Letter

Issues in Pain Prediction – More Gain than Pain

Matthias Zunhammer,[1,*] Ulrike Bingel,[1] and Tor D. Wager[2,3]

Recently, Hu and Iannetti [1] discussed pitfalls of using machine learning (ML) for predicting pain from neuroimaging data. We agree that it is important to establish dos and don'ts for using ML in pain neuroimaging to avoid spurious conclusions. However, we believe that several opinions they offer are based on flawed premises; here, we provide a brief counterpoint.

First, Hu and Iannetti caution that 'pain signatures' obtained from ML procedures may suffer 'the same problem of reverse inference' as earlier brain mapping approaches, and state that any valid brain response pattern for pain must be 'unique for pain'. We agree that reverse inference is challenging, and along with others have been vocal critics of overstated reverse inference claims for many years [2]. However, reverse inferences are not inherently flawed, and can be useful if validated successfully [3]. Many medical decision-making tests are based on them, and rely on Bayes' rule: for example, the probability of disease D given observed test results T, or $P(D|T)$, is $P(T|D) \times P(D)/P(T)$. The same applies when T is a marker derived from neuroimaging and the outcome is pain incidence or intensity. Ruling out potential alternative explanations is an open-ended challenge, requiring many tests across studies (e.g., [4,5]), but ultimately an empirical one.

Second, Hu and Iannetti postulate that signal magnitude cannot contribute meaningfully to neural signatures of pain, and that brain signals must be normalized to remove their mean intensity. We agree that it is interesting to understand whether predictive signals are driven mainly by topography or by response magnitude.

