Published in final edited form as:

J Pain. 2015 May ; 16(5): 472–477. doi:10.1016/j.jpain.2015.02.002.

Comparison of machine classification algorithms for fibromyalgia: Neuroimages versus self-report

Michael E. Robinson, Ph.D.^a, Andrew M. O'Shea, MSc.^a, Jason Craggs, Ph.D.^a, Donald D. Price, Ph.D.^b, Janelle E. Letzen, M.S.^a, and Roland Staud, M.D.^c

^aDepartment of Clinical and Health Psychology, University of Florida, 1225 Center drive, Gainesville, FL, 32611

^bDepartment of Oral and Maxillofacial Surgery, University of Florida, 1395 Center drive, Gainesville, FL, 32603

^cCollege of Medicine, University of Florida, 1600 SW Archer road, Gainesville, FL, 32603

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 has not been compared. This study used commonly reported ML algorithms to measure differences between “objective” neuroimaging data and “subjective” self-report (i.e., mood and pain intensity) in their ability to discriminate between individuals with and without chronic pain. Structural MRI data from 26 individuals (14 individuals with fibromyalgia, 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 (i.e., anger, anxiety, depression, frustration, 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 53–76%, whereas mood and pain intensity ratings ranged from 79–96% and 83–96%, 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, accurate, and continues to be clinically vital.

Keywords

machine learning; MRI; self-report; fibromyalgia; pain biomarkers

© 2015 Published by The American Pain Society.

Corresponding Author: Michael E. Robinson, Ph.D., University of Florida, 1225 Center drive, Room 3131, P.O. Box 100165, Gainesville, FL 32610-0165, Tel: 352-273-6153, merobin@ufl.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure: The authors have no conflicts of interest to declare.

Introduction

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^{3,17,25}. 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 (fMRI) data fell within a “good” range of reliability, whereas the participants’ self-reported pain ratings fell within an “excellent” range of reliability¹³. This finding corroborates our argument that the question of self-report reliability is unsupported, and acts as a limitation for using machine learning 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 compared to those without pain, or to provide a substitute for self-report measures of pain¹⁹. 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-report¹⁹. Although claims made by neuroimaging classification studies have important clinical implications, these methods have not directly tested whether neuroimaging data outperforms 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 machine learning 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 females diagnosed with fibromyalgia (FM; mean age = 44.1 years) according to the American College of Rheumatology criteria²⁶, 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 an MPRAGE scanning protocol: 170 1-mm sagittal slices, matrix (mm) = $256 \times 256 \times 170$, repetition time = 8.1 ms, echo time = 3.7 ms, field of view (mm) = $240 \times 240 \times 170$, flip angle = 8° , voxel size = 1 mm^3 . Data were processed through the automated subcortical segmentation stream in FreeSurfer V.5.1.0⁷, which was used to measure volumes of 55 neuroanatomical regions that were included for further analysis with our machine learning (ML) algorithms (Table 1). The software takes into account aspects of the collected MRI data, as well as previously established characteristics of MRI data in general (e.g., signal intensity information of subcortical vs. cortical brain regions), to determine the probability that each discrete neuroanatomical region is correctly labeled⁷. Previous research has shown 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 (VAS) on the day of the MRI. VAS ratings were acquired for five mood variables (i.e., depression, anxiety, frustration, anger, and fear), as well as pain intensity, for a total of six VAS ratings. Mood was chosen as a feature of interest because there is a strong association between mood disturbance and individuals with fibromyalgia².

Machine Learning Model Preparation

Machine learning is an increasingly popular method of classifying data into discrete groups. The input for classifier functions is a set of examples, called features (i.e., independent variable), and the outputs are a class (i.e., dependent variable), or discrete group that the example belongs to¹⁵. 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 $\times$ Participants (1×26).

In building our model, we took two 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 datasets, we did not perform supervised attribute selection, because it has been shown to yield optimistically biased classification results²⁰.

Additionally, we created a dataset 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 dataset contained 55 features and included the five mood features and 50 pseudorandom numbers ranging from 0–100.

