RESEARCH ARTICLE

CrossMark

Structural brain changes versus self-report: machine-learning classification of chronic fatigue syndrome patients

Landrew S. Sevel[1] · Jeff Boissoneault[1] · Janelle E. Letzen[1] · Michael E. Robinson[1] · Roland Staud[2]

Abstract

Chronic fatigue syndrome (CFS) is a disorder associated with fatigue, pain, and structural/functional abnormalities seen during magnetic resonance brain imaging (MRI). Therefore, we evaluated the performance of structural MRI (sMRI) abnormalities in the classification of CFS patients versus healthy controls and compared it to machine learning (ML) classification based upon self-report (SR). Participants included 18 CFS patients and 15 healthy controls (HC). All subjects underwent T1-weighted sMRI and provided visual analogue-scale ratings of fatigue, pain intensity, anxiety, depression, anger, and sleep quality. sMRI data were segmented using FreeSurfer and 61 regions based on functional and structural abnormalities previously reported in patients with CFS. Classification was performed in RapidMiner using a linear support vector machine and bootstrap optimism correction. We compared ML classifiers based on (1) 61 a priori sMRI regional estimates and (2) SR ratings. The sMRI model achieved 79.58% classification accuracy. The SR (accuracy = 95.95%) outperformed both sMRI models. Estimates from multiple brain areas related to cognition, emotion, and memory contributed strongly to group classification. This is the first ML-based group classification of CFS. Our findings suggest that sMRI abnormalities are useful for discriminating CFS patients from HC, but SR ratings remain most effective in classification tasks.

Keywords Machine learning · Gray matter · Self-report · Chronic fatigue · Classification

Introduction

Chronic fatigue syndrome (CFS) is a condition mainly characterized by disabling fatigue of unknown origin for at least 6 months (Gunn et al. 1993). In addition, CFS patients commonly report neurologic, psychiatric, and physical symptoms (Fukuda et al. 1994). Although the etiology of CFS is not fully understood, central nervous system (CNS) dysfunction (Chen et al. 2008) as well as sensitization of peripheral fatigue pathways (Staud et al. 2015) have been suggested. Magnetic resonance imaging (MRI) studies have

✉ Roland Staud
staudr@ufl.edu

[1] Department of Clinical and Health Psychology, University of Florida, Gainesville, FL, USA

[2] Department of Medicine, College of Medicine, University of Florida, PO Box 100277, Gainesville 2610-0277, FL, USA

described significant differences between patients with CFS and healthy controls (HC) across measures of structural and functional neuroimaging, as well as neurochemical factors (Chen et al. 2008).

Several studies have reported abnormal grey matter volumes in patients with CFS. Whereas one study observed global grey matter reduction in CFS (de Lange et al. 2005), others described select reductions within prefrontal (Okada et al. 2004; de Lange et al. 2008) and occipital cortices (Puri et al. 2012) as well as within angular and parahippocampal gyri (Puri et al. 2012). Whether structural brain features can be used to classify individuals with CFS from HC, however, has not yet been tested.

Supervised machine learning (ML) classifiers are specific application of ML technology that uses individual feature values to predict the discrete and pre-assigned class to which a given example belongs. The performance of a given model can be assessed by characterizing its success in correctly predicting group membership, also allowing for the comparison of performance between feature sets (Pereira et al. 2009). Model performance evaluation methods can yield estimates such as sensitivity (i.e., true positive rate) and

specificity (i.e., true negative rate), and allows for the assessment of positive predictive value (PPV, i.e., probability that positive result occurs in a patient who has the condition) and negative predictive value (NPV, i.e., probability that a negative result occurs in an individual who does not have the condition). This method has been used in the classification of individuals with chronic pain (Ung et al. 2014; Labus et al. 2015) and depressed mood (Nouretdinov et al. 2011) with success. Self-report (SR) is the gold standard for classification of most chronic pain and psychiatric conditions, but only few studies have compared neuroimaging and self-report data for patient classification (Robinson et al. 2015).

The present study had several aims. First, we examined the performance of structural neuroimaging data in classifying individuals with and without CFS (i.e., sMRI model). Our interest was in determining whether structural regions with previously observed functional abnormality were predictive of patient status. Second, we examined the performance of a model solely composed of self-reported symptoms (i.e., SR model). Finally, we measured how well each model (i.e., sMRI and SR) would perform in a range of application settings using ecologically valid base rates (i.e., prevalence) by applying Bayes' theorem (Lopez Puga et al. 2015). This method may produce more representative estimates of classification metrics, given that calculations of PPV and NPV depend on the condition's prevalence in the context of specific settings, where the classifier may be applied. For example, a classifier may be used as a diagnostic tool in the clinic, where prevalence is high, or as a screening tool in the general population, where prevalence is much lower (Grimes and Schulz 2005; Robinson et al. 2016).

