Automated Machine Learning in Brain Predictive Modelling: A data-driven approach to Predict Brain Age from Cortical Anatomical Measures

Jessica Dafflon (jessica.de_faria_dafflon@kcl.ac.uk)
Centre for Neuroimaging Sciences, Institute of Psychiatry,
Kings College London, UK

James H. Cole (james.cole@kcl.ac.uk)
Centre for Neuroimaging Sciences, Institute of Psychiatry,
Kings College London, UK

Federico Turkheimer (federico.turkheimer@kcl.ac.uk)
Centre or Neuroimaging Sciences, Institute of Psychiatry,
Kings College London, UK

Robert Leech (robert.leech@kcl.ac.uk)
Centre for Neuroimaging Sciences, Institute of Psychiatry,
Kings College London, UK

Mathew A. Harris (Mat.Harris@ed.ac.uk)
Division of Psychiatry,
University of Edinburgh, UK

Simon R. Cox (Simon.Cox@ed.ac.uk)
Department of Psychology,
University of Edinburgh, UK

Heather C. Whalley (Heather.Whalley@ed.ac.uk)
Division of Psychiatry,
University of Edinburgh, UK

Andrew M. McIntosh (andrew.mcintosh@ed.ac.uk)
Division of Psychiatry,
University of Edinburgh, UK

Peter J. Hellyer (peter.hellyer@kcl.ac.uk)
Centre for Neuroimaging Sciences, Institute of Psychiatry,
Kings College London, UK

Abstract

The use of machine learning (ML) algorithms significantly increased in neuroscience. However, from the vast extent of possible ML algorithms, which one is the optimal model to predict the feature of interest? What are the best parameters for such a model? Given the plethora of possible answers to these questions, in the last years, automated machine learning (autoML) has been gaining attention. Here, we used TPOT which is a tree-based pipeline optimisation tool that scans a model space of models, their hyperparameters and finds the model with the highest accuracy. To explore autoML approaches and evaluate their efficacy within neuroimaging datasets, we choose a problem that has been the focus of previous extensive study: brain age prediction. Without any prior knowledge, TPOT could scan through the model space and create pipelines that outperform the state-of-the-art accuracy for Freesurfer-based models (MAE: 4.89 years) using only the cortical thickness and subcortical volume information. It also suggests interesting ensembles that do not match the current most used models for brain prediction but generalise well to an unseen dataset (MAE: 4.94 years). Thus, TPOT can be used as a data-driven approach to find ML models that accurately predict brain age.

Keywords: predictive modelling; automated machine learning; age prediction, neuroimaging;

Introduction

The last few decades have seen significant progress in neuroimaging methodologies and techniques focused on identifying the brain features, structural or functional associated with brain disease states. With the advance of ML algorithms, which learn to identify significant and generalisable structures

in datasets, the field of neuroimaging is moving towards becoming a predictive science, where inference can be made on the individual level rather than the general behaviour of a group of individuals (Glaser, Benjamin, Farhoodi, & Kording, 2019; Liem et al., 2017; Yarkoni & Westfall, 2017). The aim of predictive modelling is to use ML algorithms to learn patterns in a large dataset and subsequently build a model to predict an independent variable of interest. The model's performance is then evaluated in an independent dataset and new predictions can be generated when passing an unseen dataset to the trained model (Cole & Franke, 2017; Glaser et al., 2019; Liem et al., 2017; Yarkoni & Westfall, 2017). However, this raises various problems for analysis: Firstly, the richness of multivariate data, can lead to significant problems with *over*-estimation of the model, leading to a loss of generalisation; and secondly the sheer *mass* of learning approaches that are available for datasets with a vast array of different properties and approaches, provides a bewildering set of choices for the practitioner; each with advantages and disadvantages both in terms of generalisation, and computational complexity. Hence, the optimal application of ML technology requires the answer to at least three questions: For the data at hand, which one is the optimal model to predict the feature of interest? What are its best parameters? Is the chosen model and parameters overfitting to the dataset? The fact that the answer to these questions are often arbitrary and defined only on *prior*-wisdom, is a challenge for neuroimaging which continues to face a significant replication crisis.

