

# Predicting Response to the Antidepressant Bupropion using Pretreatment fMRI

Kevin P. Nguyen<sup>(✉)</sup>, Cherise Chin Fatt, Alex Treacher, Cooper Mellema, Madhukar H. Trivedi, and Albert Montillo

University of Texas Southwestern Medical Center

**Abstract.** Major depressive disorder is a primary cause of disability in adults with a lifetime prevalence of 6-21% worldwide. While medical treatment may provide symptomatic relief, response to any given antidepressant is unpredictable and patient-specific. The standard of care requires a patient to sequentially test different antidepressants for 3 months each until an optimal treatment has been identified. For 30-40% of patients, no effective treatment is found after more than one year of this trial-and-error process, during which a patient may suffer loss of employment or marriage, undertreated symptoms, and suicidal ideation. This work develops a predictive model that may be used to expedite the treatment selection process by identifying for individual patients whether the patient will respond favorably to bupropion, a widely prescribed antidepressant, using only pretreatment imaging data. This is the first model to do so for individuals for bupropion. Specifically, a deep learning predictor is trained to estimate the 8-week change in Hamilton Rating Scale for Depression (HAM-D) score from pretreatment task-based functional magnetic resonance imaging (fMRI) obtained in a randomized controlled antidepressant trial. An unbiased neural architecture search is conducted over 800 distinct model architecture and brain parcellation combinations, and patterns of model hyperparameters yielding the highest prediction accuracy are revealed. The winning model identifies bupropion-treated subjects who will experience remission with the number of subjects needed-to-treat (NNT) to lower morbidity of only 3.2 subjects. It attains a substantially high neuroimaging study effect size explaining 26% of the variance ( $R^2 = 0.26$ ) and the model predicts post-treatment change in the 52-point HAM-D score with an RMSE of 4.71. These results support the continued development of fMRI and deep learning-based predictors of response for additional depression treatments.

**Keywords:** depression · treatment response · fMRI · neuroimaging · deep learning · neural architecture search

## 1 Introduction

Major depressive disorder (MDD) has a lifetime prevalence of 6-21% worldwide and is a major cause of disability in adults [12]. Though half of MDD cases are treated with medication, there are dozens of antidepressants available and a

patient’s response to each is highly unpredictable [7]. The current standard in healthcare entails a long trial-and-error process in which a patient tries a series of different antidepressants. The patient must test each drug for up to 3 months, and if satisfactory symptomatic improvement is not achieved within this time, the clinician modifies the dosage or selects a different drug to test next. This trial-and-error process may take months to years to find the optimal treatment, during which patients suffer continued debilitation, including worsening symptoms, social impairment, loss of employment or marriage, and suicidal ideation. It has been shown that 30-40% of patients do not find adequate treatment after a year or more of drug trials [19, 22]. Consequently, a predictive tool that helps prioritize the selection of antidepressants that are best suited to each patient would have high clinical impact.

This work demonstrates the use of deep learning and pretreatment task-based fMRI to predict long-term response to bupropion, a widely used antidepressant with a response rate of 44% [15]. An accurate screening tool that distinguishes bupropion responders from non-responders using pretreatment imaging would reduce morbidity and unnecessary treatment for non-responders and prioritize the early administration of bupropion for responders.

The use of functional magnetic imaging (fMRI) measurements to infer quantitative estimates of bupropion response is motivated by evidence for an association between fMRI and antidepressant response. For example, resting-state activity in the anterior cingulate cortex as well as activity evoked by reward processing tasks in the anterior cingulate cortex and amygdala have all been associated with antidepressant response [17, 13, 16].

In this work, predictive models of individual response to bupropion treatment are built using deep learning and pretreatment, task-based fMRI from a cohort of MDD subjects. The novel contributions of this work are: 1) the first tool for accurately predicting long-term bupropion response, and 2) the use of an unbiased neural architecture search (NAS) to identify the best-performing model and brain parcellation from 800 distinct model architecture and parcellation combinations.

