scientific reports

OPEN

Predictive machine learning and multimodal data to develop highly sensitive, composite biomarkers of disease progression in Friedreich ataxia

Susmita Saha[1,2], Louise A. Corben[1,3,4], Louisa P. Selvadurai[1], Ian H. Harding[2,5,6] & Nellie Georgiou-Karistianis[1,6]✉

Friedreich ataxia (FRDA) is a rare, inherited progressive movement disorder for which there is currently no cure. The field urgently requires more sensitive, objective, and clinically relevant biomarkers to enhance the evaluation of treatment efficacy in clinical trials and to speed up the process of drug discovery. This study pioneers the development of clinically relevant, multidomain, fully objective composite biomarkers of disease severity and progression, using multimodal neuroimaging and background data (i.e., demographic, disease history, genetics). Data from 31 individuals with FRDA and 31 controls from a longitudinal multimodal natural history study IMAGE-FRDA, were included. Using an elasticnet predictive machine learning (ML) regression model, we derived a weighted combination of background, structural MRI, diffusion MRI, and quantitative susceptibility imaging (QSM) measures that predicted Friedreich ataxia rating scale (FARS) with high accuracy (R^2 = 0.79, root mean square error (RMSE) = 13.19). This composite also exhibited strong sensitivity to disease progression over two years (Cohen's d = 1.12), outperforming the sensitivity of the FARS score alone (d = 0.88). The approach was validated using the Scale for the assessment and rating of ataxia (SARA), demonstrating the potential and robustness of ML-derived composites to surpass individual biomarkers and act as complementary or surrogate markers of disease severity and progression. However, further validation, refinement, and the integration of additional data modalities will open up new opportunities for translating these biomarkers into clinical practice and clinical trials for FRDA, as well as other rare neurodegenerative diseases.

Keywords Friedreich ataxia, Machine learning, Multimodal data, Composite biomarkers, Disease progression, Neuroimaging

Friedreich ataxia (FRDA) is a rare inherited multi-system disease. The neurological component of FRDA is defined by abnormalities and degeneration of the spinal cord, cerebellum, and cerebro-cerebellar pathways, leading to significant and progressive balance, gait, upper limb coordination, speech, and swallowing deficits. FRDA poses a significant challenge in the medical field due to its debilitating nature, lack of effective therapies, and the absence of a cure[1,2]. While a recent milestone was achieved with the approval of the first disease-modifying drug for individuals with FRDA, further drug development efforts are necessary[3,4]. A significant obstacle in advancing drug development and clinical trials for FRDA lies in the scarcity of sensitive, reliable and comprehensive biomarkers capable of capturing disease progression over typical clinical trial timeframes (i.e., 24–72 weeks)[5–7]. Meeting this need is particularly challenging in FRDA due to the slow rate of symptom

[1]School of Psychological Sciences, The Turner Institute for Brain and Mental Health, Monash University, 18 innovation walk, Clayton campus, Clayton, Victoria, Australia. [2]Department of Neuroscience, School of Translational Medicine, Monash University, Melbourne, Australia. [3]Present address: Bruce Lefroy Centre for Genetic Health Research, Murdoch Children's Research Institute, Parkville, Victoria, Australia. [4]Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia. [5]QIMR Berghofer Medical Research Institute, Brisbane, Australia. [6]Ian H. Harding and Nellie Georgiou-Karistianis jointly supervised this work. ✉email: nellie.georgiou-karistianis@monash.edu

advancement, rarity of the disease, and heterogeneity in symptoms between individuals and across the disease course[8]. Clinician-rated outcome measures, which are currently considered the reference standard, often suffer from high variability due to subjective assessment[9,10]. The modified Friedreich Ataxia Rating Scale (mFARS) is currently the accepted clinician-rated outcome measure for FRDA trials[11]. However, a recent systematic review graded the evidence supporting the efficacy of FARS scores (including both FARS and mFARS) in assessing therapeutic effects in FRDA as very low-quality[12]. The positive evidence is primarily driven by the omaveloxolone study, which demonstrated a significant improvement in its primary outcome measure, mFARS[6,11]. While this may be influenced by factors such as the short trial duration, the sensitivity limitations of FARS alone, and the slow progression of FRDA, further validation of mFARS efficacy remains essential. In contrast, the most recent drug trial, vatiquinone, for FRDA, found no significant difference in the mean placebo-corrected change in mFARS from baseline over the 72-week core trial period[13]. Additionally, studies also indicated modest and short-lived placebo responses in mFARS, a lack of correlation between mFARS changes and other disease-related endpoints, and its poor reliability and/or sensitivity in some cohorts such as non-ambulatory individuals with FRDA[11,14]. Overall, there remains a critical need to validate alternative outcome measures to serve as complementary or surrogate clinical trial endpoints. In fact, the systematic review[12] suggested that combining multiple biomarkers offers a more comprehensive assessment of treatment efficacy, which is particularly crucial for slowly progressive diseases like FRDA, where subtle yet meaningful treatment effects may only emerge through integrated evaluation.

In response to this pressing need, a number of global initiatives are underway to identify objective biomarkers capable of offering greater sensitivity in tracking FRDA disease progression. Longitudinal studies such as IMAGE-FRDA[15], TRACK-FA[16], CAMPINAS[17] and MINNESOTA[18] were dedicated to discovering multimodal neuroimaging biomarkers using brain and spinal cord MRI and spectroscopy across demographically diverse participant groups. In addition, automated measures of movement and speech using digital devices and recordings, wearables, and signal processing techniques have also demonstrated superior performance in diagnosis, symptom grading, and tracking disease progression compared to conventional clinical scales[19–24]. Measures derived from biofluids and genetic assays, including frataxin level, epigenetic silencing and methylation of the FXN gene, as well as proteomic and metabolomic discovery approaches also offer promise for identifying new pharmacodynamic, monitoring and prognostic biomarkers[25]. These developments herald an exciting era for biomarker development in FRDA.

However, medical researchers accept that tracking disease progression using a single biomarker may not provide sufficient accuracy, as evidenced by studies across various diseases[26]. Consequently, it is becoming increasingly common that multiple biomarker tests are performed on each individual, and the corresponding measurements are combined into a single score to help clinicians make better diagnostic, prognostic, and disease monitoring judgements. Studies have shown that combining multiple biomarkers to measure treatment outcome after randomised clinical trials performed better than a single biomarker assessment approach in diseases like chronic fatigue syndrome or major depression[27]. In FRDA, composite biomarkers may similarly offer a more comprehensive understanding of disease pathology and enhanced sensitivity to disease progression[28,29]. However, it is important to note that while combining multiple biomarkers may improve diagnostic accuracy or treatment efficacy in a statistical sense, the clinical relevance of such combinations remains a critical question. In addition, a later study by Rummey and colleagues[30] suggested that composites are most effective when studies include a wider cohort of patients at different disease stages, as certain components may only reflect changes in specific phases. Therefore, incorporating disease-phase-defining elements into the composites, or developing personalised composite biomarkers tailored to an individual's unique biological characteristics and disease history, could be particularly advantageous in cohort studies and trials for rare and heterogeneous conditions like FRDA.

Advanced mathematical models can play a pivotal role in identifying biomarkers relevant to disease mechanisms, particularly in the context of large multivariate data[31]. The development of composite, multidimensional, individually-relevant biomarkers using pattern recognition algorithms and artificial intelligence is an emerging trend in health research[32,33]. Machine learning (ML) methods could provide several advantages over traditional statistical approaches in the development of composite biomarkers, including greater ability to handle complex multidimensional datasets without stringent distributional assumptions and making predictions without explicit programming at each step[34]. Recent ML applications in biomarker discovery across various diseases have utilised techniques such as random forest, lasso, and ensemble methods[19,22,35,36]. These examples highlight the potential for further adoption of ML in composite biomarker development studies.

