Curr Rheumatol Rep (2017) 19: 5
DOI 10.1007/s11926-017-0629-9

CHRONIC PAIN (R STAUD, SECTION EDITOR)

Biomarkers for Musculoskeletal Pain Conditions: Use of Brain Imaging and Machine Learning

Jeff Boissoneault[1] · **Landrew Sevel**[1] · **Janelle Letzen**[1] · **Michael Robinson**[1] · **Roland Staud**[2]

Published online: 31 January 2017

Abstract Chronic musculoskeletal pain condition often shows poor correlations between tissue abnormalities and clinical pain. Therefore, classification of pain conditions like chronic low back pain, osteoarthritis, and fibromyalgia depends mostly on self report and less on objective findings like X-ray or magnetic resonance imaging (MRI) changes. However, recent advances in structural and functional brain imaging have identified brain abnormalities in chronic pain conditions that can be used for illness classification. Because the analysis of complex and multivariate brain imaging data is challenging, machine learning techniques have been increasingly utilized for this purpose. The goal of machine learning is to train specific classifiers to best identify variables of interest on brain MRIs (i.e., biomarkers). This report describes classification techniques capable of separating MRI-based brain biomarkers of chronic pain patients from healthy controls with high accuracy (70–92%) using machine learning, as well as critical scientific, practical, and ethical considerations related to their potential clinical application. Although self-report remains the gold standard for pain assessment, machine learning may aid in the classification of chronic pain disorders like chronic back pain and fibromyalgia as well as provide mechanistic information regarding their neural correlates.

This article is part of the Topical Collection on *Chronic Pain*

✉ Roland Staud
staudr@ufl.edu

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

[2] Department of Medicine, University of Florida, PO Box 100277, Gainesville, FL 32610, USA

Keywords Chronic pain · Musculoskeletal · Magnetic resonance imaging (MRI) · Machine learning · Classification

Introduction

Chronic pain is a highly prevalent condition associated with significant disability and societal cost [1]. The etiology of chronic pain can vary substantially across patients and often appears to be secondary to biological factors, such as musculoskeletal injury (e.g., osteoarthritis [OA]), nerve injury (neuropathic pain), autoimmunity (rheumatoid arthritis [RA], systemic lupus erythematosus), or substance abuse (alcoholic neuropathy and chronic pancreatitis). In many cases, however, chronic pain seems to be the primary illness associated with a given condition (e.g., fibromyalgia syndrome [FM; 2]). Over the past several decades, substantial effort has been dedicated to the development of methodologies for the reliable discrimination or classification of patients with a given chronic pain condition from people without the condition (i.e., "healthy controls"), patients with similar but distinct illnesses, or both.

The goal of many studies is to accurately identify individuals with a clinical chronic pain condition in question based on potential mechanistic underpinnings. Given that some chronic pain conditions are associated with peripheral tissue pathophysiology, numerous studies have been aimed at reliable discrimination of individuals with clinical pain using peripheral measures including structural (e.g., knee degeneration, lumbar disc pathology; [3]), functional (e.g., gait abnormality, inflammatory processes in rheumatoid arthritis; [4, 5]), and genetic data [6]. Because chronic pain often exists as a primary symptom without remarkable tissue abnormalities, there is an increasing interest in their neural correlates, i.e., mechanistic or structural abnormalities derived from structural or functional magnetic resonance (MRI) brain imaging studies

of chronic pain (e.g., grey matter density/volume, white matter integrity, or functional connectivity between pain-related brain regions). In addition, such information can be used to determine whether central nervous system (CNS) abnormalities are suitable to appropriately classify individuals with chronic pain.

In this regard, a major goal of many MRI-based chronic pain classification studies is the development of objective biomarkers for each condition that do not rely on self-report [7, 8•, 9]. Brain-based biomarker studies of chronic pain leverage the extensive neuroimaging literature describing the critical role of certain brain regions in the sensory (e.g., thalamus, primary and secondary somatosensory cortex, posterior insula, basal ganglia), affective (e.g., anterior insula, hypothalamus, anterior cingulate cortex, amygdala, hippocampus), and cognitive-evaluative (e.g., dorsolateral prefrontal cortex, anterior cingulate cortex, thalamus) aspects of the pain experience (for review, see [10]). Although our review focuses exclusively on brain imaging biomarkers, it is worth noting that a substantial literature has also applied this approach to peripheral mechanisms. For instance, measurement of exhaled organic volatiles [11] and joint ultrasound [12] have been suggested as useful approaches for the automated classification and/or differential diagnosis of rheumatoid arthritis.

Some proponents of brain imaging-based approaches to illness classification claim that brain biomarkers could act as a "surrogate" for pain self-report [13]. This could result in greater sensitivity to measure efficacy of analgesic treatments [9] by avoiding potential pitfalls associated with the use of subjective patient and provider reports. Furthermore, such brain biomarkers could serve as targets for novel treatments [14, 15]. In particular, objective markers of chronic pain may play an important role in pain classification of patients with cognitive/psychological dysfunction and in individuals who are unable to communicate, or in cases of deception [16•, 17], as well as aid in the adjudication of legal claims of injury-related pain disorders [18]. These purported advantages have spurred interest in the development of biomarkers, although it is important to note that self-report remains the gold standard for pain measurement. Numerous studies have indicated that self-report measures of pain and psychosocial factors have excellent classification accuracy and reliability [8•, 19•, 20]. Indeed, evidence suggests that chronic pain patients can be distinguished from healthy controls with greater than 90% accuracy based on personality factors [21], perceived pain and functional disability [22], and simple visual analog scale (VAS) measures of affect [23••]. Questionnaire-based approaches have shown similar accuracy for the separation of FM and RA [24] and FM, RA, and OA [25].