Finding the best model and its hyper-parameters in a systematic, timely and computationally efficient way is the aim of autoML (Hutter, Kotthoff, & Vanschoren, 2019). This approach takes advantage of the complexity in the underlying dataset, and while searching for the best model it optimises the *performance*, whilst simultaneously attempting to maximise the generalisability of resulting predictions.

In this paper, we explore a genetic autoML approach and evaluate its efficacy for both identifying and predicting patterns within neuroimaging datasets. As a test-case, we choose to look at a problem that has been the focus of previous extensive study: the use of structural brain data (in this case, cortical thickness) to predict the subject's age. We first analysed the TPOT (Olson, Bartley, Urbanowicz, & Moore, 2016; Olson, Urbanowicz, et al., 2016) performance, then we analyse the ML models suggested by TPOT and how well they generalised to a validation set.

Methods

For this analysis, T1-weighted MRI scans from N=10,307 healthy subjects (age range 18-89 years, mean age = 59.40) were obtained from 14 publicly available datasets (Cole et al., 2017) and the UK Biobank (Sudlow et al., 2015). All subjects were screened to exclude those with major neurological or psychiatric diseases and were divided into train (n=1,030), test (n=464) and validation (n=8,813) set in a pseudo-random fashion to ensure that age and sex distribution was balanced in all 3 groups.

The feature space under analysis consisted of 116 features that describe the thickness of different sub-cortical and cortical brain areas and were segmented using the Desikan-Killiany atlas (Desikan et al., 2006) and the Freesurfer recon-all (Dale, Fischl, & Sereno, 1999) (Freesurfer version v6.0) from the individual's T1-weighted brain images.

To perform automatic machine learning we used TPOT (Olson, Bartley, et al., 2016; Olson, Urbanowicz, et al., 2016), which is a tree-based pipeline optimisation tool that uses genetic programming to search different hyper-parameters combinations and choose the most suitable model and its parameters for solving a classification or regression machine learning problem with high accuracy. It does so by finding the models with the best cross-validated performances on the training set for each generation and applying local perturbations (e.g., mutation and cross-over). This process is repeated for a specified number of generations, and the best performing model is returned to the user. One particularly interesting feature from TPOT is that it creates ensembles, that is, it combines the prediction from different ML together into a single pipeline, in order to enhance the model's accuracy.

The model space consisted of a pool of 11 commonly used linear and non-linear ML algorithms commonly used to predict brain-age (Aycheh et al. (2018); Becker, Klein, Wachinger, Initiative, et al. (2018); Cole, Leech, Sharp, and Alzheimer's Disease Neuroimaging Initiative (2015), Table 1). We also extended the current TPOT software to include Gaussian Process Regressors (Pedregosa et al., 2011) and Relevance Vector Machines [1], as those are common methods used fro brain-age prediction.

Results

The analysed method is stochastic so for simplicity here we are showing only the results obtained for a predefined random seed.

Analysis of TPOT Performance

Our first step was to evaluate if TPOT was able to increase the prediction accuracy over generations. Fig 1 illustrates the change in prediction performance evaluated using the MAE (Mean Accuracy Error, lower values represent a better accuracy) for the different models for every generation. During the first two generations, TPOT is still exploring the model space and there is a large variance of the model's MAEs. As we can see by the decrease in the error bars, after the first 60 generations, the model pool becomes more accurate. However, throughout out the entire analysis we can observe three main groups of models: a group that poorly predict age (MAE $>$ 35 years), a second group where the accuracy oscillates around 15 $<$MAE $<$ 2, and a third group MAE $<$ 10.