This study seeks to investigate the effectiveness of ML techniques in developing clinically significant, fully objective, multi-domain, personalised, and highly sensitive composite biomarkers for monitoring the progression of FRDA. By utilising multimodal neuroimaging data (including structural, diffusion, and susceptibility imaging) along with background and clinical information (i.e., demographics, genetics, and disease history) from the longitudinal IMAGE-FRDA study[15,37,38], the goal is to determine the optimal combination and weighting of individual biomarker components. Ultimately, the goal is to develop a composite biomarker that surpasses traditional clinical scores by providing a more comprehensive, accurate, and sensitive representation of disease-related changes over time.

Methods

Participants

Data were collected as part of IMAGE-FRDA, a single-site longitudinal neuroimaging study conducted at Monash University, Monash Health and the Murdoch Children's Research Institute in Melbourne, Australia from 2013 to 2016. Initially, the study enrolled 37 individuals with FRDA and 37 control participants. However, for this analysis, we included 31 individuals with FRDA and 31 controls who completed two assessments approximately

two years apart (mean = 2.02 years, SD = 0.18 years; see Table 1) and had available demographic information and clinical scores at both time points. The remaining six participants from each group had incomplete imaging data at visit 1 and entirely missing data at visit 2. Individuals with FRDA had confirmed homozygosity for a GAA repeat expansion in intron 1 of *FXN*, with no clinically reported comorbid neurological or psychiatric diagnoses. All participants provided informed consent, and the study was approved by the Monash Health Human Research Ethics Committee (13201B) and carried out in accordance with the Declaration of Helsinki[39]. Disease severity was assessed using the original FARS (scored from 0 to 167; higher scores indicate greater severity) and Scale for the Assessment and Rating of Ataxia (SARA) (0–40).

Image acquisition

MRIs were acquired using a 3T Siemens Skyra (Siemens, Erlangen, Germany) with a 32-channel head coil (Monash Biomedical Imaging, Melbourne). Neuroimaging modalities included T1-weighted magnetization-prepared rapid gradient echo (MPRAGE), T2*-weighted dual-echo gradient-recalled echo (GRE), diffusion-weighted, and T2-weighted images with and without magnetisation transfer pulse. The details of image acquisition protocols are provided in previous studies[15,38,40]. Whole-brain T1-weighted images were acquired over 4 min and 26s with 176 sagittal slices, voxel size $1 \times 1 \times 1$ mm^3, matrix size = 256×256, field of view = 256×256 mm^2, echo time = 2.55 ms, and repetition time = 1,540 ms. Whole-brain diffusion-weighted images were acquired over1 3 min and 35 s using an echo-planar spin-echo sequence, with b = 3,000, 64 gradient directions, 60 axial slices, voxel size = $2.1 \times 2.1 \times 2.1$ mm^3, matrix size = 122×122, field of view = 256×256 mm^2, echo time = 108 ms, and repetition time = 11.6 s. The parameters for the GRE acquisition were: acquisition time, 11.5 min; repetition time, 30 milliseconds; echo time, 7.38 and 22.14 milliseconds; flip angle, 15; field of view, 230×230 mm^2; voxels, 0.9 mm isotropic; 160 axial slices.

Several participants' structural images were excluded due to poor quality, as identified through visual inspection. Specifically, for visit1, exclusions included one control and one FRDA participant, while for visit2, exclusions included two control and two FRDA participants.

Variables

In this study, a multimodal suite of variables was utilised (Table 2). These included:

(i) "Background" demographic, disease history and genetic characteristics, including age at scan, age at symptom onset, sex, disease duration, length of the expanded genetic GAA triplet repeat in *FXN* (shorter repeat = GAA1, longer repeat = GAA2).
(ii) Clinical severity, with the FARS used as the primary measure of disease severity in this paper, and the SARA used for external validation.
(iii) Regional brain volume from structural MRI in 11 regions of interest in the cerebellum and brainstem.
(iv) Diffusion tensor imaging (DTI) metrics of fractional anisotropy (fa), mean diffusivity (md), axial diffusivity (ad), and radial diffusivity (rd) in each of the three cerebellar peduncles; and.
(v) Volume and mean magnetic susceptibility of the left and right dentate nuclei derived using quantitative susceptibility mapping (QSM).

All MRI measures used in this study were derived using methods and outcomes described in our prior investigations[15,37,38,40–42]. For the volumetric measures of the cerebellum, we aggregated the original set of 28 lobule level parcellations into 5 lobe-level parcels to reduce data dimensionality, also as previously described[43].

Based on an initial exploration of FARS and SARA scores using linear models, disease duration was divided into two distinct variables: disease_duration_to_10 and disease_duration_10_up, to better capture changes in disease severity. As shown in supplementary Figure S1, this piecewise approach introduces a bend at the 10-year

Variables	Control ($n = 31$)	FRDA ($n = 31$)
Age at v1(yrs)	38.36 ± 13.33	37.01 ± 13.57
Age at v2 (yrs)	40.44 ± 13.35	38.98 ± 13.58
Age at disease onset (yrs)		19.55 ± 8.77
Disease duration at v1(yrs)		17.34 ± 10.65
Disease duration at v2(yrs)		19.45 ± 10.73
GAA1 repeat length		538 ± 227
GAA2 repeat length		867 ± 239
sex (M/F)	17/14	(19/12)
FARS at v1		79.52 ± 28.37
FARS at v2		86.25 ± 28.08
SARA at v1		18.92 ± 8.65
SARA at v2		20.83 ± 9.13

Table 1. Descriptive statistics of $n = 31$ control and $n = 31$ participants with FRDA. Mean ± *sd*. GAA = number of GAA repeats in intron 1 of the *FXN* gene on the smaller (GAA1) and larger (GAA2) alleles. *n* no of subjects. *sd* standard deviation. *v1* visit 1 (baseline). *v2* visit 2 (follow-up).

Category	Feature	Description
structural imaging	scp_vol, mcp_vol, icp_vol	Volumes of superior, middle, and inferior cerebellar peduncles
	midbrain_vol, pons_vol, medulla_vol	Volumes of brainstem regions
	ant_CBLM_I_V_vol, sup-post_CBLM_VI_VII_vol, inf-post_CBLM_VIII_IX_vol, floc_CBLM_X_vol, vermis_CBLM_vol	Volumes of different cerebellar lobes
Diffusion imaging	scp_fa, mcp_fa, icp_fa	Fractional anisotropy (fa) of cerebellar peduncles
	scp_md, mcp_md, icp_md	Mean diffusivity (md) of cerebellar peduncles
	scp_ad, mcp_ad, icp_ad	Axial diffusivity (ad) of cerebellar peduncles
	scp_rd, mcp_rd, icp_rd	Radial diffusivity (rd) of cerebellar peduncles
Quantitative susceptibility imaging	dentate_L_vol, dentate_R_vol	Volume of left and right dentate nucleus
	dentate_L_susceptibility, dentate_R_susceptibility	Susceptibility values of the dentate nucleus
Background	FARS, SARA	Clinical ataxia rating scales
	sex, age_at_scan, age_at_onset	Basic demographic and disease-related information
	disease_duration_to_10, disease_duration_10_up	Disease duration categorised by severity progression pattern
	GAA1, GAA2	Genetic information (GAA repeat lengths)

Table 2. Multidomain, multimodal variables of interest for ML model development and subsequent composites. *vol* volume, *scp* superior cerebellar peduncle, *mcp* middle cerebellar peduncle, *icp* inferior cerebellar peduncle, *ant* anterior, *sup-post* superior-posterior, *inf-post* inferior-posterior, *floc* flocculonodular lobe, *CBLM* cerebellum, *I-X* represents cerebellar lobules, *L* left, *R* right, *fa* fractional anisotropy, *md* mean diffusivity, *ad* axial diffusivity, *rd* radial diffusivity, *GAA1* GAA1 repeat length; *GAA2* GAA2 repeat length.

mark, allowing the model to reflect potential shifts in progression dynamics more accurately. The distribution of FARS and SARA scores is shown in Supplementary Figure S2.