The purpose of this critical review is to examine the development of brain biomarkers for chronic musculoskeletal pain disorders using machine learning (ML) in a manner which should be useful to physicians and other health care providers. Therefore, we will restrict our review to musculoskeletal conditions for which chronic pain is a primary complaint. The conceptual, scientific, and ethical issues related to this approach, however, will also apply to biomarker studies for other chronic conditions (e.g., depression).

Machine Learning Algorithms

Existing methods regarding the use of structural and functional brain imaging to discriminate patient groups from healthy controls have largely relied on ML algorithms, which provide an automated approach to making predictions about previously unknown data. ML algorithms can be broadly classified as supervised or unsupervised. Supervised methods develop models for classifying observations according to a known outcome (e.g., a clinical diagnosis). In contrast, unsupervised methods attempt to discern patterns or structure in data without guidance or pre-existing labels. Although unsupervised ML methods may be of interest for the development of subgroups within existing conditions or new diagnostic classification criteria based on brain abnormalities, ML studies in chronic pain to date have mostly relied on supervised methods. Support vector machines (SVM; [26, 27]), a type of ML algorithm that uses a training data set composed of one or more features to determine an optimal boundary separating a set of cases, have been most often used [23••, 28, 29]. However, many other ML techniques have been developed and applied to the automated classification of chronic pain patients [23••, 30]. Detailed discussions of the strengths and limitations of these methods are beyond the scope of this review, but basic descriptions of the most common ML algorithms applied in the study of chronic pain, as well as detailed references, are provided in Table 1.

Scientific and Clinical Utility of Chronic Pain Biomarkers

In this section, we will discuss theoretical, practical, and ethical issues surrounding pain biomarker development and use, with a particular emphasis on clinical utility.

Choice of Classification Algorithm

In an ideal scenario, the best algorithm for detecting a chronic pain condition would be one that matches the underlying model or structure that distinguishes that condition from healthy individuals or other conditions (e.g., if a linear decrease of prefrontal cortical thickness is the underlying mechanism of a condition, then a linear model would be sufficient). However, given the complex and often heterogeneous nature of chronic pain conditions, this information is frequently not

Table 1 Basic descriptions of common supervised machine learning algorithms

Algorithm	Basic approach	Reference(s)
Logistic regression (LR)	Uses linear combination of predictors and regression coefficients to make categorical prediction	[31]
Naïve Bayes (NB)	Individuals categorized according to probability that they have a certain value for each feature independently included in prediction model	[32]
k-Nearest Neighbors (kNN)	Categorizes individuals based on the most common class of the *k* closest observations for each model feature	[33]
Multilayer perceptron (MLP)	Determines appropriate weighting for features for multiple layers of decision units (perceptrons) needed to perfectly classify training data, then applies learned rule to new cases	[34]
Support vector machine (SVM)	Determines a linear or non-linear maximum-margin hyperplane (i.e., feature-based cutoff) that maximizes separation between groups	[26, 27]
Decision tree (e.g., J48/C4.5)	Categorizes individuals in a series of decisions based upon optimal feature cutoff values	[35]
Least absolute shrinkage and selection operator (LASSO)	A variant of ordinary least squares regression that constrains the summary of absolute regression coefficients across features in order to identify the most informative features and optimal cut points for classification	[36]

available. It is often difficult, if not impossible, to determine a priori which algorithm might perform best in a certain data set [37]. This problem frequently necessitates the comparison of multiple different algorithms for any single classification task (for examples, see [23••, 38]). As a general rule, simple models may not perform as well on training data but will likely generalize better when assessed on testing data, whereas more complex models will tend to perform well on training data but will not perform as well on testing data. More complex models or those that blend inputs for many different algorithms (e.g., ensemble methods) may also lead to challenges in interpretability, requiring additional pipelines to distill results in ways that are meaningful and useful for clinicians. In general, large sample sizes are often needed to increase generalizability and prevent overfitting [39]. A recent study suggests that accuracy values obtained from smaller samples are opportunistically biased [40]. Therefore, proposed biomarkers must demonstrate strong performance in generalizable samples.

Biomarker Reliability and Feasibility of Implementation

A neuroimaging-derived biomarker cannot be reliably used for scientific or clinical purposes if the measures composing it do not demonstrate high test-retest reliability (i.e., the reproducibility of results over time). This concern is especially relevant given recent evidence that the test-retest reliability of functional connectivity metrics may vary widely depending on the brain regions examined [19•]. For this reason, the reliability of neuroimaging-based measures should not be taken for granted. This concern also applies to populations where typical brain function may be perturbed due to injury or illness (e.g., post-stroke [41]).

