ambulatory participants with FRDA (Age = 26.8 ± 1.6 years old, body-mass-index (BMI) = 22.9 ± 0.7 , average disease duration = 13.6 ± 1.1 years) were recruited at the Children's Hospital of Philadelphia/ Perelman School of Medicine at the University of Pennsylvania (Philadelphia, PA, USA) – See Table 1 for participants demographics, clinical, and biological characteristics. The inclusion criteria for the study required participants to meet several conditions. Non-ambulatory status was defined as a score of 5 or higher on the FARS disability scale. Both male and female children aged 12 years and older, as well as adults of any age, were eligible for enrollment. A genetically confirmed diagnosis of FA was required for participation, though for carrier/control cheek swab and blood samples, genetic confirmation was not necessary. Participants with a clinically or genetically confirmed diagnosis of FA, pending confirmatory genetic testing through a commercial or research laboratory, were also eligible.

Participants were excluded from the study if they were unable or unwilling to provide informed consent or had any acute or ongoing medical conditions that would interfere with the study's assessments. Additionally, if the investigator determined that the participant was unlikely or unable to comply with the study protocol, they were excluded from participation.

It is important to note that around 8 subjects were non-ambulatory and defined as those with a Friedreich's Ataxia Rating Scale (FARS) disability score of 5 or higher.

All participants over the age of 18 gave written informed consent, and the study received approval from the Institutional Review Board as a sub-study of CHOP IRB 2609, Friedreich Ataxia Clinical Outcome Measures, in accordance with the Declaration of Helsinki. Parental or guardian permission, which constituted informed consent, was required for participants under the age of 18, along with child assent where appropriate. Additionally, the study is registered on ClinicalTrials.gov under the identifier

Table 1 | Participants demographics, clinical and biological characteristics

Characteristic	FA (n = 39)
Demographics	
Age in years (Mean ± SE)	26.8 ± 1.6
Sex, Female (%)	19 (48.7%)
Ethnicity, Caucasian (%)	39 (100%)
Race, Hispanic (%)	1 (2.5%)
Height in cm (Mean ± SE)	166.9 ± 1.7
Weight in kg (Mean ± SE)	64.5 ± 2.6
BMI in kg/m ² (Mean ± SE)	22.9 ± 0.7
Right-handed, %	32 (82%)
Non-ambulatory, %	8 (20.5%)
Clinical characteristics	
Disease duration in years (Mean ± SE)	13.6 ± 1.1
mFARS score (Mean ± SE)	40.0 ± 2.0
FA-ADL score (Mean ± SE)	13.2 ± 0.9
Dominant Hand 9 HPT score (Mean ± SE)	71.8 ± 10.5
Non-Dominant Hand 9 HPT score (Mean ± SE)	58.8 ± 6.6
25FWT in seconds * (Mean ± SE)	17.6 ± 3.7
Biological characteristics	
GAA repeats (Mean ± SE)	581.4 ± 40.1
FXN (% of control individuals) * (Mean ± SE)	40.0 ± 2.0

FXN and 25FWT data marked with an asterisk (*) were available for 28 participants. mFARS modified Friedreich's Ataxia Rating Scale, FA-ADL Friedreich Ataxia Rating Scale Activity of Daily Living, 9 HPT 9-Hole Peg Test, GAA Guanine-adenine-adenine, FXN frataxin, SE Standard Error.

NCT06016946. The in-clinic assessments included the standard FRDA assessments evaluated by the neurologist, including mFARS, FA-ADL, 25FWT, and 9 HPT.

During 25FWT, participants were instructed to walk along a marked path of 25 feet, which was measured and laid out with visible start and finish lines. It was conducted in a quiet, well-lit hallway free from obstructions, ensuring a consistent environment for all participants. Each participant was asked to start from a stationary position, with the front foot placed at the starting line. On the command 'go,' the timer was started, and the participant walked towards the finish line at a comfortable and safe pace. Assistive devices were allowed if typically used by the participant in daily walking. The test was performed twice to ensure consistency, with a brief resting period in between to prevent fatigue. The time taken to complete each trial was recorded in seconds using a stopwatch. The final score was calculated as the average of the two-timed trials, providing a measure of the participant's functional walking speed. Moreover, FXN and GAA levels were sourced from the clinical charts of the participants, ensuring that the data were current and reflected their health status during the study period. The cohort presents a group of patients with moderate disease but short of loss of ambulation.

At-home monitoring of physical activity and upper limb goal-directed movements

Physical activity and the upper limb GDMs were monitored over 7 consecutive days using the PAMSys™ pendant^{16,17} and PAMSys ULM™ wrist sensor^{18,19}, respectively (BioSensics LLC, Newton, MA USA). Those sensors contain a 3-axis accelerometer (sampling frequency of 50 Hz) and built-in memory for recording long-term data. The PAMSys pendant is specifically designed to be worn around the neck at the sternum level using a lanyard and magnetic closure. This design ensures easy wearability and removal while minimizing the risk of choking. Participants were instructed to wear the pendant under their shirts to secure its placement and reduce any potential interference. Participants wore the PAMSys ULM™ wrist sensor on their dominant hand. The PAMSys™ is a patented wearable sensor (U.S. Patents # 8,206,325, 9,005,141, and 9,901,209) for precision actigraphy with sensitivity and specificity exceeding 90%^{16,20}. Additionally, it provides a broad range of sensor-derived metrics, which include features and parameters related to (1) posture, including percentage of sitting, standing, walking, and lying; (2) locomotion, including daily steps, walking bouts, and step duration variability, and cadence, and (3) postural transitions, including the number of sit-to-stand transitions and stand-to-sit transitions during activities of daily living. The sensor-derived metrics from the PAMSys ULM wrist sensor include features and parameters related to GDMs of upper extremity function including counts and duration, as well as velocity and accelerometer features. The sensors detect the time periods that they are not worn. In our analysis, we only considered days when the sensors were worn by the participant for at least 18 h. During the clinical visit, participants received instructions on using the sensors and were directed to wear them continuously.

