Generative embedding for model-based classification of fMRI data

Kay H Brodersen[1,2] · Thomas M Schofield[3] · Alexander P Leff[3] · Cheng Soon Ong[1] · Ekaterina I Lomakina[1,2] · Joachim M Buhmann[1] · Klaas E Stephan[2,3]

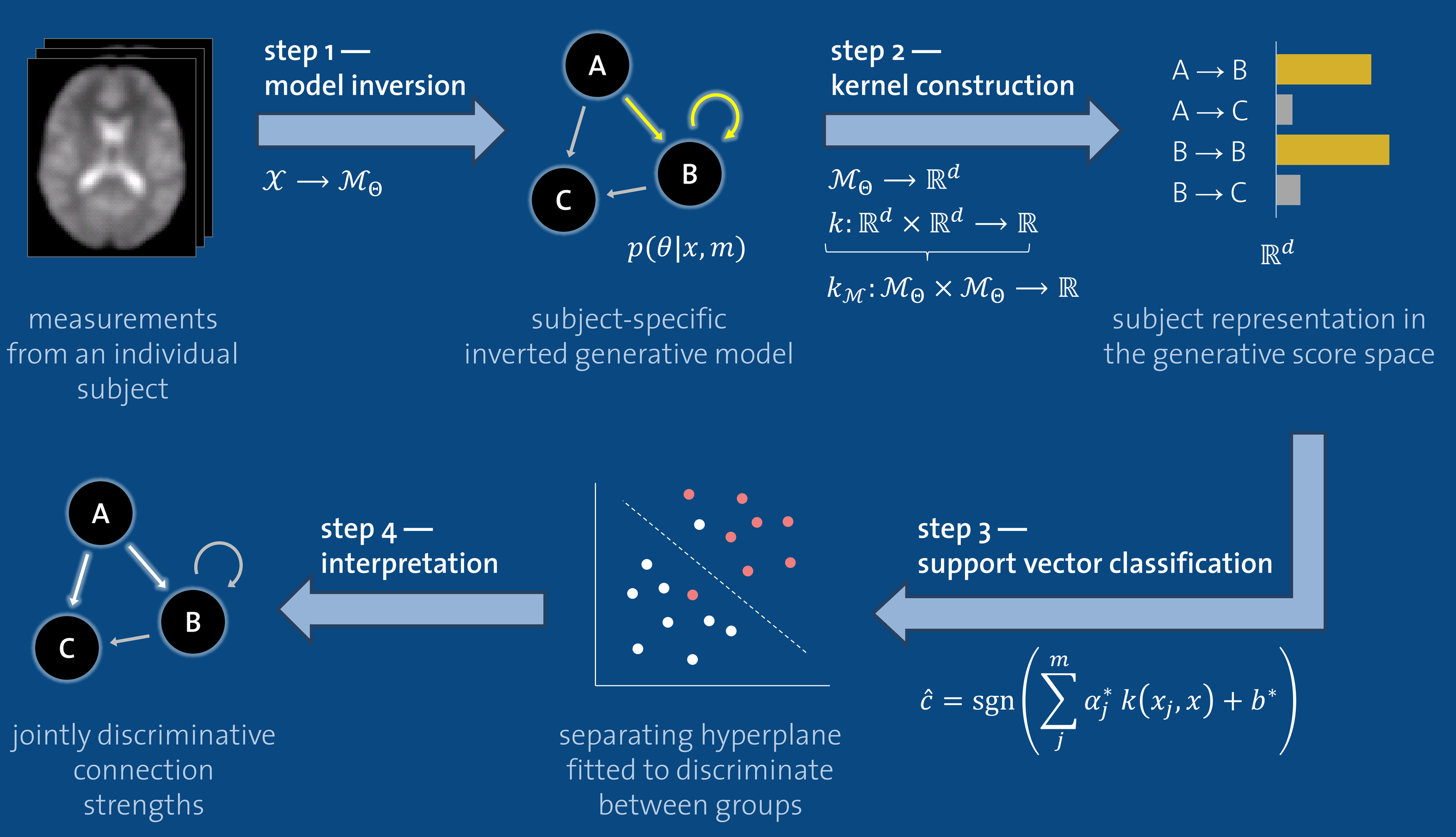
[1] Department of Computer Science, ETH Zurich, Switzerland [2] Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Switzerland [3] Wellcome Trust Centre for Neuroimaging, University College London, United Kingdom