Even a biomarker that is meeting or exceeding criteria for appropriate use as a diagnostic tool will not be clinically useful if it cannot be readily assessed in clinic settings. As noted by Woo and Wager [42], all parameters of the biomarker and its associated testing procedure should be rigorously standardized so that all users collect data comparable to the standard used to derive the measure. Depending on the demands associated with collecting certain biomarkers (especially those derived from BOLD activity during task-based fMRI), such rigor may be difficult to achieve in clinical settings. Often the additional time, nuanced design, and/or special expertise required in the acquisition of reliable data will limit some biomarkers' applicability and usefulness.

What Constitutes Adequate Biomarker Performance?

A major issue in the application (and reporting) of ML-based brain biomarkers for chronic pain is determining whether a given marker's accuracy is adequate. Developing criteria for optimal biomarker performance depends on the setting in which it will be used, as well as the cost of misclassification. Pure research-oriented use cases (e.g., identifying potential markers, phenotypic subgroups, or treatment targets; discerning disease mechanisms) may allow for less stringent performance criteria. In contrast, clinical tasks (diagnosis, prognosis, treatment planning) carry risk of harm (misdiagnosis, delaying treatment, providing inaccurate or unnecessary treatment) and will therefore require considerably more stringent performance criteria for use in patient care. Although the risk of medical error from decisions based upon ML-derived biomarkers may be no greater than those commonly accepted in current practice, the "black-box" nature of these markers raises liability concerns in the face of medical errors promulgated by ML-derived biomarkers [43] and may increase physicians' reluctance to use even those that are well validated.

Ultimately, any given biomarker's practical applicability will depend on a case-by-case assessment of its cost,

deployability, and accuracy in the context of its intended use. For instance, in clinical settings, biomarker performance should be judged by taking into consideration whether it is intended to replace physicians in a role they can already perform well (e.g., differentially diagnose mechanistically and/or symptomatically distinct conditions) or, alternatively, provide novel information that clinicians would not otherwise have. This approach may help identify mechanistically distinct clinical subgroups, guide treatment decisions, indicate prognosis, or separate conditions with significant mechanistic or symptomatic overlap. In cases where biomarker application provides useful clinical information where it would not otherwise be available, higher cost, more difficult deployability, and/or relatively poorer performance may be quite tolerable [43].

Impact of Chronic Pain Prevalence on Biomarker Performance

Ethical and practical issues surrounding potential under- or overtreatment of pain remain even in cases where biomarker sensitivity (i.e., probability of a positive biomarker given an individual has the condition) and specificity (i.e., probability of a negative biomarker given an individual does not have the condition) reach very high levels, depending on the given use case. This is because biomarker positive predictive value (PPV) and negative predictive value (NPV), as previously noted [44•], depend on the prevalence (i.e., base rate) of the condition of interest in a particular setting (i.e., Bayes' theorem). Specifically, biomarker PPV increases with prevalence, while NPV increases as prevalence decreases. This relationship is illustrated graphically in Fig. 1 using the sensitivity (81%) and specificity (75%) of a structural brain biomarker for pain we have previously reported [23••]. As a result, even the highest performing chronic pain biomarkers previously reported (e.g., 92% sensitivity, 92% specificity [16•]) will have high rates of false positives in application settings with low prevalence and high rates of false negatives in settings with high prevalence. Investigators need to take use case into account when reporting biomarker performance and judging its adequacy.

Finally, self-report of pain symptomatology is ultimately the basis for diagnosing and assessing the severity of chronic pain conditions (so-called "Gold Standard"). Therefore, the predictive utility of biomarkers is necessarily limited by the accuracy of the existing diagnostic gold standard [8•, 45]. It is conceivable that subgroups within a diagnostic category could be identified solely based upon physiological measurements using unsupervised ML methods and later validated upon self-report, thereby circumventing this issue. However, to our knowledge, no previous neuroimaging-based biomarker studies for chronic pain have used this approach.

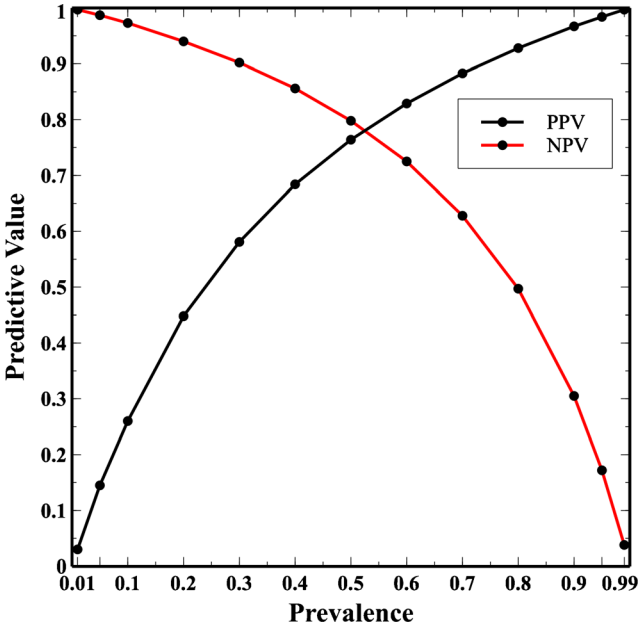

Fig. 1 Illustration of the relationship between chronic pain prevalence and the positive/negative predictive value of biomarkers using sensitivity (81%) and specificity (75%) values derived from Robinson et al. [23••]. Positive predictive value (PPV) increases with prevalence, while negative predictive value (NPV) decreases as prevalence increases. Thus, during real-world application of brain biomarkers for chronic pain, performance will depend largely on the expected proportion of individuals with the condition in question

Pain Biomarkers Based on Brain Imaging

Structural Biomarkers

Chronic pain brain biomarkers derived from structural MRI are dependent on the assumption that chronic pain conditions are associated with abnormalities in brain structure that either pre-date or are the result of pain chronification. Brain structure can be characterized in terms of grey matter features, which reflects integrity of neuronal cell bodies, or white matter features that reflect axonal integrity. Note that although white matter perturbations have been demonstrated in a variety of chronic pain conditions [46–48], to our knowledge, ML algorithms have yet to be applied to these data in order to generate chronic pain classifiers. Thus, we will focus on studies utilizing grey matter features.

