Predicting nicotine dependence from multimodal MRI data

Pablo Lasarte[1], Vladimir Omelyusik[1], Spencer Upton[2], Nicholas Henigman[2], Brett Froeliger[2], Satish S. Nair*[1]

1. Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, 65211. 2. Department of Psychological Sciences, University of Missouri, Columbia, MO, 65211.

Mizzou

LBP023.07

Motivation

Nicotine dependence (ND) is a major public health issue linked to significant changes in brain circuitry. Traditional analyses often use generalized linear models (GLMs) to identify neural correlates behind the disease. We investigate whether non-linear machine learning (ML) models can predict ND with high accuracy and interpretability using multimodal magnetic resonance imaging (MRI) data.

Methods

Subjects and recordings

Anatomical and resting-state functional MRI (fMRI) recorded from 276 subjects across multiple studies. Data was processed with DeepPrep *(Ren et al., 2025, Nature Methods)*.

Regions of interest (ROIs)

68 cortical ROIs according to the Desikan-Killiany atlas *(Desikan et al., 2006, NeuroImage)*.

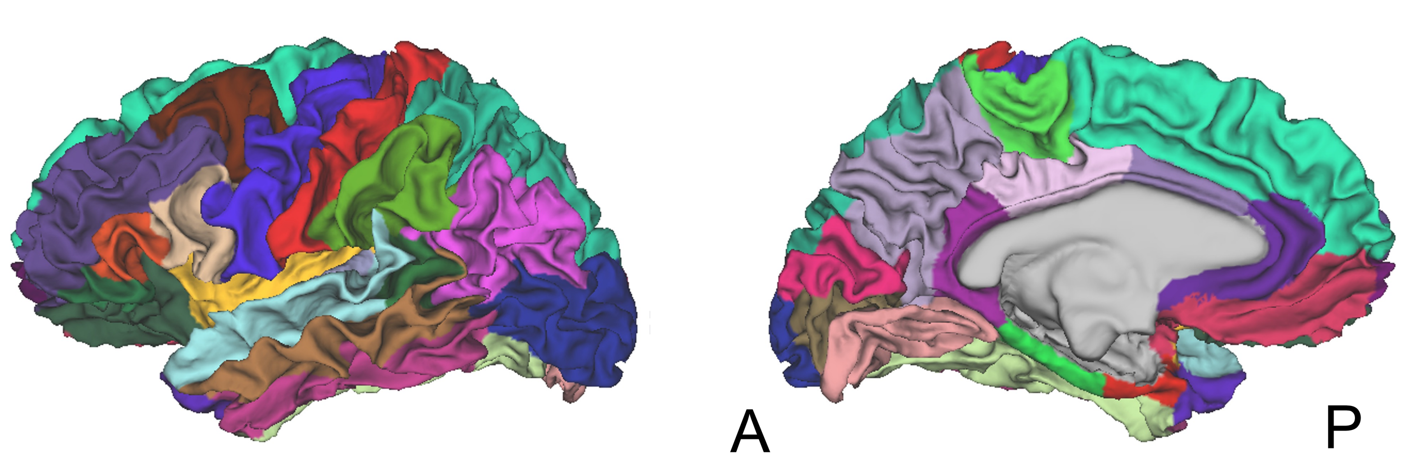


Figure 1. ROIs for one subject.

Target

Fagerstrom Test for Nicotine Dependence (FTND) value, treated as an 11-class classification problem (0 to 10).