Methods and materials

Participants

Participants comprised of 15 HC and 18 CFS patients. Two additional HC participants were initially recruited for this study; however, data quality issues resulted in their exclusion from analyses. CFS participants were required to fulfill Center for Disease Control criteria for chronic fatigue (Fukuda et al. 1994), which include unexplained, persistent, or relapsing chronic fatigue that is of new or definite onset over a period of at least 6 months. The fatigue could not be the result of ongoing exertion and had to result in substantial reduction in the previous levels of occupational, educational, social, or personal activities. In addition, CFS participants were required to have at least four of the following additional symptoms in the 6 months prior to assessment, including impaired memory/concentration, sore throat, tender cervical or axillary lymph nodes, muscle pain, multi-joint pain, new headaches, unrefreshing sleep, and post-exertional malaise

(Fukuda et al. 1994). CFS diagnosis was confirmed by a board-certified rheumatologist (RS). Subjects were excluded if they had a history of heart disease, chronic obstructive pulmonary disease, malignancy, psychiatric illnesses, or other disorders that could confound the diagnosis. CFS participants were recruited via outpatient medical clinics at the University of Florida, where they had completed the Center for Epidemiological Studies-Depression Scale (CES-D) to exclude depression (Radloff 1977), or through advertisements. Controls were excluded if they had a history of chronic fatigue, chronic pain, or psychiatric illness. All individuals were asked to sleep for at least 6 h and refrain from caffeine consumption prior to the study session. Consumption of alcohol, psychoactive substances, or any medication except vitamins was prohibited 24 h before the study. Individuals were also excluded if they were current smokers, yielded a positive pregnancy test result (women of childbearing potential only) or contained any ferromagnetic metal within the body.

Because the focus of this study was primarily conceptual and mechanistic rather than diagnostic, Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines (Cohen et al. 2016) have not been applied in the current report.

Clinical and affective ratings

Prior to MRI scanning, clinical fatigue, pain intensity, sleep quality, anxiety, anger, and depression were assessed using mechanical visual analog scales (VAS). VAS measures were selected as a comparison for MRI-based measures because of their reliability, validity, efficiency, and low cost (Price and Harkins 1987). Each scale was anchored on the left with "No (fatigue, pain, anxiety, anger, depression) at all" and on the right with "The most (fatigue, pain, anxiety, anger, depression) imaginable." Sleep quality was also assessed using a VAS with "Best sleep/No problem sleeping" on left, and "Worst sleep/no sleep" on right. It is important to note that these SR measures were not used to diagnose CFS patients. As mentioned above, patients were diagnosed according to the criteria defined by Fukuda et al. (1994). As such, the SR feature set reflects subjective symptomatology. The SR feature set does, however, serve as a proxy for the process of clinical diagnosis, and as such is an appropriate comparison to assess the clinical utility of neuroimaging-based classifiers.

Data acquisition and processing

Data were acquired on a Philips Achieva 3T MRI scanner with a 32-channel head coil. All participants completed one high-resolution, 3D anatomical, T_1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) sequence. The

parameters were as follows: 176 sagittal slices (1 mm thickness), TR = 7.2 ms, TE = 3.2 ms, FA = 8°, FOV = 240 mm, and voxel size = 1 mm^3. Total acquisition time was 4 min and 34 s.

All MRI data were processed individually through the automated subcortical segmentation and cortical parcellation tool in FreeSurfer, version 5.3 (Martinos Center for Biomedical Imaging, Charlestown, MA, USA) (Fischl et al. 2002; Fischl 2012). To increase the probability that each discrete region is correctly labelled, FreeSurfer considers aspects of the collected data and general aspects of MRI such as signal intensity of different regions, producing accurate and reliably segmentation (Jovicich et al. 2009). The software produces indices of area, thickness, and volume for cortical structures, and volume for subcortical structures, resulting in 274 unique neuroanatomical features. White matter, gray matter, and pial matter boundaries were reviewed in each subject for gross artifacts and errors in segmentation by the authors (LS and JB).

Feature selection

Of the initial 274 anatomical features, a subset of 61 features was chosen for classification in the *a priori* feature model (sMRI). The areas selected were those in which abnormalities in brain structure and function were previously observed in CFS patients compared to HC (Okada et al. 2004; Puri et al. 2012; Boissoneault et al. 2016; Gay et al. 2016). Table 1 provides a complete list of regions. The SR model included all ratings discussed above except fatigue, which was excluded due to concerns regarding circularity of analyses given the closeness of this construct to the most salient aspects of CFS.

Modeling approach

Classification analyses were performed within Rapid-Miner Studio version 7.6.001 (RapidMiner, Inc., Boston, MA, USA). Data were normalized via z-transformation. Z-transformation produces standardized values across all features by calculating the mean value per feature, subtracting individual scores, and dividing by the standard deviation. This procedure eliminates scale differences between features that may otherwise confound classifier performance (Alpaydin 2014). The performance of two feature sets in classifying CFS patients and HC was compared: (1) structural MRI composed of the 61 regions based upon functional differences (sMRI) and (2) six clinical and affective, self-reported VAS ratings (SR). Features were not further mathematically transformed in any way prior to model estimation. A linear support vector machine (SVM), as implemented by Chang's Library of Support Vector Machines (LIBSVM) (Fan et al. 2005) within