[1] https://github.com/JamesRitchie/scikit-rvm

Table 1: List of used Regression Algorithms

Algorithms	Sklearn Implementation
Elastic Net model with iterative fitting along a regularisation path	ElasticNetCV
Randomised Decision Trees on sub-samples of the dataset	ExtraTreesRegressor
k-Nearest Neighbours Regression	KNeighborsRegressor
Cross-validated Lasso using the LARS algorithm	LassoLarsCV
Linear Support Vector Regression	LinearSVR
Linear Least squares with l2 regularisation	Ridge
Random Forest Regressor	RandomForrestRegressor
Ordinary Least Squares Linear Regression	LinearRegression
Decision Tree Regressor	DecisionTreeRegressor
Gaussian process regression	GaussianProcessRegressors
Relevance Vector Machine	RVR

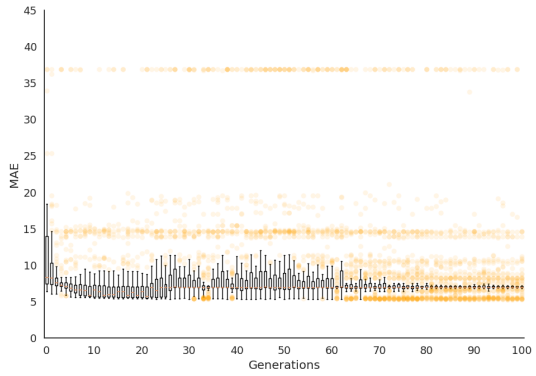

Figure 1: **Prediction performance for every generation** The prediction performance of the trained model is evaluated by the MAE. The yellow dots represent models configurations that were treated as outliers for the boxplot calculation. Note that until the 25-th generation there is high variability in the median of the predicted accuracy.

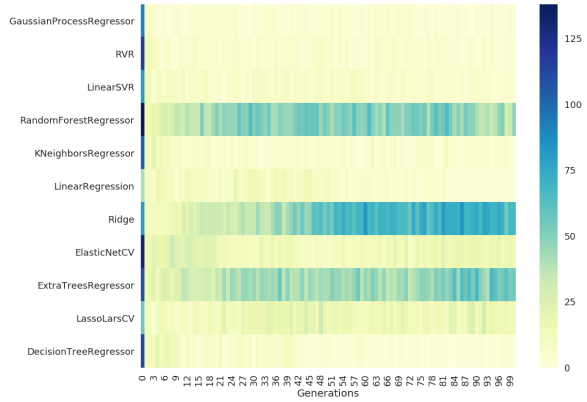

Figure 2: **Sum of the different models for every generation:** We evaluated the number of models for each one of the different model families for every generation. Random Forests, Ridge Regression and Extra Trees Regressors are the 3 families with the highest model count per generation

TPOT Suggested Models

We also analysed the presence of each model at every generation. As illustrated in Fig 2, Random Forests, Ridge Regression and Extra Trees Regressors achieve high accuracy and are passed on into future generations.

TPOT Generalisability

For this specific analysis, the ensemble of LinearRegression and two combined ExtraTreesRegressor performs with the highest accuracy. To validate whether this model shows evidence of overfitting to the training dataset we applied the obtained model to a left-out validation dataset. While the MAE on the test set was 4.89 years, the obtained MAE for the validation dataset was 4.94 years.

Discussion

In the course of this work we have demonstrated that: (1) TPOT can be used to identify a good predictive model (2) For this application there is no single analysis model that ”best” predicts age from the underlying structural imaging data; the ”best” models identified by TPOT consist of a mix of random forest, randomised decision trees and cross-validated lasso. (3) The accuracy of the models suggested by TPOT is better than recent brain age models. When comparing the accuracy of different studies, it is important to take into account the age range of the analysed sample, as age prediction in a small range has less variability than in a large range. In fact, using a sample with subjects aged 45-91 Aycheh et al. (2018) obtained a MAE of 4.02 years. While, Valizadeh, Hänggi, Mérillat, and Jäncke (2017) had a similar age range as that described in this abstract, they do not report the MAE for the entire sample and use instead 3 age groups (8-18 years, 18-