1 Summary

- Multivariate classification algorithms rely on decoding models to infer cognitive or clinical brain states from fMRI data [1]. Two major challenges for all approaches are: (i) the high data dimensionality in fMRI, and (ii) achieving mechanistic interpretability.
- We address these issues by proposing a novel generative-embedding approach that incorporates neurobiologically interpretable generative models into discriminative classifiers [2].
- Using fMRI data from aphasic patients and healthy controls, we illustrate that our approach enables more accurate classification and deeper mechanistic insights than conventional methods.
- Generative embedding may be particularly useful whenever:
  - ☑ the scientific question reduces to a classification or regression problem,
  - ☑ a generative model of the data is available,
  - ☑ model parameters can be interpreted mechanistically.

2 Generative embedding for fMRI

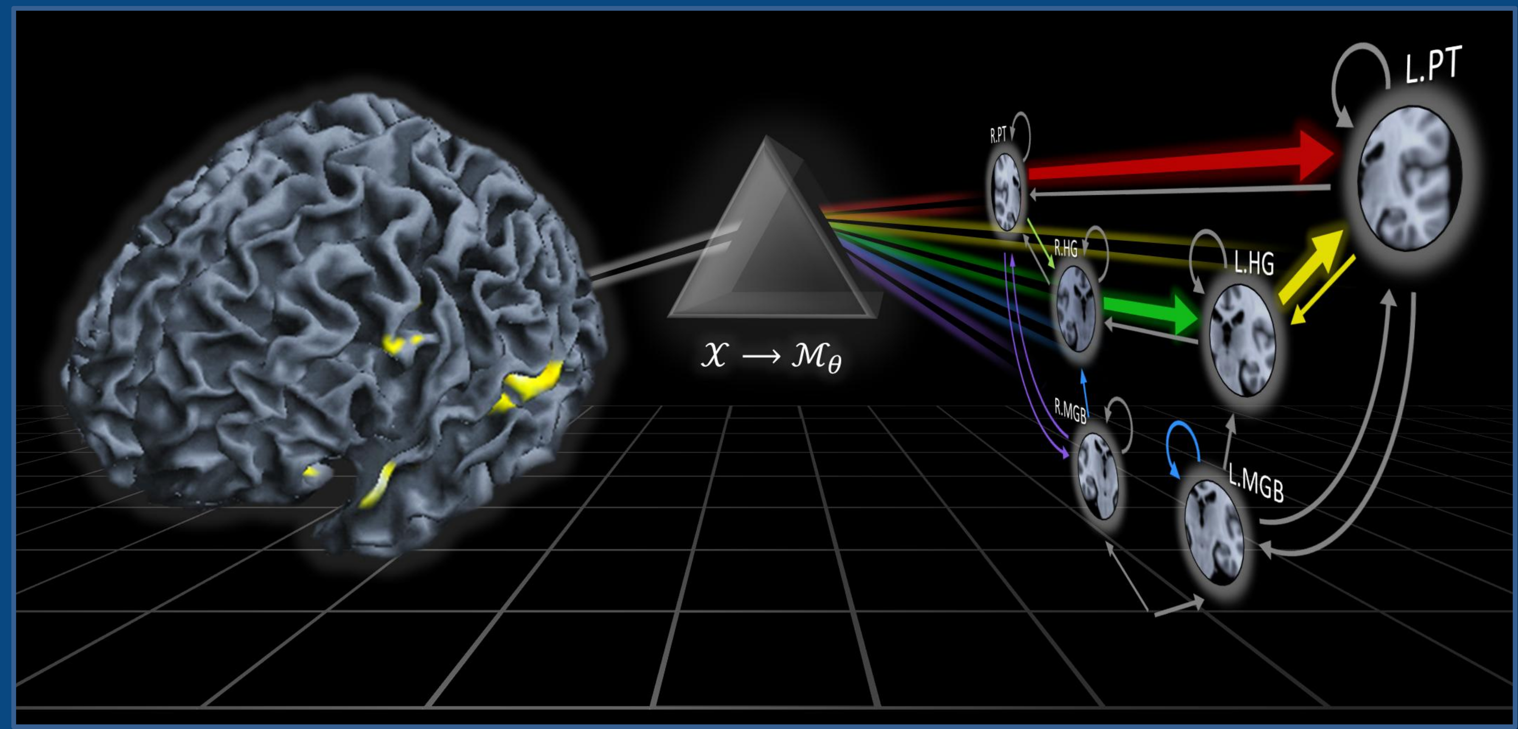
We introduce generative embedding for fMRI using a combination of dynamic causal models (DCM) and support vector machines (SVM).