Table 1 Regions used for classification analyses

Hemisphere	Region
Left	Caudate[a]
Left	Pallidum[a]
Right	Hippocampus[a]
Right	Pallidum[a]
Left	Caudal anterior cingulate cortex
Left	Caudal middle frontal cortex
Left	Insula
Left	Lingual gyrus
Left	Parahippocampal
Left	Postcentral gyrus
Left	Posterior cingulate cortex
Left	Precuneus
Left	Rostral anterior cingulate cortex
Left	Rostral middle frontal cortex
Left	Superior frontal cortex
Right	Caudal anterior cingulate cortex
Right	Cuneus
Right	Inferior parietal lobe
Right	Lateral occipital lobe
Right	Precuneus
Right	Rostral anterior cingulate cortex
Right	Superior parietal lobe
Right	Superior temporal lobe

For all regions, unless marked with superscript a, estimates of area, thickness, and volume were included. Those marked with an asterisk represent subcortical regions for which only estimates of volume were used. A total of 61 neuroanatomical features were included in analyses

the RapidMiner environment, was used in classification. SVMs classify labelled data by estimating the hyperplane that maximizes the width of the margin separating the hyperplane and data. Margin width maximization reduces the expected risk of model error (i.e., decreases model complexity). In cases where data are not perfectly separated by a hyperplane, a soft margin may be used which allows for some cases to be on the wrong side of the margin (i.e., increased model complexity). This complexity is captured in a penalty proportional to how far cases are on the wrong side of the margin. The regularization parameter, *C*, balances the trade-off between model complexity and error frequency. This parameter was tuned by standard grid-search optimization according to Hsu et al. (2008). SVMs have previously been successfully applied to the classification of chronic low back and pelvic pain (Bagarinao et al. 2014; Ung et al. 2014). To aid in feature weight interpretation, weights were squared and ranked (Guyon et al. 2002). Broadly, greater values correspond to features that were more influential in the classification of each case.

Performance assessment

For all models, accuracy [the sum of true positives (TP) and true negatives (TN) divided by the sum of TN, TP, false positives (FP) and false negatives (FN); $(TN + TP)/(TN + TP + FN + FP)$], sensitivity [$TP/(TP + FN)$], and specificity [$TN/(TN + FP)$] were calculated. For a brief review of these performance measures in the context of brain biomarkers for chronic pain, see (Robinson et al. 2016). To adjust for model optimism, bootstrapping was used to reduce bias in performance estimates (Cenzer et al. 2013). Similar to model assessment in cross-validation and with holdout samples, this approach is a viable alternative to guard against overfitting and may produce more conservative estimates of model performance (Friedman et al. 2009; Smith et al. 2014; Walsh et al. 2017). This approach is particularly germane for small sample sizes, such as those regularly used in human neuroimaging investigations. To estimate and correct for model optimism the following steps are used: (1) a classification model is trained on the full set of available data; (2) a set of bootstrap data sets are collected based upon the original data (200 replications were used in the current study); (3) models are estimated for each bootstrap sample replication; (4) the previously estimated bootstrap replication-derived model is then applied to the original data set; (5) differences in model performance between the models estimated in Steps 3 and 4 are calculated and averaged (the resulting average metric reflects the optimism of the original model derived in Step 1); and (6) optimism-corrected performance estimates are calculated by subtracting the optimism metric found in Step 5 from the performance metrics achieved by the original model estimated in Step 1. All performance metrics are reported with this correction taken into account.

Prevalence-sensitive performance evaluation

PPV [$TP/(TP + FP)$] and NPV [$TN/(TN + FN)$] were additionally calculated for each of the three models. PPV and NPV reflect the probability of true positive and true negative tests, respectively, and unlike sensitivity and specific, depend on the prevalence of the condition(s) in question in the setting where the classifier is applied. To evaluate classifier performance as a function of differential CFS prevalence across a variety of potential application contexts, PPV and NPV were calculated with incrementally varying prevalence in equal increments from 10 to 90%, including 1 and 99%. Doing so allows for intuitive evaluation of model performance in different settings (e.g., screening of the general population or acting as a diagnostic marker

in the clinic) in which the prevalence of CFS may vary greatly (Robinson et al. 2016).

Results

Participant characteristics

Participant demographics are displayed in Table 2. HC and CFS patients did not differ significantly in age ($t_{31} = 0.28$, $p = .78$). HC reported more years of education than CFS patients ($t_{31} = 2.36$, $p = .03$). All subjects were female with the exception of two males in the CFS group. In the CFS group, 83.3% identified as white, 11.1% as hispanic, and 5.6% as black. In the HC group, 80% of participants identified as white, 13.3% as hispanic, and 6.7% black.

Model classification performance

Performance metrics for the primary models tested (sMRI and SR) are displayed in Table 5; all estimates displayed are optimism corrected. The model produced for the a priori sMRI feature set achieved an accuracy of 79.58% (sensitivity = 82.00%, specificity = 76.70%; see Table 3 for feature weights). The SR feature set achieved considerable greater values on all performance metrics (95.59% accuracy, sensitivity = 95.44%, and specificity = 95.77%). Mean VAS ratings for each group on all six measures are displayed in Table 4.