Statistics and reproducibility

All statistical analyses were performed using standard methods in Python (SciPy and statsmodels libraries) and IBM SPSS (Version 29.0.1.0) Software. Continuous variables are presented as mean ± standard error of the mean (SEM), unless otherwise specified.

Reliability analysis. The reproducibility of our findings is supported using Intraclass Correlation Coefficients (ICCs) to assess the test-retest reliability of the sensor-derived metrics. By dividing the 7-day monitoring period into two distinct segments, we evaluated the consistency of the sensor measurements over time. The majority of metrics demonstrated high ICC values (ICC > 0.75), indicating good to excellent reliability, thus confirming that the wearable sensors provide stable and reproducible measurements across the monitoring period. ICC values were calculated using a two-way random effects model, which considers

both measurement effects and subjects as random. This model choice is suitable for data where the aim is to assess the reliability of measurements over different time periods. ICC values above 0.75 are typically interpreted as indicating good reliability, values between 0.5 and 0.75 suggest moderate reliability, and values below 0.5 indicate poor reliability²¹.

Correlation analysis. Spearman or Pearson correlations (chosen based on the scale of measurement) were conducted to demonstrate the relationship between sensor-derived metrics and clinical measures and biological biomarkers of disease severity. Furthermore, Benjamini-Hochberg (BH) correction method was applied to all correlation analyses to adjust for multiple comparisons. This method involves ranking the p-values from all tests according to their size and then adjusting these values according to their rank and the total number of comparisons. By employing this technique, we maintained the false discovery rate (FDR) at a predetermined level of 5% for this study. Consequently, only hypotheses with adjusted *p* values below this threshold were deemed statistically significant.

Machine Learning Model Development

Machine Learning Model Development. We trained multiple stochastic gradient descent regressor models to predict clinical and biological outcomes using different combinations of input features. Model 1 uses demographic information, disease duration, and GAA levels as input features. Model 2 includes only sensor-derived metrics related to physical activity as input features. Model 3 combines demographic information, disease duration, GAA levels, and sensor-derived physical activity metrics. Model 4 incorporates demographic information, disease duration, GAA levels, and sensor-derived GDM metrics. Finally, Model 5 uses a comprehensive set of input features, including demographic information, disease duration, GAA levels, sensor-derived physical activity, and GDM metrics. For the models where GAA levels are the target outcome, they are not used as input features within the same model. Instead, these predictions rely on sensor-derived metrics and demographic data.

It is important to note that when GAA levels were the target outcome for prediction, they were not used as an input feature within the same models. Instead, the predictions were made using a combination of sensor-derived metrics and demographic data. Additionally, individuals classified as non-ambulant were included in the analysis despite their limited walking capabilities. These participants could perform essential tasks such as moving short distances within the home or using the restroom. By incorporating these measures, we ensured that the analysis captured the full range of physical abilities present in the study population, allowing for a comprehensive assessment of functional mobility across all participants, regardless of their ambulatory status.

A comprehensive feature selection technique, known as a wrapper exhaustive approach, systematically assesses every conceivable combination of features to pinpoint the most suitable subset of input features for constructing a predictive model²².

Feature selection. A comprehensive wrapper exhaustive approach was implemented for feature selection. This technique initially considered all potential predictors, including demographics, disease duration, GAA levels, and sensor metrics. We systematically assessed every possible combination of these features to identify the most predictive subset. Each subset was evaluated using an SGD regressor, selected for its effectiveness in large-scale data handling and optimization.

Model validation. To assess the model's performance rigorously, we implemented a leave-one-out cross-validation strategy, which systematically evaluates the model's ability to generalize by leaving out one data point at a time from the training set and assessing its prediction accuracy²³. This approach minimizes the risk of over-fitting and ensures that our model's performance is robust and reliable.

Performance metrics. The model performance was evaluated in terms of coefficient of determination (R^2), the correlation between the prediction and actual measurement (r), and the mean absolute error (MAE). These metrics provided a comprehensive evaluation of how well each model predicted the outcomes relative to actual measured values.

Optimization and finalization. Once the optimal feature set was determined, the final model was retrained using only these selected features to finalize its configuration. This process ensures that the chosen features genuinely enhance the predictive accuracy of the model, optimizing data utilization for reliable and robust predictions.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Study Participants

31 participants were 100% compliant in wearing the sensors at all times for 7 consecutive days. The remaining 8 participants had an average non-wear duration of 72.3 ± 14.5 min, indicating an average 95% compliance. Similarly, for compliance in wearing the wrist sensor, 37 participants were 100% compliant, and the remaining 2 participants had an average compliance of approximately 90%. Figure 1 illustrates the correlations between sensor-

