

[Print this Page](#)

## Presentation Abstract

Program#/Poster#: 679.04/NNN3

Presentation Title: Using sensitivity analysis to interpret machine learning applied to fMRI data

Location: Halls B-H

Presentation time: Tuesday, Nov 12, 2013, 4:00 PM - 5:00 PM

Topic: ++G.07.a. Data analysis and statistics

Authors: **\*A. FLOREN, B. NAYLOR, D. RESS;**  
Univ. of Texas at Austin, Austin, TX

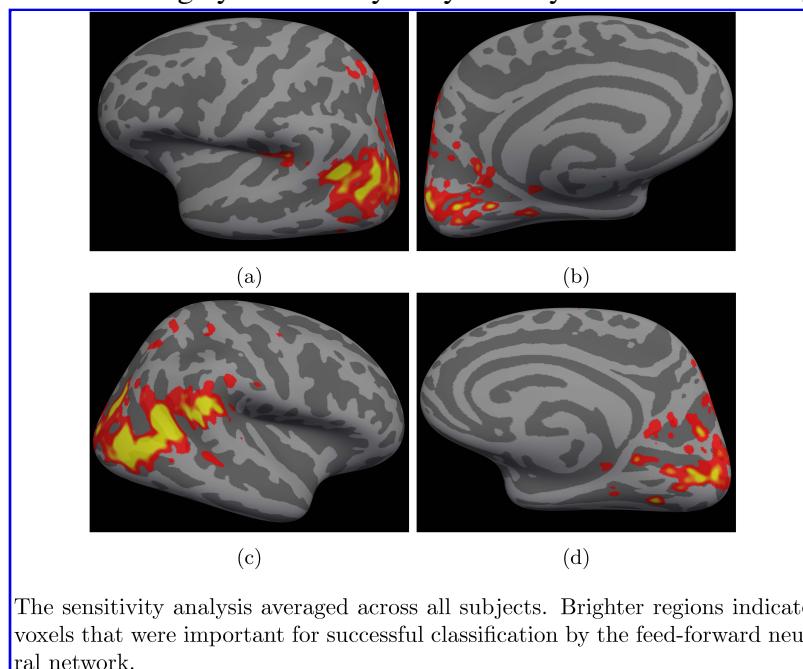
Abstract: Recently, machine learning has been employed as an analysis tool in fMRI. Algorithms are trained to classify or predict various stimuli from voxel time series. These flexible algorithms can potentially interpret complex and distributed patterns of activation in the brain. However, relating their results to brain functional anatomy is difficult.

Classifier performance is bounded by the mutual information between the voxels and stimulus. If some set of voxels can be used to predict the stimulus then some, but not necessarily all, of those voxels must be driven by that stimulus. However, poor classification performance of a single voxel is insufficient to disregard that voxel because of potential conditional dependencies; two voxels may be able to better predict the stimulus when taken jointly rather than independently. This is consistent with the concept of population encoding.

Starting with an input set of voxels, we want to isolate the minimum distributed network that best encodes the stimulus without evaluating all permutations. Here, we measure the sensitivity of the classifier outputs to its inputs. In feed-forward neural networks, this can be solved analytically. We train the network, calculate the sensitivity of the network to each voxel, and remove the voxels with low sensitivity. We repeat this process until the

network performance begins to degrade.

We tested our approach on a complex visual stimulus. Whole-brain fMRI data was collected while subjects viewed a virtual world that contained a variable number (1–6) of animated characters. We trained a feed-forward neural network to predict the number of characters present in each frame; classification performance was >50%, significantly above chance (~17%). Sensitivity analysis and voxel culling were performed as described above. Sensitivity values were then smoothed and averaged across subjects and projected onto the cortical surface (Figure). The network of classification voxels is almost entirely clustered in visual cortex, indicating that character number is largely encoded by early sensory and association regions.



The sensitivity analysis averaged across all subjects. Brighter regions indicate voxels that were important for successful classification by the feed-forward neural network.

Disclosures:

**A. Floren:** None. **B. Naylor:** None. **D. Ress:** None.

Keyword(s):

FUNCTIONAL MRI

MACHINE LEARNING

VISION

Support:

ARO W911NF-12-1-0160