To further clarify model performance, we also assessed the performance of a whole brain sMRI feature set with the same modelling approach (total number of features = 235, sMRI-W). This model achieved nearly identical performance to our original, *a priori*-based sMRI model (accuracy = 79.44%, sensitivity = 82.00%, specificity = 76.70%, for feature weights refer to supplemental material Table S1).

Feature importance

Feature weights are squared to aid in assessing importance in classification and are displayed in Table 3 for both models.

sMRI Model There was a greater representation of the left hemisphere regions among those most highly ranked. These included the left caudal anterior cingulate cortex (area and

Table 2 Demographic characteristics by group

	CFS	HC
Age (mean ± SD)	48.33 (12.67)	47.13 (11.67)
Education (mean ± SD)	14.25 (2.43)	16.00 (1.66)
Sex (F/M)	16/2	15/0

CFS chronic fatigue syndrome, *HC* healthy controls

Table 3 SVM feature weights

Feature set		
Feature	Weight	Rank
SR		
Sleep quality	0.179	1
Pain intensity	0.159	2
Depression	− 0.020	3
Anger	− 0.016	4
Anxiety	− 0.009	5
sMRI		
Left caudal anterior cingulate area	0.86	1
Left postcentral thickness	0.85	2
Left caudal anterior cingulate volume	0.71	3
Right superior temporal thickness	0.68	4
Left insula thickness	0.61	5
Left rostral middle frontal area	0.57	6
Left rostral middle frontal volume	0.57	7
Left lingual volume	0.56	8
Left caudate	0.53	9
Left lingual area	0.52	10
Right caudal anterior cingulate volume	0.43	11
Left precuneus thickness	0.42	12
Right superior parietal thickness	0.42	13
Right pallidum	0.41	14
Right lateral occipital area	0.41	15
Right superior parietal volume	0.40	16
Right caudal anterior cingulate thickness	0.40	17
Left caudal middle frontal area	0.39	18
Left pallidum	− 0.38	19
Left postcentral volume	0.37	20
Left caudal middle frontal volume	0.37	21
Right lateral occipital volume	0.35	22
Right superior temporal volume	0.34	23
Left superior frontal area	0.34	24
Right superior parietal area	0.33	25
Left posterior cingulate area	0.31	26
Left rostral anterior cingulate area	0.29	27
Left precuneus volume	0.29	28
Left insula volume	0.28	29
Right caudal anterior cingulate area	0.27	30
Right superior temporal area	0.27	31
Left postcentral area	0.26	32
Left posterior cingulate volume	0.26	33
Left insula area	0.26	34
Left superior frontal volume	0.24	35
Left rostral anterior cingulate volume	0.24	36
Right rostral anterior cingulate volume	0.24	37
Right precuneus area	0.23	38
Right inferior parietal area	0.23	39
Left rostral anterior cingulate thickness	0.23	40
Right precuneus volume	0.21	41

Table 3 (continued)

Feature set		
Feature	Weight	Rank
Left precuneus area	0.20	42
Left parahippocampal volume	− 0.20	43
Right inferior parietal volume	0.18	44
Right cuneus area	− 0.17	45
Left caudal middle frontal thickness	0.16	46
Left caudal anterior cingulate thickness	0.15	47
Right lateral occipital thickness	0.12	48
Left rostral middle frontal thickness	0.12	49
Left parahippocampal area	0.11	50
Left posterior cingulate thickness	− 0.11	51
Right rostral anterior cingulate thickness	0.09	52
Left lingual thickness	0.08	53
Right cuneus volume	− 0.08	54
Left parahippocampal thickness	− 0.08	55
Right cuneus thickness	− 0.07	56
Right inferior parietal thickness	0.07	57
Right rostral anterior cingulate area	0.06	58
Left superior frontal thickness	0.04	59
Right hippocampus	0.03	60
Right precuneus thickness	0.00	61

sMRI structural MRI feature set, *SR* self-report feature set

volume; cACC), post central gyrus thickness, rostral middle frontal cortex (area and volume), caudate, lingual gyrus (volume and area), and insular cortex. The right superior temporal lobe thickness, cACC volume, and superior parietal thickness were also particularly salient in the classification of patients.

Self-Report Model For the SR Model, the most salient feature weights were sleep quality and pain intensity (Weights for all features are displayed in Table 3). Their contribution to participant classification exceeded that of depression, anger, and anxiety ratings by nearly an order of magnitude.

Prevalence-sensitive performance evaluation

PPV and NPV values for each model can be seen in Table 5. PPV and NPV were calculated for each feature set with prevalence varying from 1 to 99% (Table 6). In general, as prevalence increases, PPV increases and NPV decreases. When comparing feature sets, the sSR feature set performed the best, followed by sMRI across all prevalence rates.