Machine learning models

First, we developed elasticnet[44] regression models to find the optimal combination of multi-domain variables that exhibited the strongest predictive association with FARS scores. A series of models were trained to predict FARS scores using different combinations of the neuroimaging and background variables (Table 2). Outlier detection was conducted for each feature set using Tukey's Fences method[45], identifying data points as outliers if they exceeded three times the interquartile range (IQR) beyond the upper or lower quartiles. Each participant's visit data at a particular time point was treated as one sample, with each participant potentially contributing a maximum of two samples. Only samples with a complete set of background and neuroimaging features at any visit were included in the cross-sectional models, while those containing one or more outlier features were excluded from the analysis. The final FARS prediction models were trained on cross-sectional data from 34 samples.

Our first goal was to identify the best-performing model with the optimised hyperparameter sets and the most predictive combination of input variables for prediction of FARS scores. Due to the small sample size, the dataset was not split into independent training, validation and testing sets. Rather, model performance was evaluated using repeated cross-validation, with 10 folds in each of 10 repetitions. Each "fold" contained 90% of the total data as training and 10% of the data as validation. To prevent artificially inflating the model's performance, we ensured that each participant was included in either the training or validation set, but not both, through a participant-based split. The hyperparameters of the models were tuned during the cross-validation, using a grid search as follows: alpha (mixture) parameters between 0 and 1 were tried in steps of 0.1, and lambda parameters between 0 and 20 were tried in steps of 0.1. Alpha determines the relative weight of the ridge and lasso penalties, while lambda determines the overall strength of the regularisation.

R^2 and root mean square error (RMSE) performance metrics were estimated by pooling the predictions from all the validation sets in the 100 model fits, as recommended by the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement article (Collins et al., 2015). The RMSE of each model quantifies the difference between predicted and actual scores, which is particularly sensitive to outliers or large deviations. The smaller the RMSE values, the better the prediction. Additionally, the R^2 performance metric measures the proportion of variance in the FARS scores that is predictable from the input variables in the model. R^2 values generally range from 0 to 1, with a value of 1 indicating perfect prediction. We compared the model performances and selected the optimal model based on cross-validation performance metrics, rather than relying on the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) scores. This approach enabled a better understanding of the model's generalisation ability without making any specific assumptions about model parameters. Regression models were fitted to the data using the glmnet package[46] and tidymodels framework[47] in R (R Core Team, 2013).

Variable importance

We calculated estimates of relative importance for each variable in the model prediction task using standardised coefficients, which represent the standardised weight of each predictor. Before model training, the predictors and targets were centred and scaled to have a mean of zero and a standard deviation of one. We generated

standardised coefficient plots using 201 bootstraps from the entire dataset and optimised model hyperparameters from cross-validation. Bootstrapping provides a robust estimation of the uncertainty in feature coefficients, irrespective of dataset distribution characteristics. In these plots, the least influential predictor trends towards 0, while the most influential predictor trends towards 1 or −1. For a particular model and a predictor, we present the coefficient values obtained through bootstrapping, the 95% percentile interval, the coefficient value from training the model with the entire dataset, and the percentage of non-zero coefficients among the bootstraps.

Composite biomarker development and sensitivity analysis

The best performing predictive models, along with the optimal combinations of predictor variables, were then fit to the full dataset. The absolute coefficients from these models were used to generate weighted composites for each predictor combination. Subsequently, we assessed the sensitivity to disease progression over 2 years for each individual predictor, target variable, and the derived composites using paired t-tests and Cohen's d effect sizes (d) of the differences between visit 2 and visit 1 scores, where:

$$d = \frac{mean(visit2-visit1)}{sd(visit2-visit1)};\ sd = standard\ deviation$$

For each individual or composite variable, d-scores were calculated using only the participants who had available data for both Visit 1 and Visit 2 for that specific variable. For instance, the full set of measures was available for $n=15$ participants; therefore, all statistical analyses for the background and all_neuroimaging composite were performed using this subset. In contrast, FARS scores were calculated using data from $n=31$ participants. Additionally, d-scores for the neuroimaging composites were measurable in both the control and FRDA cohorts.

An overview of the methods is presented in (Fig. 1).

Model evaluation for SARA scores

In addition to evaluating the model's predictive performance for FARS scores, we also assessed its ability to predict SARA scores and measured the disease sensitivity of surrogate composite scores for SARA. This evaluation followed the same procedures as for the FARS, described above, and was conducted using $n=31$ samples (visits) to ensure consistency and robustness in our approach.

Results

Prediction performance of the models and optimal predictive input combinations

The cross-validation performance measures (Table 3; Fig. 2) showed that the highest FARS prediction was obtained with elasticnet with the combination of background and all_neuroimaging (structural, diffusion and

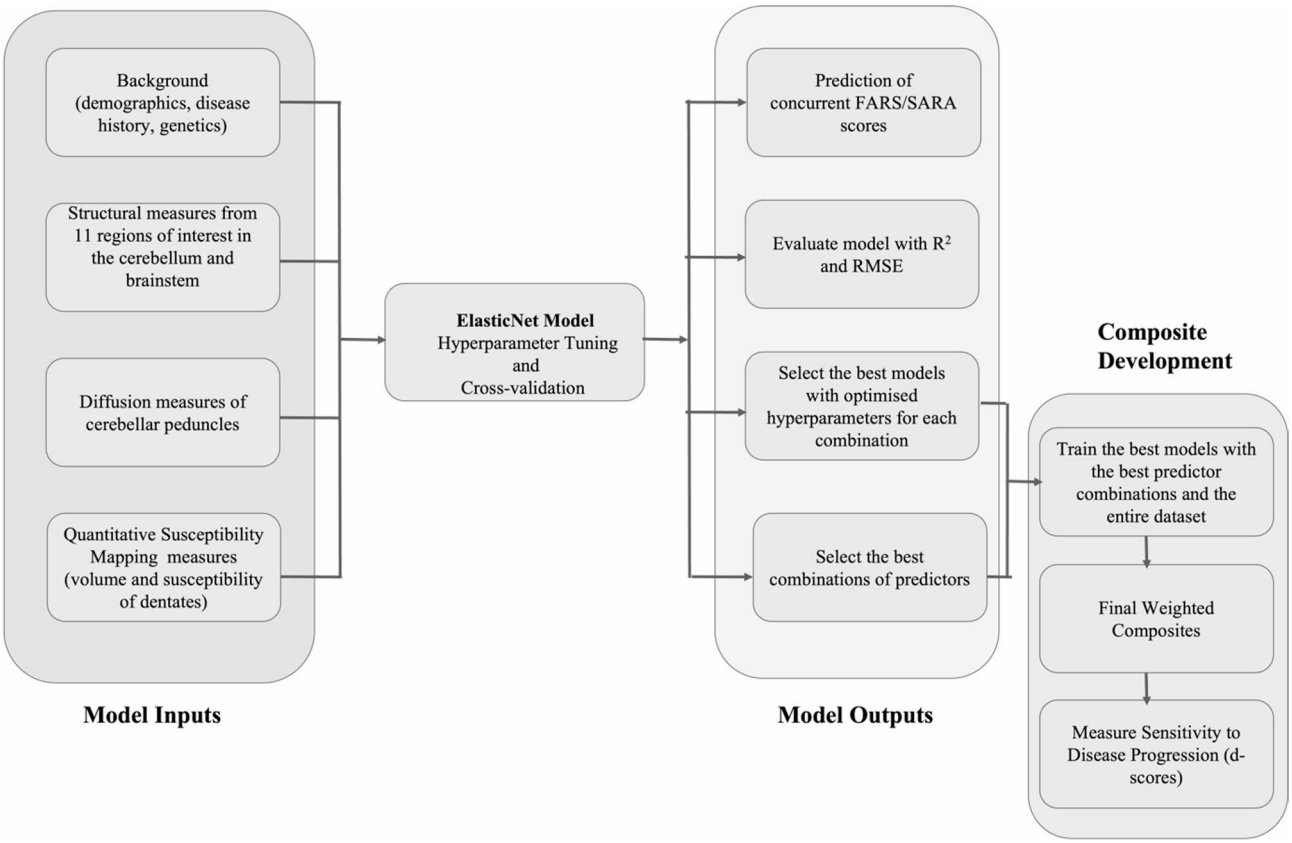

Fig. 1. Overview of methods.