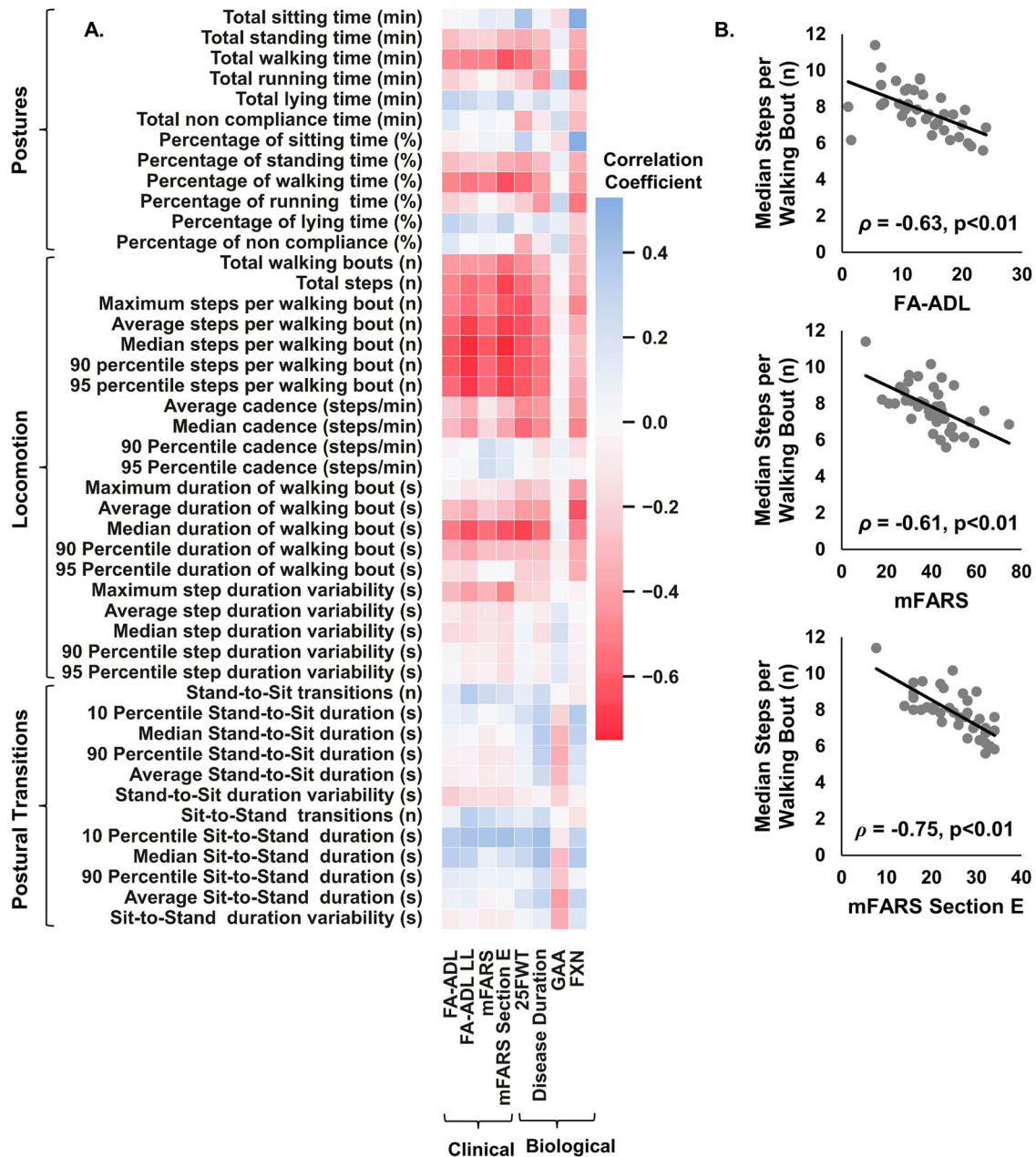


Fig. 1 | Correlations between sensor-derived physical activity metrics and clinical scores. A Correlation map displaying correlation coefficients (r or ρ) between 45 sensor-derived measures of physical activity and clinical scores of FRDA, as well as disease duration, GAA and FXN. Spearman correlation analysis was used to quantify the relationship between sensor-derived metrics and clinical scores, including FA-ADL, FA-ADL LL, mFARS, and mFARS Section E. Meanwhile, Pearson correlation

analysis was employed to evaluate the association between sensor-derived metrics and biological outcomes, such as 25FWT, GAA, FXN, and disease duration. **B** Median steps per walking bout versus FA-ADL, mFARS, and mFARS Section E. Data represent $n = 39$ participants. All physical activity metrics are averaged daily values measured over 7 consecutive days.

derived physical activity metrics measured by the PAMSys pendant with clinical outcomes (i.e., FA-ADL, FA-ADL LL, mFARS, and mFARS section E) and biological outcomes (i.e., Disease duration, GAA and FXN).

Physical activity monitoring

The correlation analysis of sensor-derived physical activity metrics with clinical scores and biological outcomes (i.e., GAA and FXN) revealed several significant relationships, Supplementary Data 1. Notably, the percentage of sitting time was significantly positively correlated with FXN ($r = 0.64$, $p = 0.002$), indicating that higher sitting time is associated with higher FXN levels. Conversely, the percentage of standing time showed a moderate negative correlation with FXN ($r = -0.47$, $p = 0.064$), although this did not reach significance.

Locomotion metrics demonstrated strong negative correlations with clinical scores, particularly FA-ADL, FA-ADL LL, mFARS, and mFARS Section E. Total walking time (ρ ranging from -0.46 to -0.63), total steps (ρ ranging from -0.48 to -0.67), and various step bout metrics, such as average, median, and 90th percentile steps per bout, showed significant negative correlations with these clinical scores. This indicates that higher walking activity is associated with lower clinical impairment scores. Specifically, median steps per walking bout correlated strongly with FA-ADL ($\rho = -0.63$, $p < 0.001$), FA-ADL LL ($\rho = -0.74$, $p < 0.001$), mFARS ($\rho = -0.61$, $p < 0.001$), and mFARS Section E ($\rho = -0.75$, $p < 0.001$).

Postural transition metrics generally showed weak correlations with clinical scores, GAA, and FXN. However, the 10th percentile sit-to-stand duration showed a moderate positive correlation with mFARS Section E ($\rho = 0.42$, $p = 0.040$), suggesting that faster transition times might be associated with better clinical outcomes in specific motor functions. Additionally, 28 participants were able to perform 25FWT. The median step count per walking bout as measured by the PAMSys pendant demonstrated a significant correlation with the 25FWT, underscoring the relevance of this physical activity measure in reflecting walking capabilities in FRDA patients.