Grey matter structure can be assessed in several ways. One common approach is the use of voxel-based morphometry (VBM) on high-resolution T1-weighted images. VBM produces estimates of grey matter density for each voxel and subject after warping to fit a standardized template brain [49, 50]. Another typical technique is to assess grey matter characteristics in cortical (thickness, volume, surface area, or mean curvature) or subcortical (volume) structures by assigning a neuroanatomical label to each brain voxel using a probabilistic

atlas [51, 52]. Though other techniques and measures are available for assessing grey matter structure, efforts to construct biomarkers for chronic pain conditions based on grey matter structure have largely relied on grey matter density [29, 53] or thickness/volume/surface area/curvature [23••, 30].

Structural brain biomarkers for chronic pain build on a significant literature demonstrating both atrophy and hypertrophy in chronic pain patients in numerous pain-related brain regions. For example, chronic low back pain (cLBP) has been associated with lower grey matter density compared to controls in the dorsolateral prefrontal cortex (DLPFC), a region associated with cognitive/evaluative function and pain modulation [54]. At the same time, FM has been associated with both increased grey matter volume in striatum, orbitofrontal cortex, and cerebellum [55], and lower volume in anterior cingulate cortex, amygdala, thalamus, superior temporal gyrus, supplementary motor cortex, and insula [55–58]. Although not directly contradicting one another, these studies differed somewhat with regard to regions showing significant differences between FM patients and controls, with differences in anterior cingulate cortex being the most consistently observed [56, 58]. The potential utility of these measures as features for classification algorithms to distinguish chronic pain patient groups depends on the assumption that reliable commonalities and differences in grey matter morphometrics can be detected between chronic pain conditions. It is also worth noting that the goal of chronic pain biomarker studies is distinct from between-group analyses of structural differences because they are focused specifically on identifying optimal combinations of features that best separate patient groups and/or healthy controls.

Classifier Studies Based on Structural Brain Abnormalities

ML classifier studies based on structural brain features have been conducted in several chronic pain conditions, including CPP, IBS, FM, and cLBP, using samples from 26 to 160 participants. In each case, samples were composed of ~50% splits of patients to controls. Performance of these potential biomarkers has differed substantially between studies, with sensitivity and specificity ranging from 65 to 81%. Although brain regions discriminating patient groups from normal controls varied, certain areas were more frequently reported. These included precentral gyrus (primary motor cortex), postcentral gyrus (primary somatosensory cortex), amygdala, and cuneus. In addition, many discriminating regions were convergent with those identified in previous studies focused on identifying structural differences between chronic pain patients and controls (e.g., DLPFC, amygdala, cingulate cortex, insula, etc.,). For further detail regarding sample size, algorithm(s) used, classifier performance, and brain structures with the greatest contribution to performance for studies using structural brain biomarkers, see Table 2.

Functional Brain Biomarkers

Whereas, high-resolution neuroanatomical data is the basis for studies proposing structural neuroimaging biomarkers of chronic pain, brain activity measures are used to derive functional neuroimaging biomarkers. Functional MRI (fMRI), which refers to noninvasive methods for measuring brain function, has two common variants. The first detects changes in cerebral blood oxygenation (BOLD), and the second, arterial spin labeling (ASL), measures changes in regional cerebral blood flow (rCBF). Although neither of these methods directly measure neuronal activity, they are both considered surrogate markers of such activity (for review of these technologies, see [61]). These methods excel in their ability to provide spatial information about where changes in BOLD response or rCBF occur over the course of time; however, they are intrinsically limited in temporal resolution, or accuracy in determining exactly when the neuronal activation occurred, due to the physiological characteristics of blood oxygenation (BOLD) and/or extended repetition times (ASL; [62]). Although ASL has several theoretical advantages over BOLD regarding the measurement of brain function [63], most brain classification studies for chronic pain to date have relied on BOLD fMRI.