Models were then built using six learning algorithms, or classifiers, using the software Weka⁸. We chose the following models due to their popularity among classification papers. First, we used naïve Bayes¹¹, 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 estimator¹², which takes a linear combination of predictors and regression coefficients to predict a categorical class (i.e. patient versus control). Third, we used a 3-nearest-neighbors instance based classifier¹, which examines a specified number of neighbors (i.e., 3) for each feature to determine the categorical class. Fourth, we used a multi-layer perceptron classifier, which is a feed-forward artificial neural network that assumes simple features work in tandem to produce a complex output⁴. Fifth, we used a sequential minimal optimization support vector machine^{10,16}, which aims to find a maximum margin hyperplane, or a subspace of dimension, that separates the classes. Finally, we used a J48 decision tree¹⁸, which uses information between predictors to split classes using the most informative features.

Evaluation of Machine Learning Models

Models were evaluated using ten iterations of ten-fold cross validation on each dataset, with a different random seed chosen each time. Models were evaluated on a variety of statistics, including overall classification accuracy percentage, sensitivity, specificity, kappa, F1, and area under the receiver operating characteristic curve (AUC). To test the relative accuracy of our six learning models, the all were compared to a base rate classifier, which overly simplifies the model to classify 100% of the sample as having FM. In addition to testing our neuroimage dataset, we also used these methods to test our self-report datasets for pain intensity and all five mood features.

Results

Comparison of Classifiers

Our six learning classifiers were first compared to the base rate classifier our neuroimaging, mood, and pain intensity datasets. The neuroimaging dataset resulted in two of six learning classifiers significantly outperforming the base rate classifier ($p < .05$). Classifiers used on the self-reported mood and pain intensity datasets, however, all significantly outperformed the base rate classifier ($p < .05$). Directly comparing the datasets, self-reported pain intensity outperformed the neuroimaging dataset for peak and mean accuracy across all six learning classifiers. Self-report mood and pain intensity datasets were similar in terms of predictive efficacy. Tables 2 and 3 provide a full breakdown of accuracy rates.

The model with the highest accuracy was the J48 decision tree used on the self-reported mood datasets. Specifically, this classifier showed the highest accuracy for correctly classifying individuals into FM or HC using anger, with only one out of 27 FM participants misclassified as a HC. Features were ranked to determine which were most valuable for classification in each dataset. An information gain ratio evaluator¹⁸ implemented in Weka was used to determine rank. In the mood dataset, the most informative feature was anger. In the neuroimaging dataset, the most informative feature was left amygdala volume. Effect

sizes of the most informative features in the neuroimaging and mood datasets were calculated; pain intensity is also included for reference (Table 4).

Relationship between Informative Features and Pain Intensity

Because the diagnosis of FM is currently largely reliant on individuals' report of widespread pain, we wanted to examine whether reported pain intensity was related to the most predictive features identified in our neuroimaging (i.e., left amygdala volume) and self-reported mood (i.e., anger) used to classify participants into diagnostic groups (i.e., FM or HC). Left amygdala volume accounted for approximately 15% of the variance in pain intensity, while self-report anger accounted for approximately 45% of the variance in pain intensity.

Discussion

The present study examined the use of several machine learning (ML) algorithms on structural MRI and self-report (i.e., pain intensity and mood) datasets to classify individuals as belonging to fibromyalgia (FM) or healthy control (HC) groups. Overall, we were able to classify individuals using features, or input variables, within these datasets, but found that self-report features generally outperformed neuroimaging features. Additionally, we found that among neuroimaging features, left amygdala volume was the most predictive feature for classifying individuals, whereas anger was the most predictive self-report feature. To the best of our knowledge, this is the first study to examine FM classification using structural MRI features.

Our results align with previous work showing successful classification using neuroimaging features, and expand on this work by adding self-report features to directly test previous claims that “objective” neural biomarkers could outperform self-report for diagnosis. Ung and colleagues²² previously described successful classification of chronic low back pain from HC with 76% accuracy (26% greater than base rate) using a support vector machine (SVM) model trained by structural MRI features. Additionally, Sundermann and colleagues²¹ reported 73.5% accuracy in classifying FM from HC using SVM classifiers with resting state functional MRI connectivity features. We were able to produce similar results in accurately classifying FM patients and HC using a J48 decision tree (75.50% accuracy; 22.17% above base rate) and SVM (72.17% accuracy; 18.84% above base rate) classifiers tested with 100 iterations of cross-validation.