Datasets by feature type

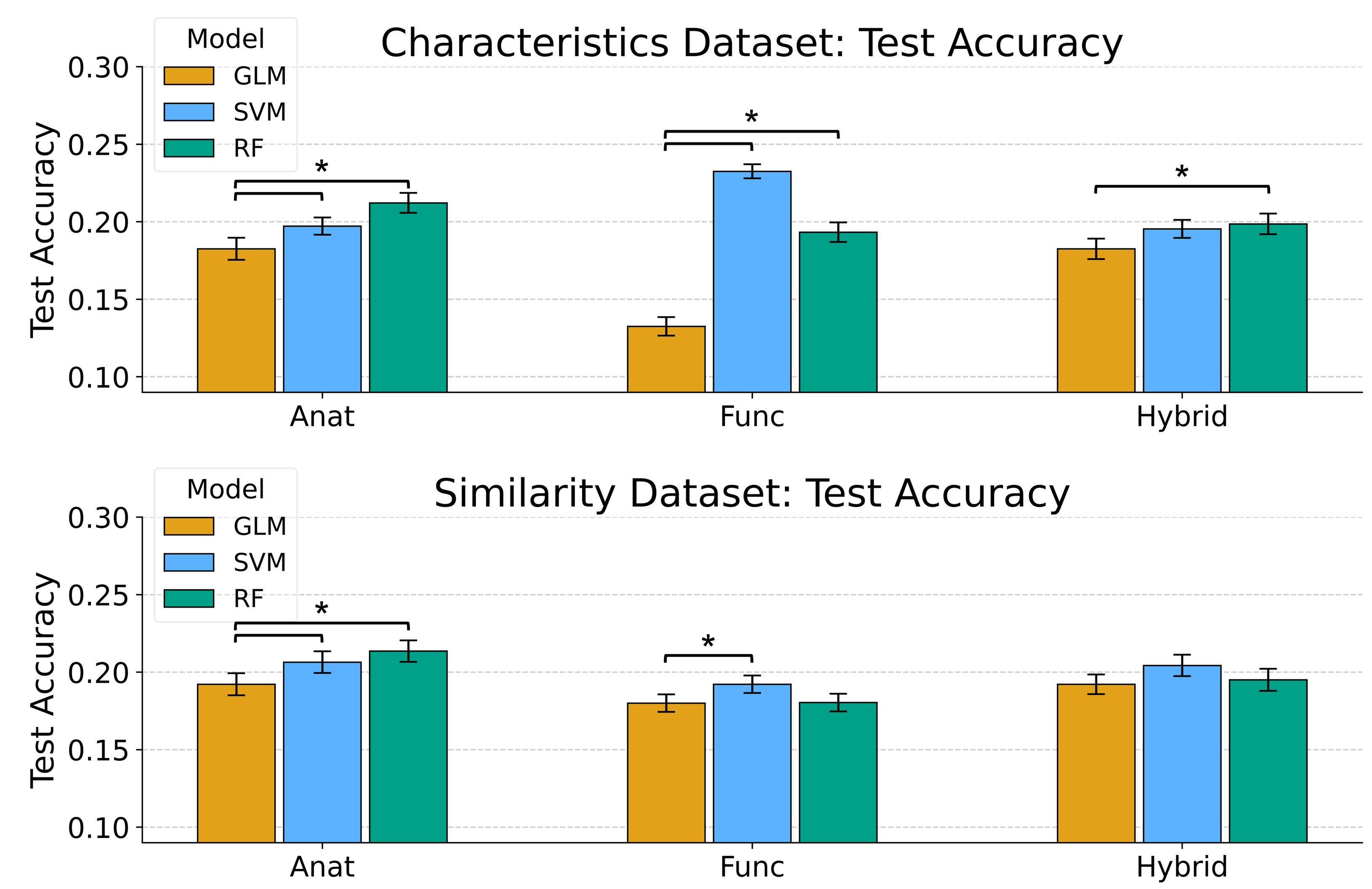
Characteristics (614 features). 8 anatomical features per ROI and time-averaged blood-oxygen-level dependent (BOLD) signals per ROI.

Similarity (4558 features). Pearson correlation between anatomical features and between BOLD signals of all ROI pairs.