Goal-directed movements

The correlation analysis of sensor-derived GDM metrics with biological outcomes and clinical scores also revealed several significant relationships, see Fig. 2 and Supplementary Data 2. The velocity features of GDMs showed strong correlations with clinical scores. Median velocity was significantly negatively correlated with mFARS ($\rho = -0.56$, $p = 0.004$) and 9-HPT ($\rho = -0.60$, $p = 0.001$), suggesting that lower median velocities are associated with higher impairment. Maximum velocity and velocity root mean squared also showed similar negative correlations with these clinical scores. Conversely, velocity dominant frequency was positively correlated with mFARS ($\rho = 0.60$, $p = 0.001$) and 9-HPT ($\rho = 0.66$, $p < 0.001$), indicating that higher dominant frequencies

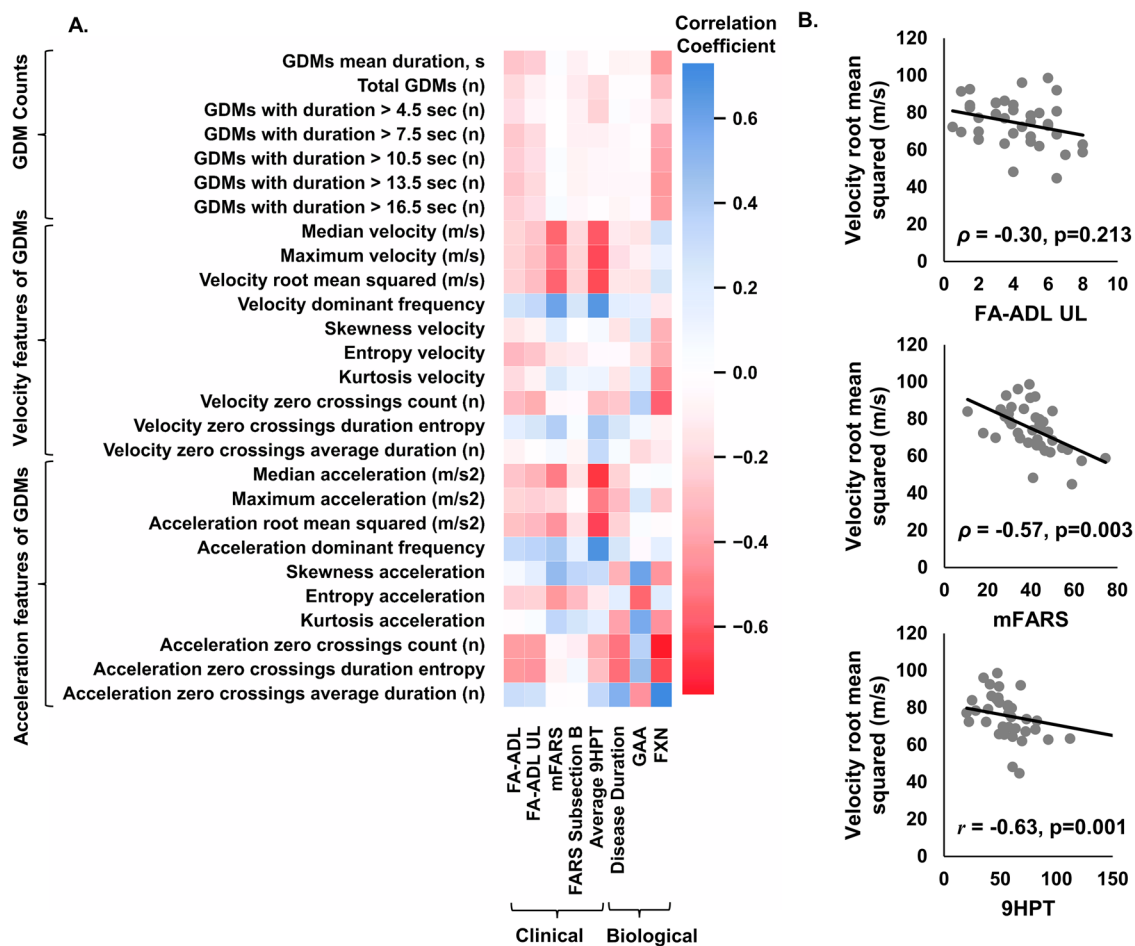


Fig. 2 | Correlations between sensor-derived GDM and clinical scores and biological biomarkers. A Correlation map displaying correlation coefficients (r or ρ) between 27 sensor-derived measures of GDM and clinical scores of FRDA, as well as disease duration, GAA and FXN. Spearman correlation analysis was used to quantify the relationship between sensor-derived metrics and clinical scores, including FA-ADL, FA-ADL UL, mFARS, and mFARS Subsection B. Meanwhile, Pearson

correlation analysis was employed to evaluate the association between sensor-derived metrics and biological outcomes, such as GAA, FXN, average HPT, and disease duration. B Velocity root mean squared of GDMs versus FA-ADL, mFARS, and 9 HPT. Data represent $n = 39$ participants. All GDM metrics are averaged daily values measured over 7 consecutive days.

are associated with lower impairment. Acceleration features also revealed significant correlations. Median acceleration was negatively correlated with mFARS ($\rho = -0.50$, $p = 0.013$) and 9-HPT ($\rho = -0.68$, $p < 0.001$), indicating that higher impairments are associated with lower median accelerations. Skewness acceleration had a significant positive correlation with mFARS ($\rho = 0.49$, $p = 0.016$) and a significant negative correlation with FXN ($\rho = -0.59$, $p = 0.010$), suggesting that higher skewness is associated with greater clinical impairment and lower FXN levels. Additionally, kurtosis acceleration was positively correlated with GAA ($r = 0.59$, $p = 0.004$) but negatively correlated with FXN ($r = -0.58$, $p = 0.013$). Other significant findings include the negative correlations of entropy acceleration with GAA ($r = -0.53$, $p = 0.013$) and positive correlations of kurtosis acceleration with GAA ($r = 0.59$, $p = 0.004$). Acceleration zero crossings count was significantly negatively correlated with FXN ($r = -0.71$, $p = 0.001$), while acceleration zero crossings duration entropy showed significant negative correlations with FXN ($r = -0.73$, $p < 0.001$) and significant positive correlations with GAA ($r = 0.44$, $p = 0.062$). Overall, the analysis indicates that several GDM metrics, particularly those related to velocity and acceleration features, are significantly correlated with clinical scores and FXN levels. GAA, on the other hand, showed mixed correlations, with some metrics, such as entropy acceleration and kurtosis acceleration, exhibiting significant relationships.