Discussion

To our knowledge, this study represents the first investigation of the utility of sMRI features for the classification of CFS patients versus HC (Fig. 1). We used a bootstrap-based

Table 4 Mean VAS ratings by group

	HC mean (SD)	CFS mean (SD)	t_{31}	p	Cohen's d
Fatigue	7.87 (9.35)	52.33 (20.17)	7.85	< .0001	2.83
Sleep quality	15.33 (21.81)	56.11 (19.32)	5.70	< .0001	1.98
Pain intensity	1.87 (5.46)	43.28 (22.39)	6.98	< .0001	2.54
Depression	4.47 (15.67)	32.44 (26.42)	3.60	.001	1.29
Anxiety	2.67 (6.11)	39.94 (27.65)	5.10	< .0001	1.87
Anger	0.13 (0.52)	24.11 (24.22)	4.20	.0006	1.40

HC healthy control, *CFS* chronic fatigue syndrome

Table 5 Classification results

Feature set	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
sMRI	79.58	82.00	76.70	81.96	76.70
SR	95.59	95.44	95.77	96.74	95.43

PPV positive predictive value, *NPV* negative predictive value, *sMRI* structural MRI feature set, *SR* self-report feature set

optimism correction, which allowed us to create a model that both uses all available data and maintains minimal bias is performance estimation. Our data suggest that sMRI features alone are able to classify CFS patients with ~80% accuracy, which is comparable to that obtained for other chronic musculoskeletal pain conditions, including FM (Robinson et al. 2015), irritable bowel syndrome (Labus et al. 2015), chronic low back pain (Baliki et al. 2011; Ung et al. 2014), and knee osteoarthritis (Baliki et al. 2011). Results were nearly identical when classifier performance was assessed on a reduced sample excluding the two male CFS patients and another randomly determined CFS patient, resulting in equivalent sample size and sex composition between groups. As expected, the modified models displayed a comparative decrease in sensitivity and increase in specificity in relation with the original estimates (see the Appendix for additional details). This suggests that the slightly unbalanced sample size and inclusion of men in the CFS group did not meaningfully bias classifier performance. Results also suggest

that performance is largely unchanged with the inclusion of a whole brain feature set versus our theoretically derived feature set. However, future replication of our model performance in other samples is needed.

When interpreting SVM feature weights, it is important to emphasize that these cannot be interpreted in the same way as voxel weights in standard, GLM-based mass-univariate analyses. SVM feature weights do not support inferences regarding underlying biological processes that separate groups (Haufe et al. 2014). Rather, these weights exist in the context of the entire multivariate pattern, in which the direction and magnitude of the weight indicate influence on classifier decision (Mansson et al. 2015). Furthermore, although weights assigned to individual features reflect the relative importance of those features in calculating the dot product used to predict class membership (i.e., which side of the hyperplane each case falls on), they do not necessarily reflect the biological states that may distinguish CFS patients from healthy controls. They also do not imply that features with relatively lower weights do not meaningfully contribute to classification decisions. Features were selected from the previous work that displayed structural and functional connectivity abnormalities in CFS patients (Okada et al. 2004; Boissoneault et al. 2016, 2018; Gay et al. 2016; Wortinger et al. 2016). Performance metrics for the sMRI model suggest initial convergence among those regions that display abnormalities and those that aid in structure-based classification of CFS patients. As noted above, although all features contributed to the observed performance estimates,

Table 6 Predictive values by model accounting for variable prevalence

Prevalence (%)											
	1	10	20	30	40	50	60	70	80	90	99
Positive predictive value											
sMRI	3.43	28.11	46.80	60.13	70.12	77.87	84.07	89.14	93.34	96.94	99.71
SR	18.56	71.48	84.94	90.63	93.77	95.76	97.13	98.14	98.90	99.51	99.99
Negative predictive value											
sMRI	99.76	97.46	94.46	90.86	86.47	80.99	73.96	64.61	51.58	32.13	4.13
SR	99.95	99.47	98.82	98.00	96.92	95.45	93.33	90.00	84.00	70.00	17.50

sMRI structural MRI feature set, *SR* self-report feature set, *sMRI + SR* combined structural MRI and self-report feature set
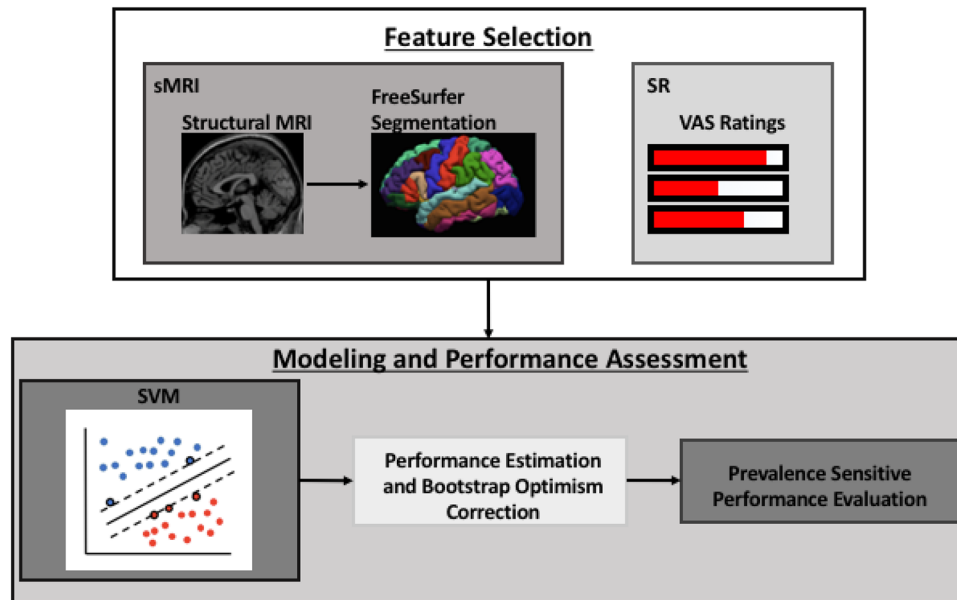
Fig. 1 Analysis pipeline used to classify CFS patients and HC. Two models using different features were compared: (1) FreeSurfer segmented regional estimates based upon previously detected functional abnormalities and (2) clinical and affective VAS ratings. Across feature sets, classification was performed with SVM and performance evaluated following bootstrap optimism correction. PPV and NPV were then calculated based upon incrementally increasing estimates of CFS prevalence to aid in assessing clinical utility of the results. *sMRI* structural MRI feature set, *VAS* visual analogue scale, *SR* self-report VAS rating feature set, *SVM* support vector machine, *PPV* positive predictive values, *NPV* negative predictive value, *CFS* chronic fatigue syndrome, *HC* healthy control