Because functional neuroimaging provides a global measure of brain activity, participants' mental state during data collection is always important to note. For example, brain activity can be captured during a goal-directed task (e.g., participants are asked to continuously rate their levels of clinical pain or undergo acute painful stimulation), or during wakeful rest (i.e., resting state) in which participants are not instructed to engage in a particular task. Raw data collected during these scans typically undergoes preprocessing, or various image and signal correction techniques, to improve the ratio of desired signal to undesired noise [64]. Subsequently, preprocessed data are statistically analyzed at the individual-participant level to determine either 1) whether a significant change in signal occurred for a given region or across the whole brain (i.e., activation), or 2) the coherence in activation among spatially distinct brain regions over time (i.e., functional connectivity, FC). Statistically analyzed activation or FC values can be selected across the whole brain, or limited to hypothesis-driven regions of interest (i.e., a priori ROIs). In classification studies proposing functional neuroimaging biomarkers, individual-level activation or FC information is then

Table 2 Characteristics of machine learning-based classifier studies using brain imaging results

Study	Pain condition	Accuracy (%)	Sensitivity (%)	Specificity (%)	Manuscript N	Manuscript base rate (%)	Machine learning classifier	A priori regions of interest?	Major regions contributing to classification
Structural biomarkers									
Bagarinao et al. [28]	CPP	73	70	73	66	50	SVM	No	>Density in L SFG, L PCG, R PCG, L PaHcG, R PaHcG, amygdala, hippocampus in CPP
Labus et al. [30]	IBS	70	65	75	160	50	sPLS-DA	No	>Thickness in L PCG, lesser thickness in R insula; >mean curvature in R PreCG, lesser in L SCG, R temporal sulcus, and L intraparietal sulcus; and < surface area in L fusiform gyrus in IBS
Robinson et al., [23••]	FM	76	81	75	26	54	LR, MLP, Bayes, SVM, J48	No	< Volume in L amygdala in FM
Ung et al. [29]	cLBP	76	76	75	94	50	SVM	No	< GM density in R amygdala, L MOG, and R cuneus; greater GM density in R cerebellum, R/L middle/superior temporal gyrus, L PreCG, L PCG, R calcarine sulcus, and R DLPFC in cLBP
Dynamic biomarkers									
Callan et al. [16•]	cLBP	92	92	92	26	50	SLR	Yes	< L PCG activation and > L IPC in cLBP during pain
Sundermann et al. [59]	RA FM	79 (vs. HC) 62 (vs. HC) 79 (vs. RA)	NR	NR	50	34 32	SVM SVM kNN	Yes	NR
López-Solà et al. [60••]	FM	93	92	94	72	51	SVM, LR	Yes	> Activity in insula, PCC, precuneus, DMPFC, parahippocampal gyrus < Activity in DLPFC, primary/secondary visual and auditory cortex, lateral cerebellum, basal ganglia, and hypothalamus

SFG superior frontal gyrus, *PCG* postcentral gyrus, *PaHcG* parahippocampal gyrus, *PreCG* precentral gyrus, *SCG* subcollosal gyrus, *MOG* medial orbital gyrus, *DMPFC* dorsomedial prefrontal cortex, *DLPFC* dorsolateral prefrontal cortex, *IPC* inferior parietal cortex, *RA* rheumatoid arthritis, *FM* fibromyalgia, *CFS* chronic fatigue syndrome, *FM* fibromyalgia, *IBS* irritable bowel syndrome, *cLBP* chronic low back pain, *LASSO* least absolute shrinkage and selection operator, *OA* osteoarthritis, *CPP* chronic pelvic pain, *LR* logistic regression using ridge estimator, *MLP* multilayer perceptron, *J48* J48 decision tree, *PPV* positive predictive value, *NPV* negative predictive value, *SLR* sparse logistic regression, *sPLS-DA* sparse partial least squares for discriminant analysis, *IPC* inferior parietal cortex, *NR* not reported

entered into computational models, which classify individuals into distinct groups based on these values.

Classifier Studies Based on Functional Brain Abnormalities

Compared to structural brain abnormalities, fewer studies have used fMRI activation or functional connectivity metrics as features in classification models for chronic pain patients. Using measures of functional brain activation in pain-related brain regions during repeated 14-s blocks of experimental electrical pain induction, Callan et al. [16•] achieved 92% accuracy, sensitivity, and specificity classifying cLBP and healthy controls. The two most informative regions reported by Callan et al. [16•] have been strongly implicated in pain discrimination: (primary somatosensory cortex) and attention (inferior parietal cortex) [65]. In a separate study using resting state fMRI [59], RA and FM patients could be discriminated from HC with 79 and 62% accuracy, respectively, using measures of functional connectivity between structures associated with stimulus evaluation (i.e., the "salience network" [SN]) and internal focus/mind wandering (i.e., "default mode network" [DMN]). These networks have been implicated in pain processing (SN) or reported as perturbed in chronic pain states (DMN). Interestingly, RA and FM patients could also be discriminated from each other with 79% accuracy, suggesting that, despite potential commonalities in neuronal plasticity, the functional neural correlates of certain chronic pain disorders may be sufficiently distinct to enable their separation [59]. Finally, in a recent fMRI study using ML [60••], the investigators could discriminate FM patients from HC with 93% accuracy based on their brain activation associated with painful pressure stimuli [66] and non-painful multisensory stimulation (e.g., simultaneous auditory, tactile, and visual stimuli). Overall, the rapid improvement of fMRI-based "pain signatures" has not only helped our mechanistic understanding of chronic pain but will also benefit the diagnosis of FM and other chronic pain conditions.

Summary

Taken together, a limited number of studies (see Table 1) have tested the ability of ML algorithms to discern chronic pain patients from healthy controls using structural or functional brain abnormalities. However, interest in ML approaches for chronic pain diagnosis and classification is growing due to their purported potential to help elucidate chronic pain mechanisms, identify resilient or vulnerable subgroups, improve clinical decision making, predict treatment outcome, or augment [29] or even replace self-report (for critical discussion, see [8•]). However, as previously discussed, certain caveats apply to the practical application of biomarkers even where performance metrics may be very high.