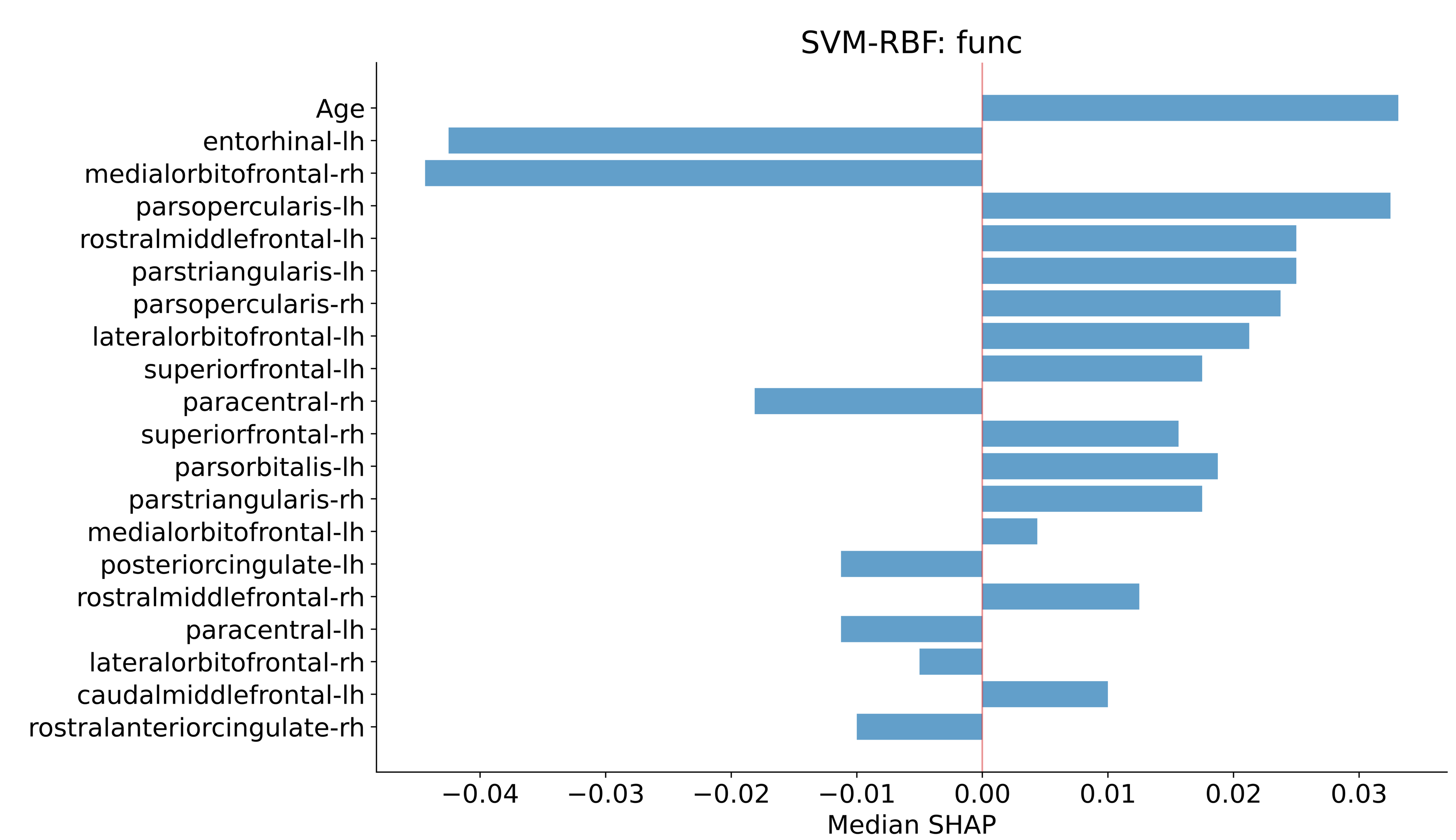
RESULTS

ML MODELS OUTPERFORM GLM

- All models predicted above chance ($p < 0.05$, paired t-test).
- Characteristics: ML > GLM, less so on the hybrid.
  Similarity: ML > GLM on individual modalities but not on the hybrid.
- The overall highest accuracy was 0.232 (SVM, Characteristics, func-only).



ML MODELS IDENTIFY ND-RELATED BRAIN REGIONS



CONCLUSIONS

ML models can achieve higher accuracy and uncover multi-modal neural correlates for predicting ND compared to classical approaches.

Datasets by MRI modality

Anatomical-only / functional-only / anatomical & functional (hybrid) features were considered for each dataset separately (6 datasets in total).

Machine learning models

Random forest (RF), support vector machine with the RBF kernel (SVM).