certain features exerted much greater influence on classification than others.

While the previous studies have identified global and focal brain abnormalities in brain perfusion, function, and structure in CFS patients (Okada et al. 2004; Boissoneault et al. 2016, 2018; Gay et al. 2016; Wortinger et al. 2016), it appears that a subset of these regions is most salient in the classification of CFS patients from healthy controls. Among these, we found that regions that most contribute to classification include aspects of midline cortical and subcortical, parietal, temporal, and occipital structures. Taken together, these regions subsume multiple functions including cognitive, emotional, reward, sensory, visual, and regulatory processes. This constellation is consistent with the range of CFS symptomatology. Functional perturbations relative to HC in these regions during wakeful rest (Boissoneault et al. 2016; Gay et al. 2016) and task performance (Miller et al. 2014) have been previously reported in CFS.

For example, a third of the 15 highest ranked features were estimates from the cACC and rostral middle frontal gyrus. These regions are involved in both cognitive and affective processes, and may be particularly important for the integration of these elements of experience. The rostral medial aspect of the PFC may be particularly involved in inhibitory functions of both cognitive and affective processes, while the cACC, particularly salient to CFS, is involved in cognitive aspects of movement, including intention (Hoffstaedter et al. 2014). In CFS patients, our group previously identified resting-state perturbations in functional connectivity among the precuneus and basal ganglia structures (Boissoneault et al. 2018), which may be related to abnormalities in attention switching or reward-related gating during attentional switching. The potential importance of these processes is confirmed in relatively high SVM weights assigned to these structures in the present analysis. Other studies have also demonstrated structural, functional, and neurometabolic perturbations in occipital lobe structures in CFS patients (Puri et al. 2002, 2012; Boissoneault et al. 2016). Our findings suggest that these structures may further aid in discriminating from among patients and controls. It is also notable that among the highest ranked features, there was a greater representation of the left hemisphere than right hemisphere estimates. Future studies should examine lateralization of structural abnormalities in chronic fatigue syndrome to aid in identifying potential hemisphere-specific discriminatory mechanisms.

As previously noted, we also tested a classification model using SR measures of affect and pain. Clinical utility of this classifier is limited, because CFS diagnosis was confirmed as part of the study screening process. Thus, the SR model used in this investigation was intended primarily as a comparator for the sMRI model, so that its potential utility in clinical settings could be gauged. Classification of CFS patients with this model was approximately 96% across performance

indices. This is consistent with a previous report in which SR measures achieved ~ 85–95% classification accuracy distinguishing patients with fibromyalgia syndrome from HC (Robinson et al. 2015). Consistent with the diagnostic criteria for CFS, which emphasize musculoskeletal pain and sleep disruption as secondary symptoms of CFS (Fukuda et al. 1994), SR measures of pain intensity and sleep quality were each strong contributors to the SR-only model. The SR model also performed considerably better than the sMRI model. This suggests that while sMRI features aid in classification, there is additional information vital to classification that is not adequately captured in the sMRI feature set. Combinations of both functional and structural features may improve performance and give key mechanistic insights.

Finally, the utility and accuracy of any ML classifier also depend on the prevalence of the condition of interest in a particular setting (i.e., the base rate, or overall likelihood of encountering an individual with the condition), and will directly affect the PPV and NPV. For instance, the base rate of patients with a specific condition is likely much higher in a specialty clinic than among the general population (Robinson et al. 2016). In high prevalence settings, a biomarker will have an excellent PPV, but poor NPV even if specificity is very high. For this reason, PPV and NPV were calculated for each model on a spectrum of CFS prevalence ranging from 1 to 99%. For reference, estimates of the prevalence of CFS in the general population range from 0.07 to 2.8% (Afari and Buchwald 2003; Vincent et al. 2012). Our results indicate that none of the models performed well at extremely high or low base rates due to expected drops in NPV and PPV, respectively. However, PPV and NPV for the SR model were above 80% for the range of base rates from 20 to 80%. It is obvious from these data that extremely accurate models will be needed to function as useful classifiers in settings where CFS is either very common (e.g., the clinic) or rare (e.g., the general population).

