The Journal of Pain, Vol 16, No 5 (May), 2015: pp 472-477
Available online at www.jpain.org and www.sciencedirect.com

Comparison of Machine Classification Algorithms for Fibromyalgia: Neuroimages Versus Self-Report

Michael E. Robinson,* Andrew M. O'Shea,* Jason G. Craggs,* Donald D. Price,† Janelle E. Letzen,* and Roland Staud‡

*Departments of *Clinical and Health Psychology and †Oral and Maxillofacial Surgery, University of Florida, Gainesville, Florida.*
‡Division of Rheumatology and Clinical Immunology, College of Medicine, University of Florida, Gainesville, Florida.

Abstract: Recent studies have posited that machine learning (ML) techniques accurately classify individuals with and without pain solely based on neuroimaging data. These studies claim that self-report is unreliable, making "objective" neuroimaging classification methods imperative. However, the relative performance of ML on neuroimaging and self-report data have not been compared. This study used commonly reported ML algorithms to measure differences between "objective" neuroimaging data and "subjective" self-report (ie, mood and pain intensity) in their ability to discriminate between individuals with and without chronic pain. Structural magnetic resonance imaging data from 26 individuals (14 individuals with fibromyalgia and 12 healthy controls) were processed to derive volumes from 56 brain regions per person. Self-report data included visual analog scale ratings for pain intensity and mood (ie, anger, anxiety, depression, frustration, and fear). Separate models representing brain volumes, mood ratings, and pain intensity ratings were estimated across several ML algorithms. Classification accuracy of brain volumes ranged from 53 to 76%, whereas mood and pain intensity ratings ranged from 79 to 96% and 83 to 96%, respectively. Overall, models derived from self-report data outperformed neuroimaging models by an average of 22%. Although neuroimaging clearly provides useful insights for understanding neural mechanisms underlying pain processing, self-report is reliable and accurate and continues to be clinically vital.

***Perspective:** The present study compares neuroimaging, self-reported mood, and self-reported pain intensity data in their ability to classify individuals with and without fibromyalgia using ML algorithms. Overall, models derived from self-reported mood and pain intensity data outperformed structural neuroimaging models.*

***Key words:** Machine learning, magnetic resonance imaging, self-report, fibromyalgia, pain biomarkers.*

Although neuroimaging was initially a tool for exploring mechanisms of pain processing, the use of neuroimaging to diagnose or detect pain conditions has become an important focus of research. A strong emphasis has been placed on classifying individuals into patient or control groups based on neuroimaging data. These classification studies typically employ sophisticated multivariate statistical approaches, which are said to provide empirically derived algorithms to discriminate between individuals with and without pain. A number of these studies have even suggested that these indices reflect "objective biomarkers" of pain, or act as a surrogate for patients' self-report.[5,6,21,22,24]

Proponents of neural "biomarkers" argue that self-report is unreliable, making objective markers of pain imperative.[3,14,25] However, implied in those assumptions would be the conclusion that brain imaging is more reliable and thus would outperform self-report in classifying individuals to patient or control samples. Regarding this question of reliability, we previously demonstrated that functional neuroimaging (ie, functional magnetic resonance imaging [fMRI]) data fell within a "good" range of reliability, whereas

Received October 21, 2014; Revised January 31, 2015; Accepted February 4, 2015.
This research was supported by grants 5R01AT001424 and 5R01NS038767 from the National Institutes of Health. The authors have no conflicts of interest to declare.
Address reprint requests to Michael E. Robinson, PhD, University of Florida, 1225 Center drive, Room 3131, P.O. Box 100165, Gainesville, FL 32610-0165. E-mail: merobin@ufl.edu
1526-5900/$36.00

http://dx.doi.org/10.1016/j.jpain.2015.02.002

the participants' self-reported pain ratings fell within an "excellent" range of reliability.[13] This finding corroborates our argument that the question of self-report reliability is unsupported and acts as a limitation for using machine learning (ML) classification indices for diagnosing or detecting pain.