Although the use of neuroimaging features for FM classification is an interesting proof of concept, the usefulness of clinical translation of these techniques is unknown. An effort has been made to find an “objective” biomarker for diagnosis, while asserting “subjective” methods such as self-report are lacking. This assertion ignores the fact that any objective measures, such as neuroimaging, need to be validated against self-report measures. In a comprehensive review on detecting biomarkers from neuroimaging data to classify psychiatric disorders, Orrù and colleagues¹⁴ raise some important challenges and future considerations. The authors state, “...it is often assumed that neuroimaging would allow more accurate diagnostic and prognostic assessment than demographic, clinical and cognitive information, but no previous studies have examined this. It would therefore be of

great interest to examine the relative diagnostic and prognostic value of neuroimaging, demographic, clinical and cognitive in the main neurological and psychiatric disorder¹⁴. We addressed these points by comparing two distinct types of data from the same individuals, structural neuroimages and self-reported mood. Mood data outperformed brain data for accuracy on every classifier tested. The best performing mood model was 20% more accurate than the best performing neuroimaging model. Even when we added non-informative noise features to equate the mood model with the neuroimaging model in terms of dimensionality, two classifiers were able to overcome the noise and produce accurate classifications at a rate greater than 90%.

Considerations and Limitations

To ensure our data were analyzed in an unbiased and rigorous manner, we avoided procedures that would contribute to possible data leakage from the training to test set, resulting in overly optimistic classification rates. One way we accomplished this was by avoiding supervised feature selection (using class information to guide selection of informative features). Additionally, we opted to use a cross validation approach, instead of partitioning the data into two datasets, due to our small sample size. Third, default model parameters, as set by the software Weka, were kept constant during training to prevent artificially inflating our results. Multiple SVM models using different combinations of parameters may result in wide variations of accuracy. For a practical example related to FM diagnosis, Sunderman and colleagues²¹ report results of various combinations of parameters used in SVM models, with accuracy rates ranging from 0% to 73.5%. We used a nested cross validation approach, in which all tuning steps are repeated in each fold of cross validation, to avoid biased estimates of accuracy²³.

The neuroimaging features in our study were limited to structural MRI data. It is possible that including additional data types, such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI), might have resulted in better accuracy. However, we do not believe that our self-report models showed better accuracy due to a failure of our neuroimaging models. Our support vector machine and J48 decision tree models built on structural neuroimaging features outperformed chance and correctly classified approximately 75% of the sample, performing comparably to previous published structural neuroimaging and resting state functional connectivity predictive models^{21,22}. Furthermore, the high classification accuracy of the self-report measures (best case of 96%, with only 1 misclassification) will be hard to outperform with any set of features.

Conclusions

Accurately classifying FM can be accomplished in the absence of pain information using either self-report mood measures or neuroimaging, although self-report was the best classifier in the present study. There are excellent reasons to study neuroimaging of pain, such as to explore central nervous system processes, elucidate underlying mechanisms involved in pain processing, and develop methods for individuals who are unable to report pain due to limited consciousness or poor neurocognitive status. However, using neuroimaging to classify or diagnose people with and without pain raises serious ethical concerns, and was even less accurate than self-report measures in this study. Furthermore,

these data strongly demonstrate that claims of unreliability or high error in measuring self-report of pain or pain-related experiences are unfounded. The high classification rates using self-report would not be possible with unreliable measures. Lauding neuroimaging as a substitute for pain self-report is not supported empirically since self-report outperforms neuroimage-based classification, and because neuroimaging of pain state is, itself, validated using self-report.

Acknowledgments

This research was supported by grants 5R01AT001424 and 5R01NS038767 from the National Institutes of Health.