65 and 65-96 years) to test the accuracy of different models. In general, Valizadeh et al. (2017) reported lower accuracy for the older group with MAE ranging between 4.90 and 14.23 years, when using only the thickness information. On the other hand, Liem et al. (2017) using only the cortical thickness reported a MAE of 5.95 years (analysed age range 19 - 92 years, mean = 58.68). One of the main advantages of the approach here proposed is that it does not make any assumptions about the underlying statistics of the dataset and does not require any fine-tuning of the model of choice but still achieve the state-of-the-art accuracy. (4) We further evaluated the decisions made by TPOT in a totally independent dataset, to explore the stability of the TPOT generated analysis pipeline and found that the predictions generalise well to this 'unseen' dataset. Therefore, the autoML approach shown in this paper can be used as a data-driven method to learn patterns in the data and to avoid common pitfalls from ML algorithms such as overfitting.

Despite the obtained high accuracy of the model using only the cortical thickness information, we think that the MAE can be further reduced by including the surface area information obtained from Freesurfer to the TPOT models. Previous studies have observed an increase in accuracy by adding the different cortical anatomical measures to the models (Liem et al., 2017; Valizadeh et al., 2017; Wang et al., 2014). The accuracy of our current approach might also improve by better splitting the test, training and validation sets. With the current settings, TPOT can only use around 2500 subject for the test and training dataset in order to explore the defined model space. As we have a large sample size available, it would be interesting to add more subjects to these groups. These two refinements will be tested in future works.

References

Aycheh, H. M., Seong, J.-K., Shin, J.-H., Na, D. L., Kang, B., Seo, S. W., & Sohn, K.-A. (2018). Biological brain age prediction using cortical thickness data: A large scale cohort study. *Front Aging Neurosci*, *10*, 252. doi: 10 .3389/fnagi.2018.00252

Becker, B. G., Klein, T., Wachinger, C., Initiative, A. D. N., et al. (2018). Gaussian process uncertainty in age estimation as a measure of brain abnormality. *NeuroImage*, *175*, 246–258.

Cole, J. H., & Franke, K. (2017). Predicting age using neuroimaging: Innovative brain ageing biomarkers. *Trends Neurosci*, *40*(12), 681-690. doi: 10.1016/j.tins.2017.10 .001

Cole, J. H., Leech, R., Sharp, D. J., & Alzheimer's Disease Neuroimaging Initiative. (2015). Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Ann Neurol*, *77*(4), 571-81. doi: 10.1002/ana .24367

Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, *163*, 115–124.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179–194.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006, Jul). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968-80. doi: 10.1016/j.neuroimage.2006.01.021

Glaser, J. I., Benjamin, A. S., Farhoodi, R., & Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. *Prog Neurobiol*, *175*, 126-137. doi: 10.1016/j.pneurobio.2019.01.008

Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). Automatic machine learning: methods, systems, challenges. *Challenges in Machine Learning*.

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Masouleh, S. K., Huntenburg, J. M., ... others (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *Neuroimage*, *148*, 179–188.

Olson, R. S., Bartley, N., Urbanowicz, R. J., & Moore, J. H. (2016). Evaluation of a tree-based pipeline optimization tool for automating data science. In *Proceedings of the genetic and evolutionary computation conference 2016* (pp. 485–492).

Olson, R. S., Urbanowicz, R. J., Andrews, P. C., Lavender, N. A., Moore, J. H., et al. (2016). Automating biomedical data science through tree-based pipeline optimization. In *European conference on the applications of evolutionary computation* (pp. 123–137).

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

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., ... Collins, R. (2015). Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, *12*(3), e1001779. doi: 10.1371/journal .pmed.1001779

Valizadeh, S. A., Hänggi, J., Mérillat, S., & Jäncke, L. (2017). Age prediction on the basis of brain anatomical measures. *Hum Brain Mapp*, *38*(2), 997-1008. doi: 10.1002/ hbm.23434

Wang, J., Li, W., Miao, W., Dai, D., Hua, J., & He, H. (2014). Age estimation using cortical surface pattern combining thickness with curvatures. *Med Biol Eng Comput*, *52*(4), 331-41. doi: 10.1007/s11517-013-1131-9

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.