Predictor variables	ML model - elasticnet			
	R^2	RMSE	Lambda (regularisation)	Alpha (mixture of ridge and lasso)
Background	0.63	17.33	1.2	1
Structural	0.2	25.6	1.6	0.1
Diffusion	0.36	22.95	15.5	0.1
QSM	-0.08	29.78	20	0
All_neuroimaging	0.55	19.13	7.6	0.2
Background + structural	0.57	18.82	2.3	1
Background + diffusion	0.77	13.85	5.7	0
Background + QSM	0.59	18.4	2.1	1
Background + structural + diffusion	0.77	13.7	2.3	0
Background + all_neuroimaging	0.79	13.19	1.3	0

Table 3. Performance measures of different ML models with various combinations of predictors.

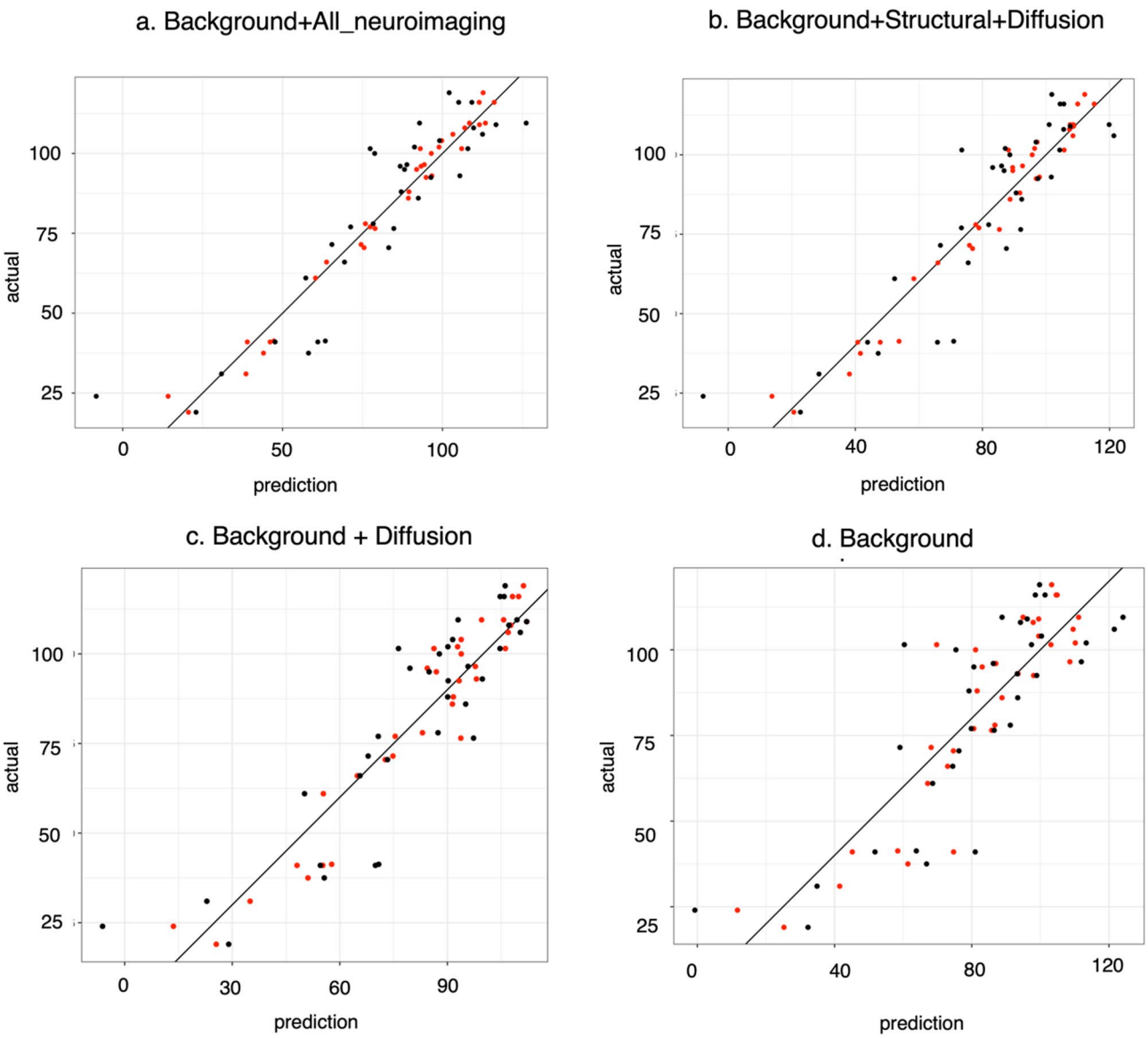

Fig. 2. Actual and predicted FARS scores from the elasticnet model with top performing input combinations, in a sequence from highest to lowest (**a**–**d**). Black dots represent the averaged predictions across the validation sets during 10-fold repeated cross-validation with optimised hyperparameters. Red dots indicate predictions from the final model using the complete dataset. Perfect predictions would align along the diagonal line.

QSM) variables ($R^2 = 0.79$; RMSE = 13.19). This was achieved by alpha = 0, meaning, the model was only using ridge regression. The next best performance was observed with a model that included background, structural, and diffusion imaging variables ($R^2 = 0.77$; RMSE = 13.7). This was closely followed by the model combining background and diffusion variables ($R^2 = 0.77$; RMSE = 13.85) and the background-only model ($R^2 = 0.63$; RMSE = 17.33). Among the neuroimaging-only combinations, the model utilising all of the imaging modalities demonstrated the strongest predictive association with FARS, whereas the model using only QSM variables failed to find a predictive pattern, indicated by a negative R^2. The RMSE values for all model predictions were significantly lower than the standard deviation of target variables (29.1), indicating that the models were performing well above random chance. Figure 2 illustrates the model predictions with the top performing combinations.

Relative contribution of variables to prediction

The standardised coefficient plots in Fig. 3 illustrate the relative contributions of each component variable in the top-performing models. As expected from the characteristics of the elasticnet model, most features included in the models had non-zero coefficients. The coefficient estimates associated with each variable were relatively variable across the bootstrap samples for all the models, as evident from the x-axis data spread in (Fig. 3). This variability is expected due to the small dataset. Positive coefficients indicate that an increase in these variables is associated with an increase in FARS scores, while negative coefficients indicate that an increase in these variables is associated with a decrease in FARS scores.