Our procedure extends the literature on generative kernels [3] and on trial-by-trial classification for electrophysiological recordings [4] to subject-by-subject classification of fMRI.

3 Clinical example: speech impairments

We illustrate the utility of our approach by a clinical example in which we classify moderately aphasic patients and healthy controls [5] using a DCM of thalamo-temporal regions during speech processing [6].

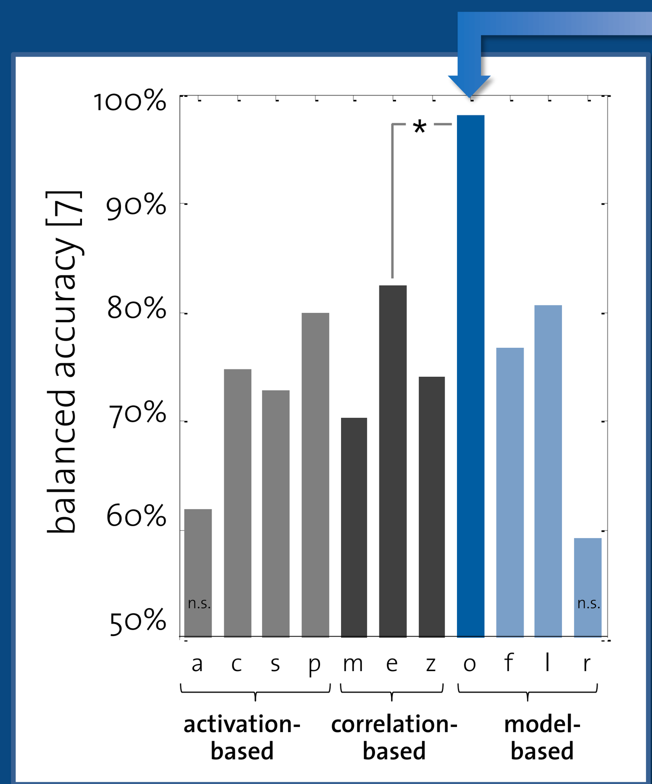


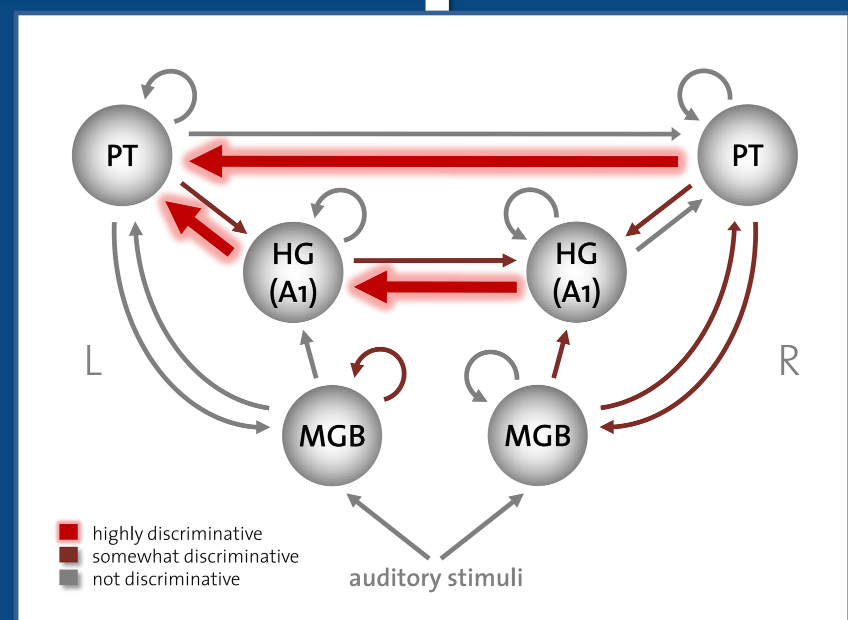
Left | In order to construct a dynamic causal model (DCM) of speech processing, we examined neural activity in response to speech. The figure shows a simple 'speech' versus 'no speech' contrast.

Right | In generative embedding, a high-dimensional activity pattern is transformed into a low-dimensional connectivity pattern. The figure shows a dynamic causal model (DCM) of speech processing whose subject-specific connection strengths served as features for classification.

4 Prediction performance

Generative embedding achieves a near-perfect balanced classification accuracy of 98%. Our approach significantly outperforms conventional activation-based and correlation-based methods.