References

1. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. *Mach Learn.* 1991; 6(1):37–66.
2. Alciati A, Sgiarovello P, Atzeni F, Sarzi-Puttini P. Psychiatric problems in fibromyalgia: Clinical and neurobiological links between mood disorders and fibromyalgia. *Reumatismo.* 2012; 64(4): 268–274. [PubMed: 23024971]
3. Apkarian AV, Hashmi JA, Baliki MN. Pain and the brain: Specificity and plasticity of the brain in clinical chronic pain. *Pain.* 2011; 152(Suppl 3):S49–S64. [PubMed: 21146929]
4. Arora R, Suman S. Comparative analysis of classification algorithms on different datasets using WEKA. *Int J Comput App.* 2012; 54(13):21–25.
5. Brown JE, Chatterjee N, Younger J, Mackey S. Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. *PLoS One.* 2011; 6(9):e24124. [PubMed: 21931652]
6. Callan D, Mills L, Nott C, England R, England S. A tool for classifying individuals with chronic back pain: Using multivariate pattern analysis with functional magnetic resonance imaging data. *PLoS One.* 2014; 9(6):e98007. [PubMed: 24905072]
7. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe Am Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron.* 2002; 33(3):341–355. [PubMed: 11832223]
8. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter.* 2009; 11(1):10–18.
9. Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B, Pacheco J, Albert M, Killiany R, Blacker D, Maguire P, Rosas D, Makris N, Gollub R, Dale A, Dickerson BC, Fischl B. 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.* 2009; 46(1):177–192. [PubMed: 19233293]
10. Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Comput.* 2001; 13(3):637–649.
11. Langley, P.; Iba, W.; Thompson, K. *Proc Natl Conf Artif Intell.* San Jose, CA: AAAI Press; 1992. An analysis of Bayesian classifiers; p. 223–228.
12. Lecessie S, Vanhouwelingen JC. Ridge estimators in logistic-regression. *Appl Stat.* 1992; 41(1): 191–201.
13. Letzen JE, Sevel LS, Gay CW, O'Shea AM, Craggs JG, Price DD, Robinson ME. Test-retest reliability of pain-related brain activity in healthy controls undergoing experimental thermal pain. *J Pain.* 2014; 15(10):1008–1014. [PubMed: 24998897]
14. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neurosci Biobehav Rev.* 2012; 36(4):1140–1152. [PubMed: 22305994]
15. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage.* 2009; 45(1):S199–S209. [PubMed: 19070668]

16. Platt JC. Sequential minimal optimization: A fast algorithm for training support vector machines. MSR-TR-98-14. 1998
17. [Accessed 2014] National Institutes of Health: Biomarkers for chronic pain using functional brain connectivity. Available at: http://commonfund.nih.gov/planningactivities/socialmedia_summary#biomarkers.
18. Quinlan, JR. C4. 5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann Publishers, Inc.; 1993.
19. Robinson ME, Staud R, Price DD. Pain measurement and brain activity: Will neuroimages replace pain ratings? *J Pain*. 2013; 14(4):323–327. [PubMed: 23548484]
20. Singhi SK, Liu H. Feature subset selection bias for classification learning. *Proc Inter Conf Mach Learn*. 2006:849–856.
21. Sundermann B, Burgmer M, Pogatzki-Zahn E, Gaubitz M, Stüber C, Wessolleck E, Heuft G, Pfleiderer B. Diagnostic classification based on functional connectivity in chronic pain: model optimization in fibromyalgia and rheumatoid arthritis. *Acad Radiol*. 2014; 21(3):369–377. [PubMed: 24507423]
22. Ung H, Brown JE, Johnson KA, Younger J, Hush J, Mackey S. Multivariate classification of structural MRI data detects chronic low back pain. *Cereb Cortex*. 2014; 24(4):1037–1044. [PubMed: 23246778]
23. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform*. 2006; 7(1):91.
24. Wager TD, Atlas LY, Lindquist MA, Roy M, Woo CW, Kross E. An fMRI-based neurologic signature of physical pain. *N Engl J Med*. 2013; 368(15):1388–1397. [PubMed: 23574118]
25. Wartolowska K. How neuroimaging can help us to visualise and quantify pain? *Eur J Pain Suppl*. 2011; 5(S2):323–327.
26. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, Tugwell P, Campbell SM, Abeles M, Clark P, Fam AG, Farber SJ, Flechtner JJ, Franklin CM, Gatter RA, Hamaty D, Lessard J, Lichtbroun AS, Masi AT, McCain GA, Reynolds WJ, Romano TJ, Russell IJ, Sheon RP. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: Report of the multicenter criteria committee. *Arthritis Rheum*. 1990; 33(2):160–172. [PubMed: 2306288]

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 machine learning algorithms. Overall, models trained with self-reported mood and pain intensity data outperformed structural neuroimaging models.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- Machine learning models were used on data from fibromyalgia patients and controls.
- We compared neuroimaging and self-report data as methods of group classification.
- Self-report outperformed neuroimaging data for determining group membership.