Variables consistently exhibited positive coefficients across the models, as evident by the 95% confidence intervals, included disease duration, left dentate susceptibility, the mean and radial diffusivity of inferior cerebellar peduncles. Among these, disease duration emerged as the highest weighted predictor in the final models trained

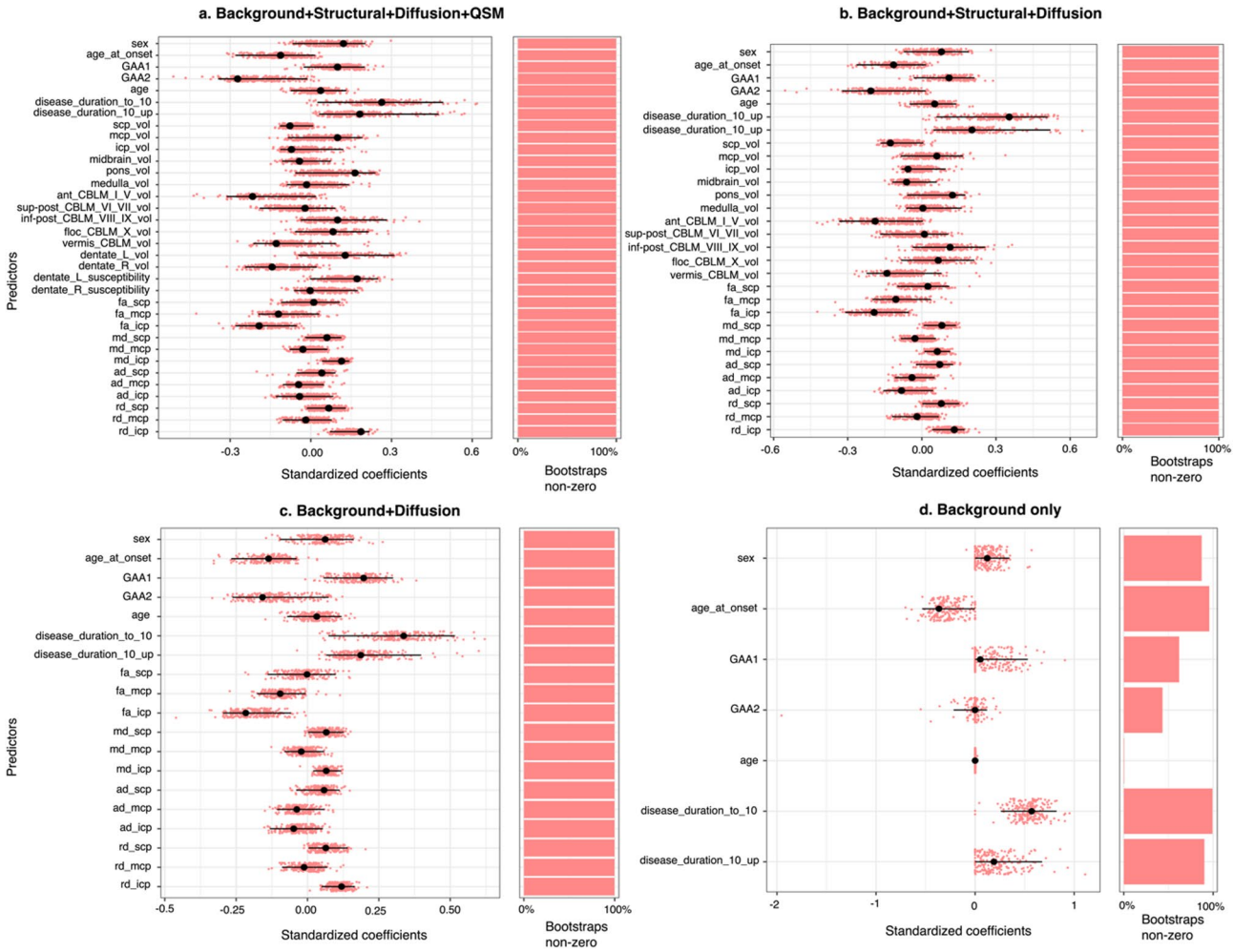

Fig. 3. Relative contributions of multidomain, multimodal variables to FARS prediction with top performing models, in a sequence from highest to lowest (**a**–**d**) performance. Red dots depict the coefficient values obtained via bootstrapping (201 iterations). Black lines show the 95% confidence intervals for each predictor's coefficient. Black blobs represent the coefficients from the model trained on the complete dataset. Bars on the right show the percentage of bootstrap samples in which a non-zero coefficient was observed.

Composites/single biomarkers	*n*	r2	mean_change	sd_change	d-score
Background + all_neuroimaging	15	0.926	8.5	7.6	1.12*
FARS	31	0.929	6.7	7.6	0.88*

Table 4. Statistical measures related to disease progression sensitivity for the composites surpassing FARS sensitivity. *n* no of participants, *sd* standard deviation, *r2* pearson correlation coefficient between visit 1 and visit 2 scores, $*p < 0.05$ (paired t-test between visit 1 and visit 2 scores).

Background components	$6.957\text{sex} - 0.369\text{age}_{onset} + 0.011\text{GAA1}_{repeatlength} - 0.034\text{GAA2}_{repeatlength} + 0.085\text{age}_{scan} + 6.129\text{DiseaseDuration}_{to_10} + 0.769\text{DiseaseDuration}_{10_up}$
	+
Structural neuroimaging components	$-0.033scp_{vol} + 0.003mcp_{vol} - 0.008icp_{vol} - 0.002midbrain_{vol} + 0.003pons_{vol} - 0.001medulla_{vol} - 0.004ant_CBLM_I_V_{vol} + 0.001inf\text{-}post_CBLM_VIII_IX_{vol} + 0.034floc_CBLM_X_{vol} - 0.11vermis_CBLM_{vol}$
	+
Diffusion neuroimaging components	$17.741scp_fa - 163.306mcp_fa - 229.242icp_fa + 65.268scp_md - 24.937mcp_md + 70.742icp_md + 38.843scp_ad - 38.297mcp_ad - 26.019icp_ad + 69.767scp_rd - 15.323mcp_rd + 110.395icp_rd$
	+
QSM neuroimaging components	$0.018dentate_L_{vol} - 0.023dentate_R_{vol} + 190.189dentate_L_{susceptibility} - 2.566dentate_R_{susceptibility}$

Table 5. Composite score equation (background + structural + diffusion + QSM), as a complement or surrogate to FARS (with the original weights of predictors estimated by the model in their original scale).

with the whole dataset. GAA1 repeat length and mean, and radial diffusivity of superior cerebellar peduncles also showed positive weights in the majority of models and across most bootstraps.

The fractional anisotropy (fa) of the inferior cerebellar peduncles consistently emerged with negative coefficients across the bootstrap models, exhibiting the same behaviour in the final models too. Additionally, age at onset, volume of the superior cerebellar peduncles, anterior cerebellum (lobule I-IV) volume, right dentate volume, and fractional anisotropy of the mid cerebellar peduncles also consistently appeared with negative weights, with only a few exceptions in certain bootstraps.

Overall, the coefficients of the variables from the final models lie within the confidence intervals derived from the bootstraps, indicating that the full model's estimates are consistent with the bootstrapped estimates. In the highest-performing model (background and all_neuroimaging), the volume of the superior cerebellar peduncle, mean diffusivity of the superior and inferior cerebellar peduncles, and radial diffusivity of the superior and inferior cerebellar peduncles had narrower confidence intervals, among the predictors with consistent positive and negative weights. This suggests that these variables had the most stable coefficient estimates across the bootstraps as well.

Disease progression sensitivity of single variables and composites

The disease progression sensitivity scores (Cohen's d) for the predictive composites were assessed, relative to the performance of the FARS scale alone (d = 0.88). Only the background and all-neuroimaging composite, which had the strongest predictive association with FARS (as shown in Table 3), exceeded the d-score of FARS (d = 1.12; Table 4 and Table S1). Among the imaging biomarkers in absence of the background variables, left dentate volume (d = −1.13) and the QSM composite (d = 1.39) also outperformed FARS (see supplementary Tables S1 and S2). However, these variables also elicited similar d-scores in the control group, with the same sign, indicating a similar direction of progression. This diminishes their effectiveness for clinical use. In contrast, right dentate susceptibility showed a high d-score (d = 0.8) in the FRDA group, closely aligning with FARS's sensitivity, while exhibiting a very low d-score (d = 0.011) in the control cohort, suggesting its potential for clinical use. Similarly, the background and structural, background and QSM, and background-only composites also showed d-scores greater than 0.7, compatible with the sensitivity of FARS as well (Table S1).

