**Supplement: Predicting Depression and Anxiety from Multi-wave Longitudinal Data using Machine Learning**

**Sensitivity Analyses: Description**

To ensure that our conclusions were robust to choice of classifier, we tested two additional classification algorithms: random forests and neural networks. Briefly, random forests are an ensemble learning method that aggregates predictions across a collection of relatively uncorrelated decision trees, each trained on a bootstrapped sample using a random subset of features and tested on unsampled observations. Neural networks, so named because they emulate the way neurons operate in the human brain, contain multiple layers of interconnected nodes (i.e., functions), starting with an input layer that feeds data through a series of transformations across multiple hidden layers to a final output layer where predictions are made. Whereas L2 penalized regression and Random Forests are built on linear functions, neural networks include non-linear functions. Like the logistic regression used in our primary results, both of these models were implemented in Python 3.7 using scikit-learn v22.2 (Pedregosa et al., 2011). For the random forest, we used “extremely randomized trees” known to performed well with limited training examples (Geurts, Ernst, & Wehenkel, 2006), and for the neural network, which typically require more training instances than learned parameters (Alwosheel, van Cranenburgh, & Chorus, 2018), we used a single hidden layer within a multi-layer perceptron (Hinton, 1989).

We included results of the main analyses using L2 penalized logistic regression in the supplemental tables for easy comparison. Additionally, to demonstrate the advantage L2 penalization affords to prediction, we also fit models using traditional (non-penalized) logistic regression as a benchmark of conventional statistical approaches in psychology. While there are countless additional models that could be chosen such as support-vector machine or convolutional neural networks, we are under powered to do model selection (Alwosheel et al., 2018; Hastie, Tibshirani, & Jermone, 2009). These results are not intended to suggest the use of such models. Rather, these supplementary experiments were a sensitivity analysis in order to validate that the same general patterns follow, and that substantially greater accuracy could not be achieved with more sophisticated models (i.e., models containing more learned parameters). Of the many choices of models with more parameters, we chose random forest and multi-layer perceptron as they represent two different types of machine learning algorithms, namely ensemble models, and neural models respectively.

**Supplement Table 1.**

***Depression Prediction Accuracy Results Using Alternative Classifiers***

|  |  |  |  |
| --- | --- | --- | --- |
|  | Combined Waves |  | Individual Waves |
| Models | A12 9 6 3 | A12 9 6 | A12 9 |  | A12 | A9 | A6 | A3 |
| Neural Networks (Multi-Layer Perceptron) |
| CCA Components Alone |   0.698 |   0.719 |   0.637 |  |   0.710 |   0.675 |   0.519 |   0.450 |
| *p* (vs. chance) | < 0.001 | < 0.001 |   0.002 |  | < 0.001 | < 0.001 |   0.348 |   0.862 |
| CCA Components + A12 Dx + Demos |   0.629 |   0.705 |   0.742 |  |   0.656 |   0.658 |   0.495 |   0.621 |
| *p* (vs. chance) |   0.002 | < 0.001 | < 0.001 |  | < 0.001 |   0.001 |   0.538 |   0.005 |
| *p* (vs. A12 Dx + Demos alone) |   0.524 |   0.064 |   0.005 |  |   0.321 |   0.278 |   0.994 |   0.598 |
| Random Forests |
| CCA Components Alone |   0.732 |   0.709 |   0.738 |  |   0.709 |   0.609 |   0.574 |   0.599 |
| *p* (vs. chance) |   < 0.001 | < 0.001 | < 0.001 |  | < 0.001 |   0.011 |   0.054 |   0.016 |
| CCA Components + A12 Dx + Demos |   0.718 |   0.694 |   0.718 |  |   0.725 |   0.594 |   0.538 |   0.622 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 |   0.023 |   0.206 |   0.004 |
| *p* (vs. A12 Dx + Demos alone) |   0.028 |   0.076 |   0.024 |  |   0.009 |   0.806 |   0.986 |   0.578 |
| L2 Penalized Logistic Regression |
| CCA Components Alone | 0.743 | 0.746 | 0.751 |  | 0.745 | 0.669 | 0.608 | 0.556 |
| *p* (vs. chance) | <0.001 | <0.001 | <0.001 |  | <0.001 | 0.003 | 0.114 | 0.413 |
| CCA Components + A12 Dx + Demos | 0.739 | 0.740 | 0.748 |  | 0.744 | 0.679 | 0.639 | 0.599 |
| *p* (vs. chance) | <0.001 | <0.001 | <0.001 |  | <0.001 | 0.002 | 0.013 | 0.082 |
| *p* (vs. A12 Dx + Demos alone) | 0.008 | 0.009 | 0.007 |  | 0.009 | 0.123 | 0.432 | 0.897 |
| Logistic Regression (without Regularization) |
| CCA Components Alone |   0.691 |   0.692 |   0.700 |  |   0.714 |   0.641 |   0.574 |   0.556 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 | < 0.001 |   0.057 |   0.114 |
| CCA Components + A12 Dx + Demos |   0.682 |   0.692 |   0.697 |  |   0.719 |   0.642 |   0.597 |   0.600 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 |   0.001 |   0.019 |   0.015 |
| *p* (vs. A12 Dx + Demos alone) |   0.159 |   0.119 |   0.092 |  |   0.041 |   0.409 |   0.823 |   0.857 |