Reliability assessment

The physical activity metrics were reported under three categories: (1) postures, (2) locomotion, and (3) postural transitions. Our analysis of sensor-derived physical activity metrics demonstrated generally high reliability across various measures of posture and locomotion as reported in Supplementary Data 3. The Intraclass Correlation Coefficients (ICCs) for the majority of the metrics, such as total sitting, standing, and walking times, as well as the number of walking bouts and total steps, consistently indicated good to excellent reliability ($ICC > 0.80$). This suggests that the wearable sensors used in the study provided stable and consistent measurements across the two segments of the 7-day monitoring period. Notably, the metrics for walking activities, including total walking time and steps, exhibited particularly high reliability, with ICCs nearing or exceeding 0.96. This high reliability extends to metrics capturing walking bout characteristics and cadence, where most measures showed ICCs above 0.90, highlighting the sensors' capability to accurately track walking parameters (i.e., sensor-derived measures of locomotion). Additionally, the summary of sensor-derived metrics for the GDM is provided in the Supplementary Data 4.

Prediction models

Supplementary Data 5 summarizes result related model performance. There is a noticeable trend where models incorporating a broader range of input features (Models 3, 4, and 5) generally outperform those with more limited inputs (Models 1 and 2) across most metrics. This suggests that the integration of both types of sensor data with demographic and disease information leads to more accurate predictions. Model 5 shows superior performance in predicting GAA and FXN levels, with R^2 values of 0.77 and 0.81, respectively, demonstrating the value of multimodal approach-based comprehensive feature set. Model 5 also excels in predicting mFARS ($R^2 = 0.73$) and mFARS Section E ($R^2 = 0.85$), indicating excellent predictive reliability and validity for these clinical measures.

Models using only sensor-derived metrics related to physical activity (Model 2) were particularly effective in predicting mFARS Section E and FA-ADL LL with relatively high R^2 values, highlighting the importance of activity data in assessing motor functions. Moreover, the inclusion of GDM metrics significantly enhanced the accuracy for GAA and FXN predictions in Model 4, showing R^2 values of 0.74 for both, pointing to the relevance of goal-directed movements in understanding biological variations. The performance metrics across the models indicate that the complexity and diversity of input features directly contribute to the robustness of the model

outputs. Notably, the comprehensive inputs in Model 5 lead to the most consistent high performance across nearly all outcome measures, establishing it as the most effective model configuration in this study. Features included after feature selection process are reported in the Supplementary Data 6.

Discussion

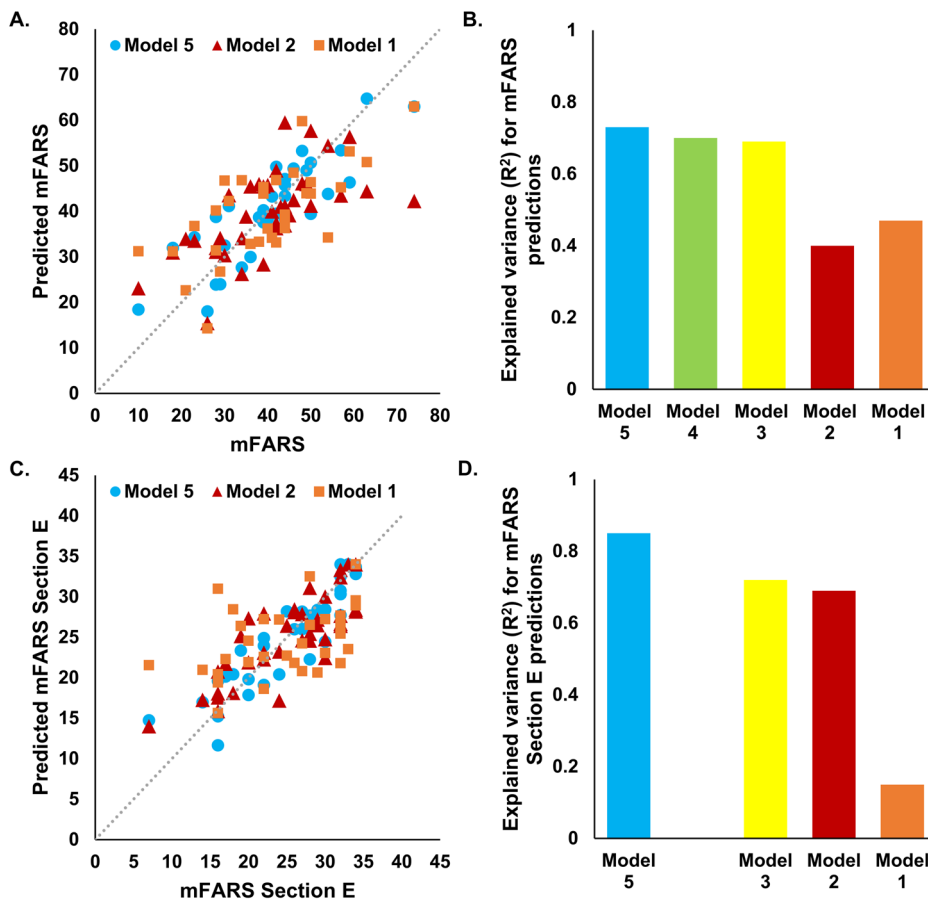
In the present study, wearable sensors for at-home monitoring and capturing motor function were used in FRDA. Sensor-derived metrics correlated with the most commonly used clinical outcomes (i.e., FA-ADL, FA-ADL UL, mFARS, and mFARS section B, Average 9 HPT) and biological outcomes (i.e., GAA and FXN). Our findings also indicate that sensor-derived measures of physical activity and upper limb function can effectively support the implementation of a data-driven approach based on machine learning for predicting disease severity. When combined with demographic and biological assessments, these sensor-derived metrics offer a comprehensive and more accurate understanding of the disease symptoms. Moreover, the diverse sub-sections of the mFARS assessment focus on specific symptoms of FRDA since it affects multiple systems in the body. Our results showed that machine learning models could predict subsections, including FA-ADL UL, FA-ADL LL, and mFARS section E. This is significant because the models specifically developed to predict the extent of impairment in each area would enable a more targeted approach to treatment and rehabilitation. Lastly, as indicated in our data (Supplementary Fig. S1), we have validated the findings of prior research, which demonstrated a correlation between the age of onset and disease severity, as well as a shorter GAA-repeat length^{24–26}.