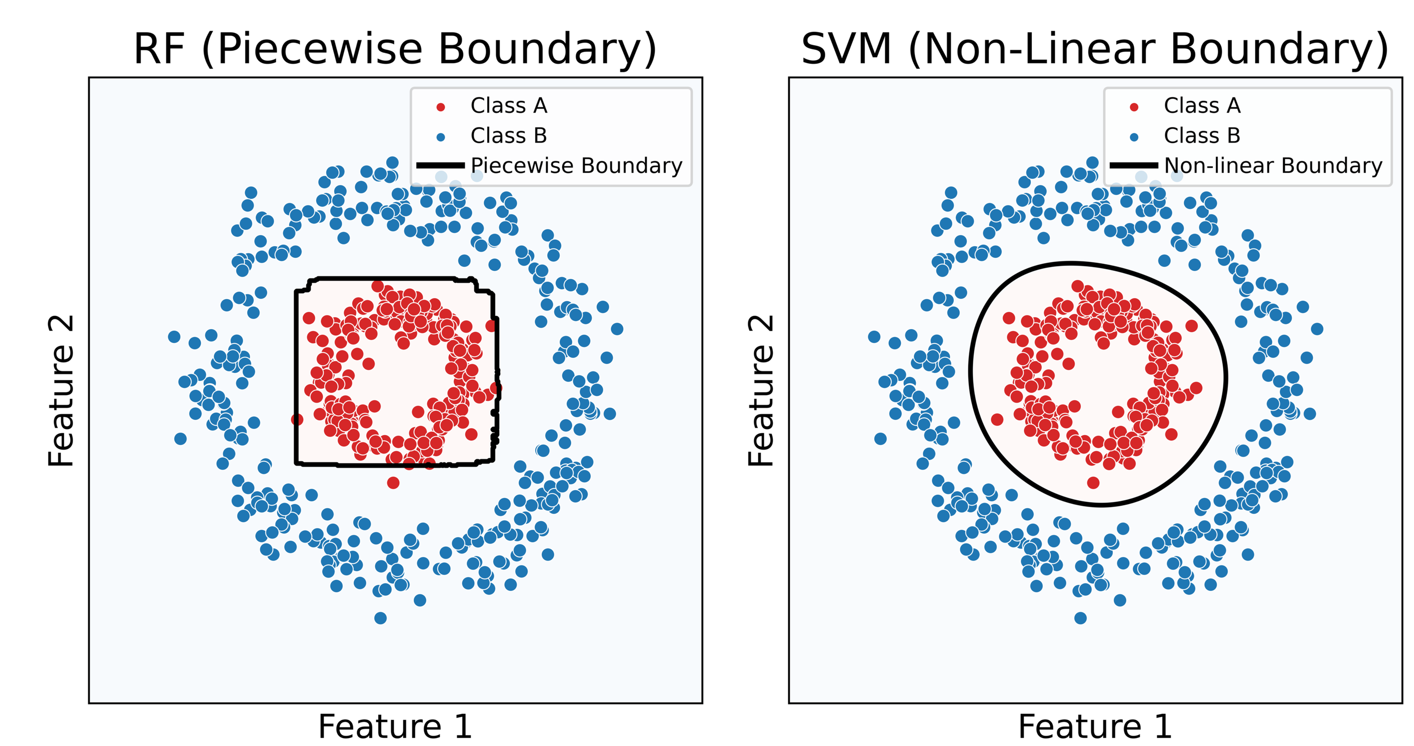


Figure 2. Example decision surfaces of the models.

Feature selection, training and evaluation

All models were trained and tested over 50 random 7:3 splits. For each partition, we selected 30 features: 28 'best' ones (train-sample F-test) together with subject's age and number of years of education. Each model was trained on this subset and evaluated on the test sample.
A random chance model (accuracy = 0.09) and a classical GLM model were used as controls.

Feature importance and directional effects

For GLM, each significant coefficient ($p < 0.05$) was assigned an importance score equal to its absolute value and a directional effect (DE) equal to its sign. Importances and effects for ML models were estimated using permutation-based median SHapley Additive exPlanations (SHAP; *Lundberg and Lee, 2017, In proceedings of NeuroIPS*).

$$SHAP_x(j) = \sum_{X \subseteq K \setminus \{x\}} \left(C_{n-1}^{|S|}\right)^{-1} \left(f_j(X \cup \{x\}) - f_j(X)\right)$$

computes SHAP for feature x and observation j in a model f with features X.