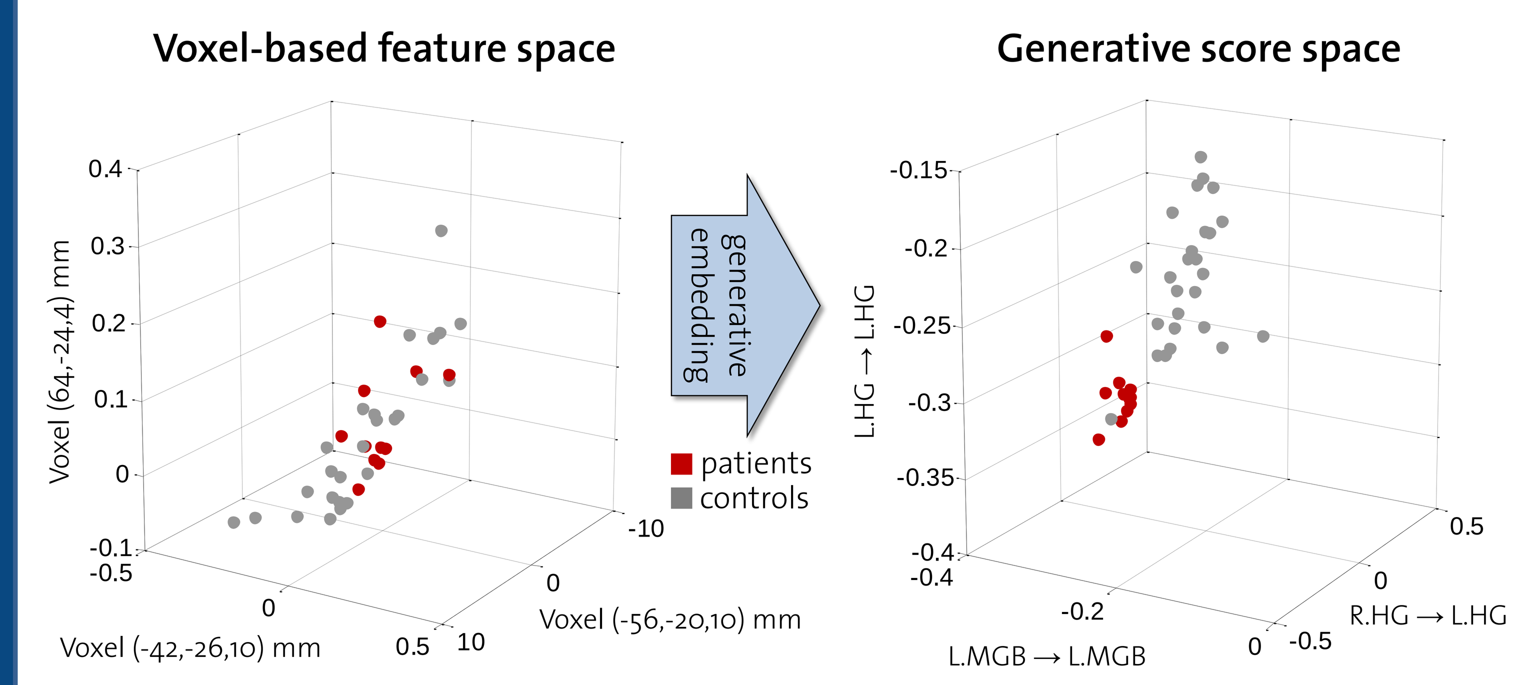


The model that enables an accuracy of 98% views speech processing as an interplay between six key regions.

Generative embedding was compared to several conventional approaches. ■ **Conventional activation-based methods:** (a) anatomical feature selection, (c) contrast feature selection, (s) searchlight feature selection, (p) PCA-based dimensionality reduction ■ **Conventional correlation-based methods:** (m) region-means correlations, (e) eigenvariates correlations, (z) eigenvariates z-correlations ■ **Generative embedding:** (o) original full model, (f) implausible feedforward model, (l) left hemisphere only, (r) right hemisphere only.

5 Induction of a generative score space

- In generative embedding, a DCM transforms the data from a high-dimensional voxel-based feature space into a low-dimensional generative score space.
- The generative score space may enable much better separability of patients and healthy controls, as shown below.



Left | The three axes represent the peaks of those three clusters that showed the strongest discriminability between patients and controls, based on a searchlight classification analysis.

Right | The three axes represent the three individually most discriminative parameters (two-tailed t-test) in the generative score space induced by a DCM of speech processing

6 Conclusions