Study limitations and future directions

Because of this study's cross-sectional nature, we were unable to determine whether perturbations in the structural features contributing to the sMRI model are cause or consequence of CFS. Longitudinal studies are required to clarify whether individuals who exhibit similar abnormalities are predisposed to developing CFS or whether these abnormalities are a consequence of CFS pathology. Our results provide additional support for the role of structural classification of CFS in regions with previously demonstrated functional alterations (Boissoneault et al. 2016, 2018; Gay et al. 2016). These regions may serve as targets for future neurometabolic studies to identify any underlying abnormalities. Similar to many classification studies of other chronic pain conditions [e.g., (Ung et al. 2014; Robinson et al. 2015)], the small

sample size of our current investigation may represent a limitation due to the risk of model overfit. However, optimism correction adds confidence to these results and decreases the likelihood of overfitting. Nevertheless, replication and extension of our study with a larger sample are needed. With the present samples, it is impossible to determine whether the features found to be important to classification are unique to CFS. Studies attempting to classify between patient groups will be necessary to increase model validity. In addition, given that a subset of CFS patients was misclassified by our sMRI models, it is possible that clinically meaningful subgroups exist within this population. Reliable identification of unbiased subgroups, however, is not feasible with the current sample size. Future studies will have to identify and validate subgroups within the CFS population based upon degree of morphological abnormality or with the inclusion of brain regions not assessed in the current investigation.

Conclusions

The results of our study demonstrate that structural brain features derived from frontal cortical and subcortical, parietal, basal ganglia, and occipital regions can be used to classify patients with CFS using ML algorithms. In addition, our findings may generate new insights into the pathogenesis of CFS. Our sMRI classification model identified CFS patients with high accuracy. Model built on self-report features was found to outperform those built on structural estimates.

References

Afari N, Buchwald D (2003) Chronic fatigue syndrome: a review. Am J Psychiatry 160:221–236

Alpaydin E (2014) Introduction to machine learning. MIT Press, Cambridge

Bagarinao E, Johnson KA, Martucci KT et al (2014) Preliminary structural MRI based brain classification of chronic pelvic pain: A MAPP network study. Pain 155:2502–2509. https://doi.org/10.1016/j.pain.2014.09.002

Baliki MN, Schnitzer TJ, Bauer WR, Apkarian AV (2011) Brain morphological signatures for chronic pain. PLoS One 6:e26010. https://doi.org/10.1371/journal.pone.0026010

Boissoneault J, Letzen J, Lai S, O'Shea A, Craggs J, Robinson ME, Staud R (2016) Abnormal resting state functional connectivity in patients with chronic fatigue syndrome: an arterial spin-labeling fMRI study. Magn Reson Imaging 34:603–608. https://doi.org/10.1016/j.mri.2015.12.008

Boissoneault J, Letzen J, Lai S, Robinson ME, Staud R (2018) Static and dynamic functional connectivity in patients with chronic fatigue syndrome: use of arterial spin labelling fMRI. Clin Physiol Funct Imaging 38:128–137. https://doi.org/10.1111/cpf.12393

Cenzer I, Miao Y, Kirby K, Boscardin WJ (2013) Estimating Harrell's optimism on predictive indices using bootstrap samples. SAS Global Forum 2013

Chen R, Liang FX, Moriya J, Yamakawa J, Sumino H, Kanda T, Takahashi T (2008) Chronic fatigue syndrome and the central nervous system. J Int Med Res 36:867–874. https://doi.org/10.1177/147323000803600501

Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, Irwing L, Levine D, Reitsma JB, de Vet HC, Bossuyt PM (2016) STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 6:e012799

de Lange FP, Kalkman JS, Bleijenberg G, Hagoort P, van der Meer JW, Toni I (2005) Gray matter volume reduction in the chronic fatigue syndrome. Neuroimage 26:777–781. https://doi.org/10.1016/j.neuroimage.2005.02.037

de Lange FP, Koers A, Kalkman JS, Bleijenberg G, Hagoort P, van der Meer JW, Toni I (2008) Increase in prefrontal cortical volume following cognitive behavioural therapy in patients with chronic fatigue syndrome. Brain 131:2172–2180. https://doi.org/10.1093/brain/awn140

Fan R, Chan P, Lin C (2005) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:27

Fischl B (2012) FreeSurfer. Neuroimage 62:774–781

Fischl B, Salat DH, Busa E et al (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33:341–355

Friedman J, Hastie T, Tibshirani R (2009) The elements of statistical learning, 2nd edn. Springer, New York

Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A (1994) The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. Ann Intern Med 121:953–959

Gay CW, Robinson ME, Lai S, O'Shea A, Craggs JG, Price DD, Staud R (2016) Abnormal resting-state functional connectivity in patients with chronic fatigue syndrome: results of seed and data-driven analyses. Brain Connect 6:48–56. https://doi.org/10.1089/brain.2015.0366