Table 1

Neuroanatomical volumes used as features in our neuroimaging dataset. T-statistics listed below brain generated from analysis in FreeSurfer. Bolded values represent significant volumes ($p < .05$).

Region	Hemisphere	T-statistic
Lateral ventricle	Left	-0.89
Inferior lateral ventricle	Left	-1.11
Cerebellum white matter	Left	0.92
Cerebellum cortex	Left	-2.55
Thalamus	Left	2.8
Caudate	Left	-1.1
Putamen	Left	0.79
Pallidum	Left	1.3
3rd ventricle	-	-1.05
4th ventricle	-	0.21
Brainstem	-	-0.67
Hippocampus	Left	-0.25
Amygdala	Left	-2.11
Cerebral spinal fluid	-	0.17
Accumbens area	Left	3.07
Ventral DC	Left	1.59
Vessel	Left	-1.04
Choroid plexus	Left	0.68
Lateral ventricle	Right	-0.48
Inferior lateral ventricle	Right	-0.96
Cerebellum white matter	Right	0.5
Cerebellum cortex	Right	-3.03
Thalamus	Right	3.09
Caudate	Right	-1.33
Putamen	Right	-0.18
Pallidum	Right	2.91
Hippocampus	Right	-0.4
Amygdala	Right	-1.94
Accumbens area	Right	-0.19
Ventral DC	Right	1.51
Vessel	Right	0.94
Choroid plexus	Right	0.35
5th ventricle	-	-1.67
White matter hypointensities	-	1.07
Non-white matter hypointensities	-	-1.01
Optic chiasm	-	-0.69

Region	Hemisphere	T-statistic
Corpus callosum posterior	-	0.27
Corpus callosum mid posterior	-	0.06
Corpus callosum central	-	1.01
Corpus callosum mid anterior	-	1.27
Corpus callosum anterior	-	1.55
Cortex	Left	-1.02
Cortex	Right	-1.19
Cortex	-	-1.11
Cortical white matter	Left	1.15
Cortical white matter	Right	1.15
Cortical white matter	-	1.15
Subcortical gray matter	-	-1.81
Total gray matter	-	-1.48
Supratentorial	-	0.2
Intracranial	-	-0.62
White matter hypointensities	Left	-
Non-white matter hypointensities	Left	-
White matter hypointensities	Right	-
Non-white matter hypointensities	Right	-

Table 2

List of accuracy rates for all 6 classifiers in each dataset. Gray shading represents classifiers that were significantly more accurate ($p < .05$) than the base rate classifier.

ML Classifier	Neuroimaging	Mood	Pseudorandom Mood	Pain Intensity
Base-rate	53.33%	53.33%	53.33%	53.33%
Logistic	63.50%	78.83%	60.17%	88.50%
MLP	68.50%	78.83%	60.17%	88.83%
SMO-SVM	72.17%	85.67%	59.83%	91.33%
IB3	53.33%	86.17%	64.17%	82.67%
J48	75.50%	96.17%	96.17%	95.83%
Naïve Bayes	64.17%	93.50%	90.50%	92.00%

^a Abbreviations: MLP=multilayer perceptron; SMO-SVM=sequential minimum optimization support vector machine; IB3=instance based 3, also referred to as k-nearest neighbors with k=3.

Table 3

Most accurate machine learning classifiers of the mood and neuroimaging datasets.

Measure	Result (J48-Mood)	Result (J48-Brain)
Accuracy	0.96	0.76
Sensitivity (FM)	0.93	0.81
Specificity (FM)	1	0.75
ROC	0.97	0.75
Kappa	0.93	0.50
F1	0.96	0.64

^a Abbreviations: FM=fibromyalgia; ROC=receiver operating characteristic.

Table 4

Effect sizes of most informative features of the mood and neuroimaging datasets, as well as pain intensity.

Feature	Cohen's D
Left amygdala volume	0.66
Anger	1.83
Pain intensity	2.84