- The first advantage of generative embedding over conventional methods is that it may provide more accurate predictions by exploiting discriminative information encoded in 'hidden' physiological quantities such as synaptic connection strengths.
- The second advantage is that it affords mechanistic interpretability of clinical classifications.
- We envisage that future applications of generative embedding may provide crucial advances in dissecting spectrum disorders into physiologically more well-defined subgroups.

Acknolwedgements
This study was funded by the University Research Priority Program 'Foundations of Human Social Behaviour' at the University of Zurich (KHB, KES), the SystemsX.ch project NEUROCHOICE (KHB, KES), the NCCR 'Neural Plasticity' (KES), the Wellcome Trust (APL), and the NIHR CBRC at University College Hospitals London (APL).

References

[1] Haynes, J. & Rees, G., 2006. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7), 523-534.
[2] Brodersen, K.H. et al., 2011. Generative embedding for model-based classification of fMRI data. *PLoS Comp Biol* (in press).
[3] Jaakkola, T.S. & Haussler, D., 1999. Exploiting generative models in discriminative classifiers. *NIPS,* 487-493.
[4] Brodersen, K.H. et al., 2011. Model-based feature construction for multivariate decoding. *NeuroImage*, 56, 601-615.
[5] Schofield, T.M. et al. (*in preparation*)
[6] Friston, K.J., Harrison, L. & Penny, W., 2003. Dynamic causal modelling. *NeuroImage*, 19(4), 1273-1302.
[7] Brodersen, K.H. et al., 2010. The balanced accuracy and its posterior distribution. *ICPR*, 3121-3124.