We observed a high level of compliance with wearable sensors for remote monitoring of physical activity and upper limb function. This signifies that FRDA patients are consistently and effectively using these devices and at-home monitoring of their physical activity and movement patterns using wearable sensors is feasible in FRDA.

Recent studies have yielded encouraging outcomes by using wearable sensors to quantify clinical evaluations such as SARA and mFARS, aiming to mitigate the inaccuracies associated with subjective clinical assessments in ataxia^{27–30}. However, these approaches still require supervised assessments, which is only feasible within clinical settings. Our study findings shares similarities with the recent work by Fichera et al. (2024), which highlights the versatility and potential of wearable sensors in enhancing our understanding of FRDA's impact on patient mobility and daily life activities³¹. While our study explores a broader range of sensor-based metrics, the reviewed study specifically focuses on detailed activity classifications like sedentary, light, and Moderate-Vigorous Physical Activity (MVPA), and uses Vector Magnitude (VM3) as a composite measure of movement intensity across three axes. Specifically, Fichera et al. (2024) demonstrated that VM3, which quantifies the total movement in three-dimensional space, is significantly correlated with clinical scales such as the Scale for the Assessment and Rating of Ataxia (SARA) and the mFARS. This aligns with our findings, which revealed significant negative correlations between locomotion-related metrics and disease severity, underscoring the relevance of three-dimensional movement analysis in capturing the functional impairments associated with FRDA.

Using wearables is of paramount importance due to the potential discrepancy between a brief examination in an outpatient clinic, where an individual with ataxia may appear to walk and maintain balance better than what caregivers report witnessing in their daily lives. Furthermore, relying on a single, or infrequently administered, mobility assessment cannot adequately capture day-to-day variations or other clinically pertinent periods of change, such as daily fluctuations in motor function or the influence of fatigue²⁴. In a large natural history study in FRDA ($n = 812$), neurological outcomes including the 25FWT showed a linear progression of the disease over a 5-year study period. The significant correlation observed between the median step count per walking bout and the 25FWT underscores the utility

Fig. 3 | Machine learning-based models for predicting mFARS. **A** mFARS predictions from Models 1 ($n = 33$), 2 ($n = 39$), and 5 ($n = 32$) versus the clinical scores. **B** The explained variance for mFARS prediction for models 1-5. **C** mFARS Section E predictions from Models 1 ($n = 33$), 2 ($n = 39$), and 5 ($n = 33$) versus the clinical scores. **D** The explained variance for mFARS Section E prediction for models 1, 2, 3, and 5. Model 4 was not trained for the prediction mFARS Section E as it uses sensor-derived GDM metrics.



of this physical activity measure in reflecting walking capabilities in patients with FRDA.

A challenge in accurately quantifying real-life physical activity lies in the need to consider environmental factors and the context of walking. For instance, the total number of steps recorded in a day can be influenced by longer walks outdoors and may not accurately represent shorter walking bouts that occur indoors. Furthermore, number of postural transitions may provide another insight about physical activity as a frail individual may lack of lower-extremity strength and perform lower postural transitions³². Therefore, different environments and walking contexts can have varying impacts on a person’s overall physical activity levels. Considering these factors, we included parameters such as different percentiles (e.g., 50th percentile) of steps per walking bout, cadence, step variability, as well as overall time percentage spent in each posture in our analysis. We observed the highest correlations (Supplementary Data 1) for median steps per walking bout.