The background and all_neuroimaging composite demonstrated a significant difference between visit 1 and visit 2 (paired t-test; $p < 0.05$) and a high correlation (r^2), underscoring its sensitivity to changes over time and consistency in its progression pattern. However, similar sensitivity and consistency were observed in most other composites in the FRDA group, except for the diffusion composite (Table S1). The all-neuroimaging composite showed high sensitivity and specificity to disease progression in the FRDA group, with a strong correlation (r2 = 0.89) and significant differences between visit 1 and visit 2 ($p < 0.05$), while in the control group, it had a negative d-score of -0.19, a weak correlation (r2 = 0.32) and no significant difference between visits. None of the other imaging composites demonstrated both high sensitivity and specificity (Table S1).

Finally, we provide the equation for calculating the composite score for the background, structural, diffusion and QSM combination in (Table 5). This composite exhibited the highest sensitivity to disease progression, as indicated by its d-score, and the strongest predictive association with FARS (as shown in Table 3).

Comparative analysis with SARA models

We also conducted the same analysis for SARA scores, with the results summarised in Supplementary Tables S3 and S4. The findings and insights were highly consistent with the FARS model outcomes. The background and

all_neuroimaging composite emerged as the top surrogate candidate composite for SARA, demonstrating both strong predictive performance for SARA (R^2 = 0.72, RMSE = 4.68, sd of SARA scores = 9.07) and sensitivity to disease progression (d = 1.09), that surpassed the performance of SARA scale (d = 0.8). Surrogate composites for SARA and FARS revealed differences in predictor combinations and their relative weights, as expected given the unique characteristics of these scales. Altogether, these results suggest that our proposed method is an effective approach for developing surrogate or complementary composites for traditional clinical scores.

Discussion

In this study, we developed clinically relevant multimodal composite biomarkers that demonstrated a strong association with clinical severity and enhanced sensitivity to short-term (2-year) disease progression compared to traditional clinical scales in FRDA. By utilising a combination of ML predictive models and statistical analyses, we identified a weighted composite of background variables, including demographics, genetics, and disease history, combined with multimodal neuroimaging (structural MRI, diffusion MRI, and QSM), which proved highly predictive of FARS scores. This composite also exhibited greater sensitivity to short-term disease progression than FARS alone or any single imaging biomarker. Furthermore, external validation using SARA scores confirmed the robustness of this approach, with similar combinations of variables showing strong predictability for clinical scales and highest sensitivity to disease progression. As such, these objective composite biomarkers may offer a viable complementary or alternative measure to traditional clinical scales in clinical trials and practice.

Recent studies in FRDA have demonstrated the potential of ML to predict disease course using clinical scales or loss of ambulation, based on baseline demographic features[48]. However, these authors suggested that incorporating biological features, such as neuroimaging outcome measures, could enhance model accuracy and utility. Additionally, previous studies[19–22,49] have shown that quantitative outcome measures, such as movement measures captured by wearables could more accurately track and predict disease progression compared with clinical scales. Our study supports and extends these findings by demonstrating higher prediction accuracy for clinical severity when neuroimaging measures are included alongside demographic and disease-related individual characteristics. An interesting observation is that, despite the acknowledged subjectivity and measurement noise in clinical scales, several studies[20], including our current study, have demonstrated that wearable-derived measures or neuroimaging composites exhibit a strong predictive relationship with FARS or SARA scores. By leveraging machine learning's ability to identify latent patterns within input features that correlate strongly with FARS ratings, despite the inherent noise in FARS, these studies successfully developed objective biomarkers that are more robust and less susceptible to measurement variability than traditional clinical scales. This approach ensures that these biomarkers remain clinically relevant while providing greater reliability and predictive accuracy than conventional clinical assessments.

Existing approaches in ataxia research have attempted to develop composite measures by combining clinical assessments of movement, balance, gait, speech, and/or visual acuity measures[28,50–52], or by incorporating genetic markers and disease duration to improve the prediction of disease status or progression[53]. The novelty of our study lies in creating composite biomarkers that integrate multimodal neuroimaging data and outperform individual clinical and neuroimaging biomarkers, as well as composites based on a single imaging modality, in both predicting clinical severity and sensitivity to disease progression. These findings clearly suggest that weighted composite biomarkers can surpass individual biomarkers. However, the relationship is not linear, implying that simply summing features with their model-generated coefficients of features does not guarantee higher sensitivity to disease progression compared to single variables. For instance, QSM composite showed higher disease progression sensitivity than the background and QSM composite. The performance of the composite biomarkers is thus contingent on the specific combination of features and the optimization of their respective coefficients. Overall, the findings from this study pave the way for the development of more robust composite scores, ultimately leading to a more holistic approach to tracking disease progression, monitoring therapeutic efficacy, and improving management strategies for this challenging condition.

The standardised coefficient plots generated by bootstrapping revealed that the relative coefficients of predictor variables could vary both in magnitude and direction across new datasets. However, the variables that most consistently predicted FARS scores were well-aligned with previous studies[52,54,55], such as, positive correlations of disease duration with FARS. Similarly, the predictive study by Hohenfeld and colleagues[48] found disease duration to be the most influential variable, independent of the ML techniques used, consistent with our findings. Significant volume reduction and longitudinal alterations in the superior cerebellar peduncle have been observed in FRDA patients, correlating with disease severity. Furthermore, grey matter alterations in cerebellar lobules I–VI have been reported in multiple studies[38,56]. Higher FARS scores and in general the disease progression were shown to be associated with reduced fa and increased md and rd values in cerebellar peduncles[57], reflecting microstructural damage in FRDA[17]. Several MRI studies in FRDA have identified a reduction in dentate nucleus volume and increased magnetic susceptibility indicating iron accumulation and showed that these changes are correlated with ataxia severity[58,59]. All of these findings are highly consistent with our results.

In our study, we aimed to retain as many variables as possible within the model by leveraging elasticnet, utilising a wide range for the mixture parameter that controlled the ratio of lasso (tends to set some coefficients to zero, effectively performing variable selection) to ridge regression (which shrinks coefficients but keeps all variables). These models are particularly well-suited for small datasets with a large number of features. Our findings indicated that elasticnet, with the mixture parameter set to 0-0.02, provided the highest predictive performance in most cases. This parameter setting leans more toward ridge regression, allowing the model to maintain a high number of predictors with non-zero coefficients. For the final composite scores, we utilised absolute coefficients from the model trained on the entire dataset, taking different scales into account to ensure

practical feasibility. However, the robustness of these coefficients requires further verification and fine-tuning in future studies using larger, more diverse samples covering a wider range of predictors and clinical scores, with appropriate splitting into training, validation, and unseen test sets for predictive models. Optimized feature selection techniques that incorporate biological insights, combined with more complex nonlinear ensemble ML methods to minimize model prediction RMSEs, could lead to more powerful composite scores. The approach emphasized in this study could be refined by selecting composite components based on broader criteria, such as the strength of features in distinguishing between control and disease cohorts, rather than solely focusing on predictability for FARS or SARA scores, which have notable limitations as previously discussed. Post-filtering the components of the composites based on biological understanding and subjecting them to continuous refinement and testing might be a suitable approach as well, for developing an efficient, concise biomarker. Additionally, longitudinal predictive methodologies such as visit-to-visit predictions could yield a more effective composite of disease progression and provide deeper insights.

A larger sample size and further validation are also necessary to address critical aspects of this analysis. Both the FARS scores and the composite biomarker showed statistically significant changes over time ($p < 0.05$), with the composite biomarker demonstrating a stronger effect than FARS, suggesting greater sensitivity in capturing disease progression. However, a direct comparison of their rate of change (t-test: $p = 0.89$) showed no significant difference. A key limitation is the relatively small changes observed in clinical scales over two years, with FARS increasing by 6.74 (sd= 7.62, $n = 31$) and SARA by only 2.15 (sd = 2.68, $n = 27$), indicating slow disease progression and high variability. This may impact the ability to detect meaningful changes and limits conclusions on the composite biomarker's superiority. Future large-scale studies would confirm these findings with greater statistical confidence.