Grimes DA, Schulz KF (2005) Refining clinical diagnosis with likelihood ratios. Lancet 365:1500–1505. https://doi.org/10.1016/S0140-6736(05)66422-7

Gunn WJ, Connell DB, Randall B (1993) Epidemiology of chronic fatigue syndrome: the Centers for Disease Control Study. Ciba Found Symp 173:83–93; (discussion 93–101)

Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46:389–422

Haufe S, Meinecke F, Gorgen K, Dahne S, Haynes JD, Blankertz B, Biessmann F (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87:96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Hoffstaedter F, Grefkes C, Caspers S et al (2014) The role of anterior midcingulate cortex in cognitive motor control: evidence from functional connectivity analyses. Hum Brain Mapp 35:2741–2753. https://doi.org/10.1002/hbm.22363

Hsu C, Chang C, Lin C (2008) A practical guide to support vector classification. BJU Int 101:1396–1400

Jovicich J, Czanner S, Han X et al (2009) MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 46:177–192. https://doi.org/10.1016/j.neuroimage.2009.02.010

Labus JS, Van Horn JD, Gupta A et al (2015) Multivariate morphological brain signatures predict patients with chronic abdominal pain from healthy control subjects. Pain 156:1545–1554. https://doi.org/10.1097/j.pain.0000000000000196

Lopez Puga J, Krzywinski M, Altman N (2015) Points of significance: Bayes' theorem. Nat Methods 12:277–278

Mansson KN, Frick A, Boraxbekk CJ et al (2015) Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. Transl Psychiatry 5:e530. https://doi.org/10.1038/tp.2015.22

Miller AH, Jones JF, Drake DF, Tian H, Unger ER, Pagnoni G (2014) Decreased basal ganglia activation in subjects with chronic fatigue syndrome: association with symptoms of fatigue. PLoS One 9:e98156. https://doi.org/10.1371/journal.pone.0098156

Nouretdinov I, Costafreda SG, Gammerman A, Chervonenkis A, Vovk V, Vapnik V, Fu CH (2011) Machine learning classification with confidence: application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. Neuroimage 56:809–813. https://doi.org/10.1016/j.neuroimage.2010.05.023

Okada T, Tanaka M, Kuratsune H, Watanabe Y, Sadato N (2004) Mechanisms underlying fatigue: a voxel-based morphometric study of chronic fatigue syndrome. BMC Neurol 4:14

Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45:S199-209. https://doi.org/10.1016/j.neuroimage.2008.11.007

Price DD, Harkins SW (1987) Combined use of visual analogue scales and experimental pain in providing standardized measurement of clinical pain. Clin J Pain 3:1–8

Puri BK, Counsell SJ, Zaman R, Main J, Collins AG, Hajnal JV, Davey NJ (2002) Relative increase in choline in the occipital cortex in chronic fatigue syndrome. Acta Psychiatr Scand 106:224–226

Puri BK, Jakeman PM, Agour M et al (2012) Regional grey and white matter volumetric changes in myalgic encephalomyelitis (chronic fatigue syndrome): a voxel-based morphometry 3 T MRI study. Br J Radiol 85:e270-273. https://doi.org/10.1259/bjr/93889091

Radloff LS (1977) The CES-D Scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1:385–401

Robinson ME, O'Shea AM, Craggs JG, Price DD, Letzen JE, Staud R (2015) Comparison of machine classification algorithms for fibromyalgia: neuroimages versus self-report. J Pain 16:472–477. https://doi.org/10.1016/j.jpain.2015.02.002

Robinson M, Boissoneault J, Sevel L, Letzen J, Staud R (2016) The effect of base rate on the predictive value of brain biomarkers. J Pain 17:637–641. https://doi.org/10.1016/j.jpain.2016.01.476

Smith GC, Seaman SR, Wood AM, Royston P, White IR (2014) Correcting for optimistic prediction in small data sets. Am J Epidemiol 180:318–324. https://doi.org/10.1093/aje/kwu140

Staud R, Mokthech M, Price DD, Robinson ME (2015) Evidence for sensitized fatigue pathways in patients with chronic fatigue syndrome. Pain 156:750–759. https://doi.org/10.1097/j.pain.0000000000000110

Ung H, Brown JE, Johnson KA, Younger J, Hush J, Mackey S (2014) Multivariate classification of structural MRI data detects chronic low back pain. Cereb Cortex 24:1037–1044. https://doi.org/10.1093/cercor/bhs378

Vincent A, Brimmer DJ, Whipple MO et al (2012) Prevalence, incidence, and classification of chronic fatigue syndrome in Olmsted County, Minnesota, as estimated using the Rochester Epidemiology Project. Mayo Clin Proc 87:1145–1152. https://doi.org/10.1016/j.mayocp.2012.08.015

Walsh CG, Ribeiro JD, Franklin JC (2017) Predicting risk of suicide attempts of time through machine learning. Clin Psychol Sci 5:457–469

Wortinger LA, Endestad T, Melinder AM, Oie MG, Sevenius A, Bruun Wyller V (2016) aberrant resting-state functional connectivity in the salience network of adolescent chronic fatigue syndrome. PLoS One 11:e159351. https://doi.org/10.1371/journal.pone.0159351