In addition to directly comparing fMRI and self-report reliability, we previously discussed theoretical, philosophical, and measurement theory–based limitations of using neuroimaging to discriminate between individuals with pain conditions and those without pain or to provide a substitute for self-report measures of pain.[19] Additionally, we disputed a number of assumptions used by proponents of brain-based classification approaches, including the reliability of self-report, the objectiveness of brain images and self-report, the validation and measurement properties of self-report and brain images, and finally the philosophical issues surrounding the substitute of brain images for self-report.[19] Although claims made by neuroimaging classification studies have important clinical implications, these methods have not directly tested whether neuroimaging data outperform self-report within this context. As such, there is a compelling need to empirically assess the relative performance of brain-based indices compared to self-report indices for the discrimination of individuals with and without pain.

In this study, we directly employ multivariate ML approaches to compare classification rates between neuroimage indices and self-report measures obtained within the same individuals during the same study visit. We tested several models commonly used in previous studies of neuroimaging classification for pain conditions on structural neuroimages, as well as self-report data of pain intensity and mood.

Methods

Participants and Study Procedures

Fourteen women diagnosed with fibromyalgia (FM; mean age = 44.1 years) according to the American College of Rheumatology criteria,[26] as determined by the study's rheumatologist, were recruited from the University of Florida and surrounding community. Twelve age- and sex-matched healthy, pain-free controls (mean age = 42.2 years) were also recruited from the community. This study was approved by the University of Florida's institutional review board, and participants provided written informed consent for their participation.

Neuroimaging Data

T1-weighted structural MRI scans were acquired from all participants using a magnetization-prepared rapid gradient-echo (MPRAGE) imaging scanning protocol: 170 sagittal slices of 1 mm, matrix = $256 \times 256 \times 170$ mm, repetition time = 8.1 milliseconds, echo time = 3.7 milliseconds, field of view = $240 \times 240 \times 170$ mm, flip angle = 8°, voxel size = 1 mm^3. Data were processed through the automated subcortical segmentation stream in FreeSurfer, version 5.1.0 (Martinos Center for Biomedical Imaging, Charlestown, MA),[7] which was used to measure volumes of 55 neuroanatomic regions that were included for further analysis with our ML algorithms (Table 1). The software takes into account aspects of the collected MRI data and previously established characteristics of MRI data in general (eg, signal intensity information of subcortical vs cortical brain regions) to determine the probability that each discrete neuroanatomic region is correctly labeled.[7] Previous research has shown that this automated procedure produces accurate and reliable results and is a popular method of segmentation within the field.[7,9]

Self-Report Data

Self-report data of mood and pain intensity were collected using visual analog scales on the day of the MRI. Visual analog scale ratings were acquired for 5 mood variables (ie, depression, anxiety, frustration, anger, and fear) and pain intensity, for a total of 6 visual analog scale ratings. Mood was chosen as a feature of interest because there is a strong association between mood disturbance and individuals with FM.[2]

ML Model Preparation

ML is an increasingly popular method of classifying data into discrete groups. The input for classifier functions is a set of examples, called features (ie, independent variable), and the outputs are a class (ie, dependent variable), or discrete group, that the example belongs to.[16] To build each model, a matrix including the number of features, or input variables, must be constructed. For the present study, the following matrices were used: Brain volumes $\times$ Participant (55×26), Mood $\times$ Participant (5×26), and Pain intensity $\times$ Participant (1×26).

In building our model, we took 2 aspects of ML into consideration: 1) supervised attribute selection and 2) the "curse of dimensionality." Supervised attribute selection is a form of data processing that uses the same data to "train" the learning classifiers. Although occasionally used on ML data sets, we did not perform supervised attribute selection because it has been shown to yield optimistically biased classification results.[20] Additionally, we created a data set to specifically mimic a common phenomenon in ML called the "curse of dimensionality," or finding a balance between having enough features for accurate classification and oversaturating the model. This data set contained 55 features and included the 5 mood features and 50 pseudo-random numbers ranging from 0 to 100.

Models were then built using 6 learning algorithms, or classifiers, using the software Weka (University of Waikato, New Zealand).[8] We chose the following models because of their popularity among classification papers. First, we used naïve Bayes,[11] which calculates the probability of data belonging to each possible class and assumes independence between predictors. Second, we used a logistic regression with a ridge estimator,[12] which takes a linear combination of predictors and regression coefficients to predict a