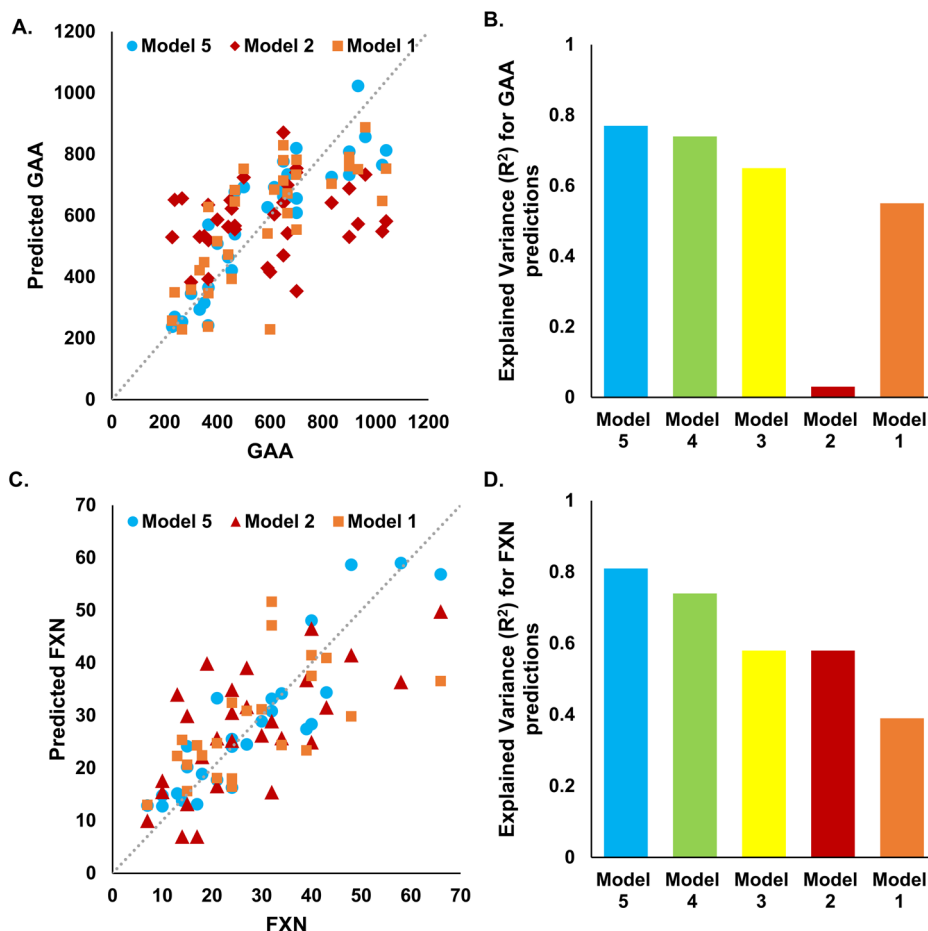
Furthermore, negative correlation of root mean squared velocity and median acceleration as measured by the PAMSys ULM wrist sensor with the 9HPT and clinical scores (i.e., mFARS) may indicate a decline in motor function as the disease progresses, leading to slower hand movements and reduced acceleration during GDMs. Further investigation would be required to determine whether the decrease in velocity and acceleration during task performance represents a compensatory strategy to accommodate for coordination deficits or if it is a consequence of muscle weakness.

We developed various machine learning models to leverage demographics information, disease duration, and GAA, as well as both sensor-derived physical activity and GDM metrics. In Model 1, which used demographics information, disease duration, and GAA as the input features, the predictive performance yielded R² values of 0.47 for mFARS and 0.39 for FXN (see Supplementary Data 5, and

Figs. 3 and 4). The most significant improvement was observed in Model 5, where the inclusion of sensor-derived physical activity and GDM metrics alongside the input features from Model 1 resulted in an increase in R² values to 0.73 for mFARS and 0.81 for FXN. This marked a 1.5-fold improvement in model performance for mFARS and a 2-fold improvement for FXN compared to Model 1. Model 2, which exclusively utilized sensor-derived physical activity metrics as input features, achieved an R² value of 0.42 for FXN. A related study conducted by Kadirvelu et al.^{27–30} reported similar predictive results. Their models, which incorporated sensor-derived metrics for the 8-minute walk and 9HPT tests, achieved R² values of 0.59 and 0.53, respectively. Notably, when demographics information, disease duration, and GAA data were included as input features in Model 3, along with physical activity metrics, and GDM metrics in Model 4, the R² values exhibited further improvement. These findings highlight the potential advantages of a multimodal approach to disease assessment and its role in enhancing the prediction of disease severity.

While our study offers encouraging insights, it is crucial to recognize several significant limitations. Primary limitation is the absence of longitudinal data for the participants in this study. Consequently, our results do not provide a comprehensive representation of the disease’s progression and its dynamic changes over time. To mitigate these limitations, future research initiatives should focus on validating our technology and methodology through a longitudinal study involving a broader and more diverse participant cohort, including individuals with different ambulatory statuses. Another limitation of our study is the absence of comprehensive neurological or clinical assessments specifically designed to test for lower velocity in motor tasks. Without these detailed evaluations, we are unable to conclusively determine whether observed reductions in velocity are due to muscle weakness or compensatory strategies for impaired motor function.

Fig. 4 | Machine learning-based models for predicting biological biomarkers. A GAA predictions from Models 1 ($n = 33$), 2 ($n = 34$), and 5 ($n = 33$) versus the actual GAA. B The explained variance for GAA prediction for models 1-5. C FXN predictions from Models 1 ($n = 23$), 2 ($n = 28$), and 5 ($n = 26$) versus the actual FXN. D The explained variance for FXN prediction for models 1-5.



Data availability

The de-identified data supporting the findings of this study are available as Supplementary Data File 8. The numerical data (source data) used to plot Figs. 1b, 2b, 3a, 3c, 4a, and 4c are included in Supplementary Data file 7.

Received: 12 January 2024; Accepted: 22 October 2024;