Personalisation of biomarkers is a critical goal of contemporary research to pave the way for precision medicine. We have demonstrated the potential for developing personalised composite biomarkers by incorporating components of an individual's unique demographics, genetics, and disease history. However, future studies should focus on achieving true personalization by generating composite biomarkers tailored to each individual's biological profile, resulting in individualised combinations and associated coefficients. This approach recognizes that the relevance of various biomarkers may change across different phases of the disease and that progression profiles in FRDA are highly heterogeneous among individuals[14,60]. This could be achieved by implementing cohort-specific training of ML models, with sufficient sample sizes within each cohort to capture the variability in disease progression and biomarker significance. Such an approach will enable the dynamic adjustment of composite biomarkers, enhancing their accuracy and clinical utility in predicting disease progression and tailoring therapeutic interventions for individuals with FRDA.

To guide the clinical application of these composite scores, we have prioritized them by equally weighting clinical relevance and disease sensitivity, and provided the corresponding equations. However, future researchers and clinicians may choose to prioritise different composites based on trade-offs between clinical relevance, disease sensitivity, and specific clinical objectives. For instance, when validating a positive FARS outcome in a clinical trial evaluating therapeutic efficacy, a FARS-relevant biomarker composite may be more suitable. However, a highly clinically relevant background-only composite, which includes time-varying factors like disease duration and age at scan, would not be appropriate, as these factors do not reflect therapeutic-related changes. Background-based biomarkers are better suited for monitoring disease progression, personalizing care, complementing subjective clinical scales, filling in missing data, or comparing cohorts at specific time points. In such cases, the background and neuroimaging composites should still be useful though. When the goal is to monitor disease progression in a natural history study or to evaluate treatment efficacy with the most sensitive measures available, clinicians may prefer composites based solely on neuroimaging, even if this comes at the expense of relevance with clincial scores. Ultimately, the specific objective, whether disease progression monitoring, therapeutic efficacy assessment, or another purpose, will determine the most appropriate composite for clinical use.

Finally, as neuroimaging is not yet a standard tool in clinical care, extensive validation is necessary before it can be widely adopted. In complex, multisystem diseases like FRDA, which remain relatively unknown and underexplored, composites may prove particularly effective, as a single measure often fails to capture the full scope of the condition. However, continued research into composite biomarkers using neuroimaging is essential for advancing their use in both research and clinical practice. A key challenge in the translation of neuroimaging composites is their cost, particularly in comparison to more cost-effective alternatives such as wearable-derived biomarkers. While future research should explore more affordable composite biomarker solutions, neuroimaging composites still offer distinct advantages. They provide a comprehensive evaluation of brain macrostructure and microstructural integrity with predictive accuracy for clinical scores that is highly comparable to wearable measures[20]. At the same time, they may offer greater sensitivity and informativeness in the early to moderate stages of the disease particularly if the progression of FRDA follows a pattern where structural and diffusion impairments emerge before functional and clinical symptoms.

Conclusion

This study marks a pioneering effort in developing clinically relevant, multimodal, highly sensitive, and fully objective composites for biomarker research and disease progression in FRDA. By employing classical ML techniques, we have demonstrated the potential to create stronger composite biomarkers of disease progression, compared to individual biomarkers, enhancing the monitoring of FRDA progression. This work lays the groundwork for future investigations into composite biomarkers for FRDA incorporating additional modalities, larger sample sizes, and more advanced ML techniques. Our findings highlight the importance of multimodal data and predictive modelling in advancing biomarker development for FRDA and potentially other rare neurodegenerative diseases.

Data availability

The original IMAGE-FRDA datasupporting the findings of this study will be made available upon reasonable-request to the corresponding author, subject to privacy considerations anddata-sharing agreements. However, minimal datasets containing all predictor andtarget feature values required to replicate the findings will be made readilyavailable to reviewers upon request, in accordance with journal policies. All codebases developed for this study will be openly accessible on a publicrepository upon publication to ensure transparency, reproducibility, and fosterfurther research collaboration.

Received: 16 October 2024; Accepted: 2 May 2025
Published online: 21 May 2025

References

1. Campuzano, V. et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**, 1423–1427 (1996).
2. Lynch, D. R., Schadt, K., Kichula, E., McCormack, S. & Lin, K. Y. Friedreich ataxia: multidisciplinary clinical care. *J. Multidiscip Healthc.* **14**, 1645–1658 (2021).
3. Lee, A. & Omaveloxolone First approval. *Drugs* **83**, 725–729 (2023).
4. Pilotto, F., Chellapandi, D. M. & Puccio, H. Omaveloxolone: a groundbreaking milestone as the first FDA-approved drug for Friedreich ataxia. *Trends Mol. Med.* **30**, 117–125 (2024).
5. Gunther, K. & Lynch, D. R. Pharmacotherapeutic strategies for Friedreich ataxia: a review of the available data. *Expert Opin. Pharmacother*. 1–11. https://doi.org/10.1080/14656566.2024.2343782 (2024).
6. Lynch, D. R. et al. Efficacy of Omaveloxolone in Friedreich's ataxia: Delayed-start analysis of the moxie extension. *Mov. Disord Off J. Mov. Disord Soc.* **38**, 313–320 (2023).
7. Therapeutics, P. PTC Therapeutics announces topline results from vatiquinone MOVE-FA registration-directed trial. (2023).
8. Pizzamiglio, C., Vernon, H. J., Hanna, M. G. & Pitceathly, R. D. S. Designing clinical trials for rare diseases: unique challenges and opportunities. *Nat. Rev. Methods Primer*. **2**, s43586–s43022 (2022).
9. Milne, S. C. et al. Interrater reliability of the scale for the assessment and rating of ataxia, Berg balance scale, and functional independence measure motor domain in individuals with hereditary cerebellar ataxia. *Arch. Phys. Med. Rehabil.* **104**, 1646–1651 (2023).
10. Regner, S. et al. Friedreich ataxia clinical outcome measures: natural history evaluation in 410 participants. *J. Child. Neurol.* **27**, 1152–1158 (2012).
11. Lynch, D. R. et al. Safety and efficacy of Omaveloxolone in Friedreich ataxia (MOXIe Study). *Ann. Neurol.* **89**, 212–225 (2021).
12. Gavriilaki, M. et al. Therapeutic biomarkers in Friedreich's ataxia: a systematic review and meta-analysis. *Cerebellum Lond. Engl.* https://doi.org/10.1007/s12311-023-01621-6 (2023).
13. Clemente, J. PTC Therapeutics announces positive results from long-term treatment studies and updates on regulatory progress for vatiquinone Friedreich ataxia program.
14. Rummey, C. et al. Natural history of Friedreich ataxia: heterogeneity of neurologic progression and consequences for clinical trial design. *Neurology* **99**, e1499–e1510 (2022).
15. Selvadurai, L. P. et al. Multiple mechanisms underpin cerebral and cerebellar white matter deficits in Friedreich ataxia: the image-frda study. *Hum. Brain Mapp.* **41**, 1920–1933 (2020).
16. Georgiou-Karistianis, N. et al. A natural history study to track brain and spinal cord changes in individuals with Friedreich's ataxia: TRACK-FA study protocol. *PLoS One* **17**, e0269649 (2022).
17. Rezende, T. J. R. et al. Longitudinal magnetic resonance imaging study shows progressive pyramidal and callosal damage in Friedreich's ataxia. *Mov. Disord Off J. Mov. Disord Soc.* **31**, 70–78 (2016).
18. Adanyeguh, I. M. et al. Brain MRI detects early-stage alterations and disease progression in Friedreich ataxia. *Brain Commun.* **5**, fcad196 (2023).
19. Corben, L. A. et al. Developing an instrumented measure of upper limb function in Friedreich ataxia. *Cerebellum Lond. Engl.* **20**, 430–438 (2021).
20. 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).
21. Krishna, R., Pathirana, P. N., Horne, M. K., Szmulewicz, D. J. & Corben, L. A. Objective assessment of progression and disease characterization of Friedreich ataxia via an instrumented drinking cup: preliminary results. *IEEE Trans. Neural Syst. Rehabil Eng. Publ IEEE Eng. Med. Biol. Soc.* **29**, 2365–2377 (2021).
22. Ngo, T. et al. Balance deficits due to cerebellar ataxia: A machine learning and cloud-based approach. *IEEE Trans. Biomed. Eng.* **68**, 1507–1517 (2021).
23. Ngo, T. et al. Federated deep learning for the diagnosis of cerebellar ataxia: privacy preservation and auto-crafted feature extractor. *IEEE Trans. Neural Syst. Rehabil Eng.* **30**, 803–811 (2022).
24. Ngo, T. et al. Technological evolution in the instrumentation of ataxia severity measurement. *IEEE Access.* **11**, 14006–14027 (2023).
25. Blair, I. A. et al. The current state of biomarker research for Friedreich's ataxia: a report from the 2018 FARA biomarker meeting. *Future Sci. OA*. **5**, FSO398 (2019).
26. Xu, T., Fang, Y., Rong, A. & Wang, J. Flexible combination of multiple diagnostic biomarkers to improve diagnostic accuracy. *BMC Med. Res. Methodol.* **15**, 94 (2015).
27. Pierce, M. & Emsley, R. A comparison of approaches for combining predictive markers for personalised treatment recommendations. *Trials* **22**, 20 (2021).
28. Bürk, K., Schulz, S. R. & Schulz, J. B. Monitoring progression in Friedreich ataxia (FRDA): the use of clinical scales. *J. Neurochem.* **126**, 118–124 (2013).
29. Lynch, D. R. et al. Measuring Friedreich ataxia: complementary features of examination and performance measures. *Neurology* **66**, 1711–1716 (2006).
30. Rummey, C., Kichula, E. & Lynch, D. R. Clinical trial design for Friedreich ataxia—Where are we now and what do we need? *Expert Opin. Orphan Drugs*. **6**, 219–230 (2018).
31. Winchester, L. M. et al. Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia. *Alzheimers Dement.* **19**, 5860–5871 (2023).
32. de Vries, R. & Sterk, P. J. eNose breathprints as composite biomarker for real-time phenotyping of complex respiratory diseases. *J. Allergy Clin. Immunol.* **146**, 995–996 (2020).
33. Kyriazakos, S. et al. Discovering composite lifestyle biomarkers with artificial intelligence from clinical studies to enable smart ehealth and digital therapeutic services. *Front. Digit. Health*. **3**, 648190 (2021).
34. Yu, W. & Park, T. Two simple algorithms on linear combination of multiple biomarkers to maximize partial area under the ROC curve. *Comput. Stat. Data Anal.* **88**, 15–27 (2015).
35. Robin, X. et al. PanelomiX: A threshold-based algorithm to create panels of biomarkers. *Transl Proteom.* **1**, 57–64 (2013).