Conclusions

Interest in the development of clinical and ML-based biomarkers for chronic musculoskeletal pain conditions derived from structural and functional neuroimages has increased substantially in recent years. Current reports describe novel biomarkers capable of separating patient groups and healthy controls with accuracies ranging from 70 to 93%. Such studies provide valuable mechanistic information regarding both the unique and common neural correlates of these conditions, with the potential to highlight differences between both musculoskeletal pain patient groups and controls to which traditional statistical approaches may not be sensitive, or identify mechanistic subgroups within certain pain conditions. However, at this time, critical theoretical, practical, and ethical concerns preclude the replacement of patient self-report for the diagnosis of chronic musculoskeletal pain with brain imaging-derived biomarkers, as self-report remains the gold standard for pain assessment [8•].

Acknowledgements This work was supported by NIH grant R01 NR014049-01

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Institute of Medicine. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. Washington, DC: National Academies Press; 2011.
2. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res. 2010;62(5):600–10. doi:10.1002/acr.20140.
3. Shamir L, Ling SM, Scott Jr WW, Bos A, Orlov N, Macura TJ, et al. Knee x-ray image analysis method for automated detection of osteoarthritis. IEEE Trans Biomed Eng. 2009;56(2):407–15. doi:10.1109/TBME.2008.2006025.
4. Kirkwood RN, Resende RA, Magalhaes CM, Gomes HA, Mingoti SA, Sampaio RF. Application of principal component analysis on gait kinematics in elderly women with knee osteoarthritis. Rev Bras Fis. 2011;15(1):52–8.