Published online: 29 October 2024

References

- Lynch, D. R., Schadt, K., Kichula, E., McCormack, S. & Lin, K. Y. Friedreich ataxia: multidisciplinary clinical care. *J. Multidiscip. Healthcare* **14**, 1645–1658 (2021).
- Strawser, C. et al. Pharmacological therapeutics in Friedreich ataxia: the present state. *Expert Rev. Neurotherapeutics* **17**, 895–907 (2017).
- Harding, I. H., Lynch, D. R., Koeppen, A. H. & Pandolfo, M. Central nervous system therapeutic targets in Friedreich ataxia. *Hum. Gene Ther.* **31**, 1226–1236 (2020).
- Lynch, D. R. et al. Measuring Friedreich ataxia: complementary features of examination and performance measures. *Neurology* **66**, 1711–1716 (2006).
- Lynch, D. R. et al. Efficacy of Omaveloxolone in Friedreich’s Ataxia: Delayed-Start Analysis of the MOXle Extension. *Mov. Disord.* **38**, 313–320 (2023).
- Rummey, C. et al. Harmonizing results of ataxia rating scales: mFARS, ARA, and ICARS. *Ann. Clin. Transl. Neurol.* **9**, 2041–2046 (2022).
- Rummey, C. et al. Natural History of Friedreich Ataxia: heterogeneity of neurologic progression and consequences for clinical trial design. *Neurology* **99**, e1499–e1510 (2022).
- Lynch, D. R. et al. Safety and efficacy of omaveloxolone in Friedreich ataxia (MOXle study). *Ann. Neurol.* **89**, 212–225 (2021).
- Rummey, C. et al. Test–retest reliability of the Friedreich’s ataxia rating scale. *Ann. Clin. Transl. Neurol.* **7**, 1708–1712 (2020).
- Rummey, C. et al. Psychometric properties of the Friedreich ataxia rating scale. *Neurol. Genet.* **5**, 371 (2019).
- Qureshi, M. Y. et al. Safety and efficacy of (+)-epicatechin in subjects with Friedreich’s ataxia: A phase II, open-label, prospective study. *J. Inherit. Metab. Dis.* **44**, 502–514 (2021).
- Lynch, D. R. et al. Safety, pharmacodynamics, and potential benefit of omaveloxolone in Friedreich ataxia. *Ann. Clin. Transl. Neurol.* **6**, 15–26 (2019).
- Perez-Lloret, S. et al. Assessment of ataxia rating scales and cerebellar functional tests: critique and recommendations. *Mov. Disord.* **36**, 283–297 (2021).
- Lynch, D. R. et al. Randomized, double-blind, placebo-controlled study of interferon- γ 1b in Friedreich Ataxia. *Ann. Clin. Transl. Neurol.* **6**, 546–553 (2019).
- Friedman, L. S. et al. Measuring the rate of progression in Friedreich ataxia: implications for clinical trial design. *Mov. Disord.* **25**, 426–432 (2010).
- Paraschiv-Ionescu, A., Buchser, E., Rutschmann, B., Najafi, B. & Aminian, K. Ambulatory system for the quantitative and qualitative analysis of gait and posture in chronic pain patients treated with spinal cord stimulation. *Gait Posture* **20**, 113–125 (2004).
- Mishra, R. K. et al. Decrease in Mobility during the COVID-19 Pandemic and its association with increase in depression among older adults: a longitudinal remote mobility monitoring using a wearable sensor. *Sensors* **21** <https://doi.org/10.3390/s21093090> (2021).

18. Lee, S. I. et al. Enabling stroke rehabilitation in home and community settings: a wearable sensor-based approach for upper-limb motor training. *IEEE J. Transl. Eng. Health Med.* **6**, 1–11 (2018).
 19. Adonay, S. N., İlkay, Y. P., Mishra, R. K., Paolo, B. & Vaziri, A. A deep learning wearable-based solution for continuous at-home monitoring of upper limb goal-directed movements. *Front. Neurol. Sect. Mov. Disord.* **ume 14**, 2023 (2023).
 20. Najafi, B., Armstrong, D. G. & Mohler, J. Novel wearable technology for assessing spontaneous daily physical activity and risk of falling in older adults with diabetes. *J. Diab. Sci. Technol.* **7**, 1147–1160 (2013).
 21. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
 22. Jović, A., Brkić, K. & Bogunović, N. A review of feature selection methods with applications. In *38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*. 1200–1205 (IEEE, 2015).
 23. Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction* (Springer, 2009).
 24. Ilg, W. et al. Quantitative gait and balance outcomes for ataxia trials: consensus recommendations by the Ataxia Global Initiative Working Group on Digital-Motor Biomarkers. *Cerebellum* **23**, 1566–1592 (2023).
 25. Mueller, A. et al. Digital endpoints for self-administered home-based functional assessment in pediatric Friedreich’s ataxia. *Ann. Clin. Transl. Neurol.* **8**, 1845–1856 (2021).
 26. Koeppen, A. H. Friedreich’s ataxia: pathology, pathogenesis, and molecular genetics. *J. Neurol. Sci.* **303**, 1–12 (2011).
 27. Kadirvelu, B. et al. A wearable motion capture suit and machine learning predict disease progression in Friedreich’s ataxia. *Nat. Med.* **29**, 86–94 (2023).
 28. Kashyap, B. et al. Objective assessment of cerebellar ataxia: a comprehensive and refined approach. *Sci. Rep.* **10**, 9493 (2020).
 29. Mohammadi-Ghazi, R. et al. Objective assessment of upper-extremity motor functions in spinocerebellar ataxia using wearable sensors. *Sensors* **22**, 7993 (2022).
 30. Zhou, H. et al. Assessment of gait and balance impairment in people with spinocerebellar ataxia using wearable sensors. *Neurol. Sci.* **43**, 2589–2599 (2022).
 31. Fichera, M. et al. Accelerometer-based measures in Friedreich ataxia: a longitudinal study on real-life activity. *Front. Pharm.* **15**, 1342965 (2024).
 32. Parvaneh, S., Mohler, J., Toosizadeh, N., Grewal, G. S. & Najafi, B. Postural transitions during activities of daily living could identify frailty status: application of wearable technology to identify frailty during unsupervised condition. *Gerontology* **63**, 479–487 (2017).
- analysis. R.M., A.S.N., D.R.L., and A.V. were responsible for result interpretation. A.E., V.P., and M.W. were responsible for clinical operations and data management. V.R.P., M.W., and D.R.L. were responsible for patient recruitment, and clinical data collection. D.L. and A.V. contributed to the study concept and design and securing funding.

Competing interests

At the time of performing the study, R.K.M., A.S.N., A.E., and A.V. were employees of BioSensics LLC, the developer and manufacturer of the wearable sensors used in this study. D.R.L. receives funding from the National Institute of Health, Friedrich’s Ataxia Research Alliance (FARA), Reata Pharmaceuticals Inc., PTC Therapeutics Inc., and Design Therapeutics. V.R.P. and M.W. declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-024-00653-1>.

Correspondence and requests for materials should be addressed to David R. Lynch or Ashkan Vaziri.

Peer review information *Communications Medicine* thanks Winfried Ilg, Tuhin Virmani and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

Author contributions

R.K.M., D.R.L., and A.V. drafted the original manuscript. R.K.M. and A.S.N. were responsible for developing the machine learning models and statistical