36. Shaw, R., Lokshin, A. E., Miller, M. C., Messerlian-Lambert, G. & Moore, R. G. Stacking machine learning algorithms for biomarker-based preoperative diagnosis of a pelvic mass. *Cancers* **14**, 1291 (2022).
37. Harding, I. H. et al. Tissue atrophy and elevated iron concentration in the extrapyramidal motor system in Friedreich ataxia: the IMAGE-FRDA study. *J. Neurol. Neurosurg. Psychiatry*. **87**, 1261–1263 (2016).
38. Selvadurai, L. et al. Longitudinal structural brain changes in Friedreich ataxia depend on disease severity: the IMAGE-FRDA study. *J. Neurol.* **268**, 1–12 (2021).
39. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *J. Am. Coll. Dent.* **81**, 14–18 (2014).
40. Ward, P. G. D. et al. Longitudinal evaluation of iron concentration and atrophy in the dentate nuclei in Friedreich ataxia. *Mov. Disord Off J. Mov. Disord Soc.* **34**, 335–343 (2019).
41. Selvadurai, L. et al. Cerebral and cerebellar grey matter atrophy in Friedreich ataxia: the IMAGE-FRDA study. *J. Neurol.* **263**, (2016).
42. Harding, I. H. et al. Brain structure and degeneration staging in Friedreich ataxia: magnetic resonance imaging volumetrics from the ENIGMA-ataxia working group. *Ann. Neurol.* **90**, 570–583 (2021).
43. Kerestes, R. et al. Reduced cerebello-cerebral functional connectivity correlates with disease severity and impaired white matter integrity in Friedreich ataxia. *J. Neurol.* **270**, 2360–2369 (2023).
44. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
45. Tukey, J. W. *Exploratory Data Analysis* 2 (Springer, 1977).
46. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
47. Kuhn, M. & Wickham, H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. (2020).
48. Hohenfeld, C. et al. Prediction of the disease course in Friedreich ataxia. *Sci. Rep.* **12**, 19173 (2022).
49. Fichera, M. et al. Accelerometer-based measures in Friedreich ataxia: a longitudinal study on real-life activity. *Front. Pharmacol.* **15**, 1342965 (2024).
50. Friedman, L. S. et al. Measuring the rate of progression in Friedreich ataxia: implications for clinical trial design. *Mov. Disord Off J. Mov. Disord Soc.* **25**, 426–432 (2010).
51. Klockgether, T. et al. Consensus recommendations for clinical outcome assessments and registry development in ataxias: Ataxia global initiative (AGI) working group expert guidance. *Cerebellum* **23**, 924–930 (2024).
52. Patel, M. et al. Progression of Friedreich ataxia: quantitative characterization over 5 years. *Ann. Clin. Transl Neurol.* **3**, 684–694 (2016).
53. Selvadurai, L. P. et al. The S-factor, a new measure of disease severity in spinocerebellar ataxia: findings and implications. *Cerebellum Lond. Engl.* **22**, 790–809 (2023).
54. Fahey, M. C., Corben, L., Collins, V., Churchyard, A. J. & Delatycki, M. B. How is disease progress in Friedreich's ataxia best measured? A study of four rating scales. *J. Neurol. Neurosurg. Psychiatry*. **78**, 411–413 (2007).
55. Pandolfo, M. Friedreich ataxia: the clinical picture. *J. Neurol.* **256**, 3–8 (2009).
56. Krahe, J. et al. Increased brain tissue sodium concentration in Friedreich ataxia: A multimodal MR imaging study. *NeuroImage Clin.* **34**, 103025 (2022).
57. Vavla, M. et al. Functional and structural brain damage in Friedreich's ataxia. *Front. Neurol.* **9**, (2018).
58. Deistung, A. et al. Quantitative susceptibility mapping reveals alterations of dentate nuclei in common types of degenerative cerebellar ataxias. *Brain Commun.* **4**, fcab306 (2022).
59. Harding, I. H. et al. Localized changes in dentate nucleus shape and magnetic susceptibility in Friedreich ataxia. *Mov. Disord Off J. Mov. Disord Soc.* https://doi.org/10.1002/mds.29816 (2024).
60. Reetz, K. et al. Progression characteristics of the European Friedreich's ataxia consortium for translational studies (EFACTS): a 4-year cohort study. *Lancet Neurol.* **20**, 362–372 (2021).

Acknowledgements
We extend our sincere thanks to all the participants of the IMAGE-FRDA study and the investigators involved, with special acknowledgment to Prof. Martin Delatycki, Prof. Gary Egan, and Prof. Elsdon Storey for their invaluable contributions to this study.

Author contributions
Susmita Saha: conception and design of the study, data preprocessing, model development, interpretation of results, manuscript writing Louise A. Corben: interpretation of results, manuscript reviewLouisa Selvadurai: data preprocessing, manuscript reviewIan H. Harding: project supervision, data preprocessing, interpretation of results, critical manuscript reviewNellie Georgiou-Karistianis: project supervision, critical manuscript review.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01047-6.

Correspondence and requests for materials should be addressed to N.G.-K.

Reprints and permissions information is available at 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) 2025