5. Frize M, Ogungbemile A. Estimating rheumatoid arthritis activity with infrared image analysis. Stud Health Technol Inform. 2012;180:594–8.
6. Frampton D, Kerr J, Harrison TJ, Kellam P. Assessment of a 44 gene classifier for the evaluation of chronic fatigue syndrome from peripheral blood mononuclear cell gene expression. PLoS One. 2011;6(3):e16872. doi:10.1371/journal.pone.0016872.
7. Apkarian AV, Hashmi JA, Baliki MN. Pain and the brain: specificity and plasticity of the brain in clinical chronic pain. Pain. 2011;152(3 Suppl):S49–64. doi:10.1016/j.pain.2010.11.010.
8.• Robinson ME, Staud R, Price DD. Pain measurement and brain activity: will neuroimages replace pain ratings? J Pain: Off J Am Pain Soc. 2013;14(4):323–7. doi:10.1016/j.jpain.2012.05.007. **This critical review summarized the limitations of neuroimaging to replace pain self-report. It also emphasized that neuroimaging will never replace pain self-report except in individuals with mental conditions or limited levels of consciousness**.
9. Wortolowska K. How neuroimaging can help us to visualise and quantify pain? Eur J Pain Suppl. 2011;5(S2):323–7.
10. Bushnell MC, Ceko M, Low LA. Cognitive and emotional control of pain and its disruption in chronic pain. Nat Rev Neurosci. 2013;14(7):502–11. doi:10.1038/nrn3516.
11. Brekelmans MP, Fens N, Brinkman P, Bos LD, Sterk PJ, Tak PP, et al. Smelling the diagnosis: the electronic nose as diagnostic tool in inflammatory arthritis. A case-reference study. PloS One. 2016;11(3):e0151715. doi:10.1371/journal.pone.0151715.
12. Rizzo G, Raffeiner B, Coran A, Ciprian L, Fiocco U, Botsios C, et al. Pixel-based approach to assess contrast-enhanced ultrasound kinetics parameters for differential diagnosis of rheumatoid arthritis. J Med Imaging (Bellingham). 2015;2(3):034503. doi:10.1117/1.JMI.2.3.034503.
13. Mackey SC. Central neuroimaging of pain. J Pain: Off J Am Pain Soc. 2013;14(4):328–31. doi:10.1016/j.jpain.2013.01.001.
14. Borsook D, Becerra L, Hargreaves R. Biomarkers for chronic pain and analgesia. Part 1: the need, reality, challenges, and solutions. Discov Med. 2011;11(58):197–207.
15. Borsook D, Becerra L, Hargreaves R. Biomarkers for chronic pain and analgesia. Part 2: how, where, and what to look for using functional imaging. Discov Med. 2011;11(58):209–19.
16.• 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. doi:10.1371/journal.pone.0098007. **This study employed supervised machine learning techniques, specifically sparse logistic regression, to train a classifier of chronic back pain based on contrast images using a leave-one-out cross-validation procedure. It correctly classified 92.3% of the chronic pain group and 92.3% of the normal controls**.
17. Lu HC, Hsieh JC, Lu CL, Niddam DM, Wu YT, Yeh TC, et al. Neuronal correlates in the modulation of placebo analgesia in experimentally-induced esophageal pain: a 3T-fMRI study. Pain. 2010;148(1):75–83. doi:10.1016/j.pain.2009.10.012.
18. Pustilnik AC. Imaging brains, changing minds: how pain neuroimaging can inform the law. Alabama Law Rev. 2015;66(5).
19.• Letzen JE, Boissoneault J, Sevel LS, Robinson ME. Test-retest reliability of pain-related functional brain connectivity compared to pain self-report. Pain. 2016;157(3):546–51. doi:10.1097/j.pain.0000000000000356. **This study examined the test-retest reliability for functional connectivity MRI (fcMRI) of pain-related brain regions and self-reported pain. Intraclass correlations coefficients for fcMRI values varied widely (range = −.174–.766), whereas intraclass correlations coefficients for VAS scores ranged from .906 to .947. Overall, self-reported pain was more reliable than fcMRI data**.
20. Nielsen CS, Price DD, Vassend O, Stubhaug A, Harris JR. Characterizing individual differences in heat-pain sensitivity. Pain. 2005;119(1–3):65–74. doi:10.1016/j.pain.2005.09.018.
21. Slesinger D, Archer RP, Duane W. MMPI-2 characteristics in a chronic pain population. Assessment. 2002;9(4):406–14.
22. Takekawa KS, Goncalves JS, Moriguchi CS, Coury HJ, Sato TO. Can a self-administered questionnaire identify workers with chronic or recurring low back pain? Ind Health. 2015;53(4):340–5. doi:10.2486/indhealth.2014-0241.
23.•• Robinson ME, O'Shea AM, Craggs JG, Price DD, Letzen JE, Staud R. Comparison of machine classification algorithms for fibromyalgia: neuroimages versus self-report. J Pain: Off J Am Pain Soc. 2015;16(5):472–7. doi:10.1016/j.jpain.2015.02.002. **Separate models representing brain volumes, mood ratings, and pain intensity ratings were estimated across several ML algorithms. Structural magnetic resonance imaging data from fibromyalgia patients and healthy controls and self-report data of pain intensity and mood were used. 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**.
24. Friend R, Bennett RM. Distinguishing fibromyalgia from rheumatoid arthritis and systemic lupus in clinical questionnaires: an analysis of the revised Fibromyalgia Impact Questionnaire (FIQR) and its variant, the Symptom Impact Questionnaire (SIQR), along with pain locations. Arthritis Res Ther. 2011;13(2):R58. doi:10.1186/ar3311.
25. Perrot S, Bouhassira D, Fermanian J, Cercle d'Etude de la Douleur en R. Development and validation of the Fibromyalgia Rapid Screening Tool (FiRST). Pain. 2010;150(2):250–6. doi:10.1016/j.pain.2010.03.034.
26. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.
27. Pontil M, Verri A. Properties of support vector machines. Neural Comput. 1998;10(4):955–74.
28. Bagarinao E, Johnson KA, Martucci KT, Ichesco E, Farmer MA, Labus J, et al. Preliminary structural MRI based brain classification of chronic pelvic pain: a MAPP network study. Pain. 2014;155(12):2502–9. doi:10.1016/j.pain.2014.09.002.
29. 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):1037–44. doi:10.1093/cercor/bhs378.
30. Labus JS, Van Horn JD, Gupta A, Alaverdyan M, Torgerson C, Ashe-McNalley C, et al. Multivariate morphological brain signatures predict patients with chronic abdominal pain from healthy control subjects. Pain. 2015;156(8):1545–54. doi:10.1097/j.pain.0000000000000196.
31. Le Cessie S, Van Houwelingen JC. Ridge estimators in logistic regression. Appl Stat. 1992;41(1):191–201.
32. Langley P, Iba W, Thompson K. An analysis of Bayesian classifiers. AAAI. 1992;90:223–8.
33. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. Mach Learn. 1991;6(1):37–66. doi:10.1023/A:1022689900470.
34. Arora R. Comparative analysis of classification algorithms on different datasets using WEKA. Int J Comput Appl. 2012;54(13):21–5.
35. Quinlan JR. C4.5 programs for machine learning. San Mateo: Morgan Kaufmann; 1992.
36. Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc B Met. 1996;58(1):267–88.
37. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med. 2001;23(1):89–109.
38. Nair SS, French RM, Laroche D, Thomas E. The application of machine learning algorithms to the analysis of electromyographic