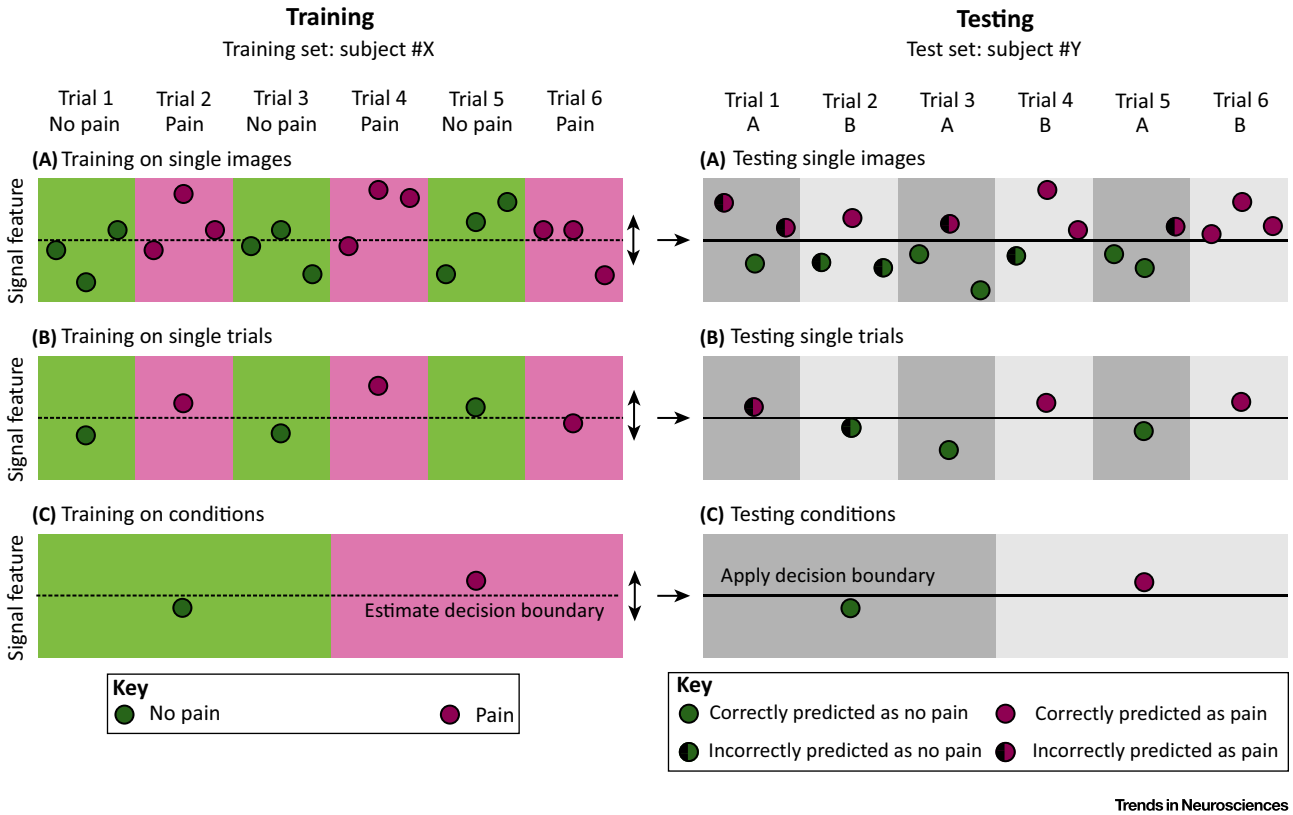

Figure 1. Valid Tests of a Classifier with Three Different Levels of Data Aggregation. (A) Prediction accuracies for single images within trials are low due to functional magnetic resonance imaging noise. (B) Multiple images taken during a trial can be used to estimate a trial-level response and predict outcomes (e.g., pain) on a trial-by-trial basis. This increases accuracy in accordance with the central limit theorem (adjusting for autocorrelation). This requires prior knowledge of when trials occur, but not the outcome value. (C) Averaging by condition is based on exactly the same principle: knowledge about how trials are grouped into similar conditions is used to average images and to increase the signal-to-noise ratio. Prior information about outcome (e.g., pain) is not required (or 'injected') to make a prediction; nonpainful and painful conditions could be labeled 'A' and 'B', as in the illustration above. One needs to only assume that the conditions represent meaningful groupings of trials – for example, repeated trials on a hypersensitive versus normal skin site. Predictions of pain are valid on a by-image, by-trial, by-condition, and by-subject level, as long as predictions are made on data independent of those used to train the model.

However, we believe that there is no reason to restrict the term 'pain signature' to measures of topography alone. At a cerebral level, pain is likely encoded by the neuronal firing 'intensity' [6] in neural populations targeted by nociceptive afferents and pain-related wide dynamic range neurons [7]. Ultimately, the topography, response magnitude, and other features such as the time course of responses and interactions between brain regions are likely to be important. Conversely, we should not be too quick to uncritically accept the null hypothesis that painful and nonpainful conditions produce equivalent activation; for example, a recent study claimed that 'pain matrix' activation is not specific for pain [8], without fully considering potential differences in activation magnitude and functional reorganization in unique patients.

Third, Hu and Iannetti take issue with the practice of averaging across multiple trials in a condition when testing ML-derived markers or 'signatures'. They claim that this practice, which we refer to as 'by-condition' testing (Figure 1), violates a 'fundamental rule' of ML by capitalizing on prior knowledge, leading to 'artificially high' prediction accuracy. This is simply not true. Testing models on a by-condition basis is common practice across many types of neuroimaging data [9], including seminal works on functional magnetic resonance imaging (fMRI) pattern recognition (e.g., [10]). Wager *et al.* [5] tested a pain-predictive fMRI 'signature' by averaging trials within conditions, and making a prediction about which is more painful. Crucially, in these and related studies, 'prior knowledge' about the painfulness of each condition is only used when assessing prediction accuracy, not when making the predictions from test data. One needs to only use prior knowledge of when events in each condition occurred and which trials belong to which conditions, not their painfulness. For example, one might test conditions involving stimulation at 47°C and 49°C, but the intensity information is not used in making the prediction – they could equally be labeled 'A' and 'B' (Figure 1).

Averaging over trials is expected to improve prediction accuracy as long as the condition labels are meaningful, by virtue of the central limit theorem (Figure 1). Recently, Lindquist *et al.* [4] applied a 'pain signature' to data from six independent data sets (N = 180), which revealed solid predictive accuracy at the single-trial level. Based on the central limit theorem, they also estimated the number of trials one would need to average to obtain a criterion level of prediction accuracy (e.g., 90%), demonstrating how averaging over trials increases predictive power.

In summary, Hu and Iannetti initiated an important discussion about the use of ML in pain research. Understanding the nuances of which procedures are valid, and which inferences can and cannot be made, is an ongoing process. Ultimately, in our view ML is a step forward from the statistical parametric mapping and region-of-interest approaches that have dominated the field until recently. Used appropriately, it has the potential to transform the neuroscience of pain, and brain–mind mapping more broadly. We must move forward constructively and with clarity.

[1]Klinik für Neurologie, Universitätsklinikum Essen, Hufelandstraße 55, 45147 Essen, Germany
[2]Department of Psychology and Neuroscience, University of Colorado, Boulder, 345 UCB, Boulder, CO USA
[3]The Institute of Cognitive Science, University of Colorado, Boulder, 345 UCB, Boulder, CO USA

*Correspondence: matthias.zunhammer@uk-essen.de (M. Zunhammer).

http://dx.doi.org/10.1016/j.tins.2016.07.004

References

1. Hu, L. and Iannetti, G.D. (2016) Painful issues in pain prediction. *Trends Neurosci.* 39, 212–220
2. Yarkoni, T. *et al.* (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670
3. Hutzler, F. (2014) Reverse inference is not a fallacy *per se*: cognitive processes can be inferred from functional imaging data. *Neuroimage* 84, 1061–1069
4. Lindquist, M.A. *et al.* (2016) Group-regularized individual prediction: theory and application to pain. *Neuroimage.* http://dx.doi.org/10.1016/j.neuroimage.2015.10.074
5. Wager, T.D. *et al.* (2013) An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* 368, 1388–1397
6. Timmermann, L. *et al.* (2001) Differential coding of pain intensity in the human primary and secondary somatosensory cortex. *J. Neurophysiol.* 86, 1499–1503
7. Treede, R.D. *et al.* (1999) The cortical representation of pain. *Pain* 79, 105–111
8. Salomons, T.V. *et al.* (2016) The "pain matrix" in pain-free individuals. *JAMA Neurol.* 73, 755–756
9. Pereira, F. *et al.* (2009) Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45 (1 Suppl), S199–S209
10. Poldrack, R.A. *et al.* (2009) Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychol. Sci.* 20, 1364–1372