Note: Cells contain area under the receiver operating characteristics curve (AUC) values. Definitions: "A" = Age; "A12 Dx" = Age 12 Depression Diagnostic Status; "Demos" = Demographics (sex, race, and ethnicity).

**Supplement Table 2.**

***Anxiety Prediction Accuracy Results Using Alternative Classifiers***

|  |  |  |  |
| --- | --- | --- | --- |
|  | Combined Waves |  | Individual Waves |
| Models | A12 9 6 3 | A12 9 6 | A12 9 |  | A12 | A9 | A6 | A3 |
| Neural Networks (Multi-Layer Perceptron) |
| CCA Components Alone |   0.743 |   0.699 |   0.773 |  |   0.781 |   0.705 |   0.688 |   0.537 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 | < 0.001 | < 0.001 |   0.176 |
| CCA Components + A12 Dx + Demos |   0.756 |   0.773 |   0.760 |  |   0.783 |   0.797 |   0.810 |   0.791 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| *p* (vs. A12 Dx + Demos alone) |   0.738 |   0.526 |   0.669 |  |   0.381 |   0.188 |   0.082 |   0.212 |
| Random Forests |
| CCA Components Alone |   0.779 |   0.779 |   0.772 |  |   0.763 |   0.670 |   0.670 |   0.516 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 | < 0.001 | < 0.001 |   0.354 |
| CCA Components + A12 Dx + Demos |   0.813 |   0.811 |   0.801 |  |   0.789 |   0.764 |   0.785 |   0.735 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 | < 0.001 | < 0.001 |   0.000 |
| *p* (vs. A12 Dx + Demos alone) |   0.066 |   0.074 |   0.133 |  |   0.279 |   0.668 |   0.338 |   0.917 |
| L2 Penalized Logistic Regression |
| CCA Components Alone | 0.777 | 0.784 | 0.787 |  | 0.788 | 0.749 | 0.711 | 0.621 |
| *p* (vs. chance) | <0.001 | <0.001 | <0.001 |  | <0.001 | <0.001 | <0.001 | <0.001 |
| CCA Components + A12 Dx + Demos | 0.807 | 0.811 | 0.812 |  | 0.810 | 0.805 | 0.799 | 0.762 |
| *p* (vs. chance) | <0.001 | <0.001 | <0.001 |  | <0.001 | <0.001 | <0.001 | <0.001 |
| *p* (vs. A12 Dx + Demos alone) | 0.046 | 0.039 | 0.025 |  | 0.027 | 0.022 | 0.069 | 0.840 |
| Logistic Regression (without Regularization) |
| CCA Components Alone |   0.728 |   0.758 |   0.784 |  |   0.796 |   0.731 |   0.695 |   0.672 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| CCA Components + A12 Dx + Demos |   0.745 |   0.768 |   0.796 |  |   0.803 |   0.790 |   0.777 |   0.791 |
| *p* (vs. chance) | < 0.001 | < 0.001 | < 0.001 |  | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| *p* (vs. A12 Dx + Demos alone) |   0.820 |   0.587 |   0.240 |  |   0.128 |   0.232 |   0.455 |   0.208 |

Note: Cells contain area under the receiver operating characteristics curve (AUC) values. Definitions: "A" = Age; "A12 Dx" = Age 12 Anxiety Diagnostic Status; "Demos" = Demographics (sex, race, and ethnicity).

References

Alwosheel, A., van Cranenburgh, S., & Chorus, C. G. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of choice modelling, 28*, 167-182.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning, 63*(1), 3-42.

Hastie, T., Tibshirani, R., & Jermone, F. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd ed.). New York, NY: Springer-Verlag.

Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence 40*(1), 185-234.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research, 12*, 2825-2830.