patterns from arthritic patients. IEEE Trans Neural Syst Rehabil Eng. 2010;18(2):174–84. doi:10.1109/TNSRE.2009.2032638.
39. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of deep learning in biomedicine. Mol Pharm. 2016;13(5):1445–54. doi:10.1021/acs.molpharmaceut.5b00982.
40. Schnack HG, Kahn RS. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. Front Psych. 2016;7:50. doi:10.3389/fpsyt.2016.00050.
41. Siegel JS, Ramsey LE, Snyder AZ, Metcalf NV, Chacko RV, Weinberger K, et al. Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. Proc Natl Acad Sci U S A. 2016. doi:10.1073/pnas.1521083113.
42. Woo CW, Wager TD. Neuroimaging-based biomarker discovery and validation. Pain. 2015;156(8):1379–81. doi:10.1097/j.pain.0000000000000223.
43. Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920–30. doi:10.1161/CIRCULATIONAHA.115.001593.
44.• Robinson M, Boissoneault J, Sevel L, Letzen J, Staud R. The effect of base rate on the predictive value of brain biomarkers. J Pain. 2016;17(6):637–41. doi:10.1016/j.jpain.2016.01.476. **Results of this study strongly suggest that many proposed brain biomarkers perform quite poorly when realistic illness prevalence base rates are taken into account**.
45. Lueken U, Zierhut KC, Hahn T, Straube B, Kircher T, Reif A, et al. Neurobiological markers predicting treatment response in anxiety disorders: a systematic review and implications for clinical application. Neurosci Biobehav Rev. 2016;66:143–62. doi:10.1016/j.neubiorev.2016.04.005.
46. Buckalew N, Haut MW, Aizenstein H, Morrow L, Perera S, Kuwabara H, et al. Differences in brain structure and function in older adults with self-reported disabling and nondisabling chronic low back pain. Pain Med. 2010;11(8):1183–97. doi:10.1111/j.1526-4637.2010.00899.x.
47. Geha PY, Baliki MN, Harden RN, Bauer WR, Parrish TB, Apkarian AV. The brain in chronic CRPS pain: abnormal gray-white matter interactions in emotional and autonomic regions. Neuron. 2008;60(4):570–81. doi:10.1016/j.neuron.2008.08.022.
48. Gustin SM, Peck CC, Cheney LB, Macey PM, Murray GM, Henderson LA. Pain and plasticity: is chronic pain always associated with somatosensory cortex activity and reorganization? J Neurosci: Off J Soc Neurosci. 2012;32(43):14874–84. doi:10.1523/JNEUROSCI.1733-12.2012.
49. Greve DN. An absolute beginner's guide to surface- and voxel-based morphometric analysis. ISMRM 19th Annual Meeting & Exhibition. 2011;7–13.
50. Ashburner J, Friston K. Morphometry. In: Ashburner J, editor. Human brain function. 2nd ed. Cambridge: Academic; 2004. p. 707–22.
51. Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans Med Imaging. 2001;20(1):70–80. doi:10.1109/42.906426.
52. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002;33(3):341–55.
53. Baliki MN, Schnitzer TJ, Bauer WR, Apkarian AV. Brain morphological signatures for chronic pain. PLoS One. 2011;6(10):e26010. doi:10.1371/journal.pone.0026010.
54. Apkarian AV, Sosa Y, Sonty S, Levy RM, Harden RN, Parrish TB, et al. Chronic back pain is associated with decreased prefrontal and thalamic gray matter density. J Neurosci: Off J Soc Neurosci. 2004;24(46):10410–5. doi:10.1523/JNEUROSCI.2541-04.2004.
55. Schmidt-Wilcke T, Luerding R, Weigand T, Jurgens T, Schuierer G, Leinisch E, et al. Striatal grey matter increase in patients suffering from fibromyalgia—a voxel-based morphometry study. Pain. 2007;132 Suppl 1:S109–16. doi:10.1016/j.pain.2007.05.010.
56. Burgmer M, Gaubitz M, Konrad C, Wrenger M, Hilgart S, Heuft G, et al. Decreased gray matter volumes in the cingulo-frontal cortex and the amygdala in patients with fibromyalgia. Psychosom Med. 2009;71(5):566–73. doi:10.1097/PSY.0b013e3181a32da0.
57. Puri BK, Agour M, Gunatilake KD, Fernando KA, Gurusinghe AI, Treasaden IH. Reduction in left supplementary motor area grey matter in adult female fibromyalgia sufferers with marked fatigue and without affective disorder: a pilot controlled 3-T magnetic resonance imaging voxel-based morphometry study. J Int Med Res. 2010;38(4):1468–72.
58. Robinson ME, Craggs JG, Price DD, Perlstein WM, Staud R. Gray matter volumes of pain-related brain areas are decreased in fibromyalgia syndrome. J Pain: Off J Am Pain Soc. 2011;12(4):436–43. doi:10.1016/j.jpain.2010.10.003.
59. Sundermann B, Burgmer M, Pogatzki-Zahn E, Gaubitz M, Stuber C, Wessolleck E, et al. Diagnostic classification based on functional connectivity in chronic pain: model optimization in fibromyalgia and rheumatoid arthritis. Acad Radiol. 2014;21(3):369–77. doi:10.1016/j.acra.2013.12.003.
60.•• López-Solà M, Woo CW, Pujol J, Deus J, Harrison BJ, Monfort J, Wager TD. Towards a neurophysiological signature for fibromyalgia. Pain. 2016. **A 'Multisensory' classifier trained on non-painful sensory stimulation revealed a brain signature for FM with high sensitivity and specificity. This study characterized individual FM patients based on symptom-related brain features**.
61. Detre JA, Wang J. Technical aspects and utility of fMRI using BOLD and ASL. Clin Neurophysiol. 2002;113(5):621–34.
62. Howseman AM, Bowtell RW. Functional magnetic resonance imaging: imaging techniques and contrast mechanisms. Philos Trans R Soc Lond B Biol Sci. 1999;354(1387):1179–94. doi:10.1098/rstb.1999.0473.
63. Chen JJ, Jann K, Wang DJ. Characterizing resting-state brain function using arterial spin labeling. Brain Connect. 2015;5(9):527–42. doi:10.1089/brain.2015.0344.
64. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004;23 Suppl 1:S208–19. doi:10.1016/j.neuroimage.2004.07.051.
65. Giesecke T, Gracely RH, Grant MA, Nachemson A, Petzke F, Williams DA, et al. Evidence of augmented central pain processing in idiopathic chronic low back pain. Arthritis Rheum. 2004;50(2):613–23. doi:10.1002/art.20063.
66. 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):1388–97.