## 2 Methods

### 2.1 Materials

Data for this analysis comes from the EMBARC clinical trial[23], which includes 37 subjects who were imaged with fMRI at baseline and then completed an 8-week trial of bupropion XL. To track symptomatic outcomes, the 52-point Hamilton Rating Scale for Depression (HAM-D) was administered at baseline and week 8 of antidepressant treatment. Higher HAM-D scores indicate greater MDD severity. Quantitative treatment response for each subject was defined as  $\Delta\text{HAM-D} = \text{HAM-D}(\text{week } 8) - \text{HAM-D}(\text{baseline})$ , where a negative  $\Delta\text{HAM-D}$  indicates improvement in symptoms. The mean  $\Delta\text{HAM-D}$  for these subjects was  $-5.98 \pm 6.25$ , suggesting a large variability in individual treatment outcomes. For

comparison, placebo-treated subjects in this study exhibited a mean  $\Delta$ HAMD of  $-6.70 \pm 6.93$ .

**Image Acquisition** Subjects were imaged with resting-state and task-based fMRI (gradient echo-planar imaging at 3T, TR of 2000ms,  $64 \times 64 \times 39$  image dimensions, and  $3.2 \times 3.2 \times 3.1$ mm voxel dimensions). Resting-state fMRI was acquired for 6 minutes. Task-based fMRI was acquired immediately afterwards for 8 minutes during a well-validated block-design reward processing task assessing reactivity to reward and punishment [8, 11]. In this task, subjects must guess in the *response phase* whether an upcoming number will be higher or lower than 5. They are then informed in the *anticipation phase* if the trial is a “possible win”, in which they receive a \$1 reward for a correct guess and no punishment for an incorrect guess, or a “possible loss”, in which they receive a -\$0.50 punishment for an incorrect guess and no reward for a correct guess. In the *outcome phase*, they are then presented with the number and the outcome of the trial.

## 2.2 Image Preprocessing

Both resting-state and task-based fMRI images were preprocessed as follows. Frame-to-frame head motion was estimated and corrected with FSL MCFLIRT, and frames where the norm of the fitted head motion parameters was  $> 1$ mm or the intensity Z-score was  $> 3$  were marked as outliers. Images were then skull-stripped using a combination of FSL BET and AFNI Automask. To perform spatial normalization, fMRI images were registered directly to an MNI EPI template using ANTs. This coregistration approach has been shown to better correct for nonlinear distortions in EPI acquisitions compared to T1-based coregistration [2, 6]. Finally, the images were smoothed with a 6 mm Gaussian filter.

Predictive features were extracted from the preprocessed task-based fMRI images in the form of contrast maps (i.e. spatial maps of task-related neuronal activity). Each task-based fMRI image was fit to a generalized linear model,

$$\mathbf{Y} = \mathbf{X} \times \boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where  $\mathbf{Y}$  is the *time*  $\times$  *voxels* matrix of BOLD signals,  $\mathbf{X}$  is the *time*  $\times$  *regressors* design matrix,  $\boldsymbol{\beta}$  is the *regressors*  $\times$  *voxels* parameter matrix, and  $\boldsymbol{\epsilon}$  is the residual error, using SPM12. The design matrix  $\mathbf{X}$  was defined as described in [11] and included regressors for the response, anticipation, outcome, and inter-trial phases of the task paradigm. In addition, a reward expectancy regressor was included, which had values of +0.5 during the anticipation phase for “possible win” trials and  $-0.25$  during the anticipation phase for “possible loss” trials. These numbers correspond to the expected value of the monetary reward/punishment in each trial. In addition to these task-related regressors and their first temporal derivatives, the head motion parameters and outlier frames were also included as regressors in  $\mathbf{X}$ .

After fitting the generalized linear model, contrast maps for anticipation ( $C_{antic}$ ) and reward expectation ( $C_{re}$ ) were computed from the fitted  $\beta$  coefficients:

$$\begin{aligned} C_{antic} &= \beta_{\text{anticipation}} - \beta_{\text{inter-trial}} \\ C_{re} &= \beta_{\text{reward expectation}} \end{aligned}$$

To extract region-based features from these contrast maps, three custom, study-specific brain parcellations (later referred to as *ss100*, *ss200* and *ss400*) were generated with 100, 200, and 400 regions-of-interest (ROIs) from the resting-state fMRI data using a spectral clustering method [5]. Each parcellation was then used to extract mean contrast values per ROI. The performance achieved with each of these custom parcellations, as well as a canonical functional atlas generated from healthy subjects (Schaefer 2018, 100 ROIs) [20], is compared in the following experiments.

### 2.3 Construction of Deep Learning Predictive Models

Dense feed-forward neural networks were constructed to take the concatenated ROI mean values from the two contrast maps as inputs and predict 8-week  $\Delta$ HAMD. Rather than hand-tuning model hyperparameters, a random search was conducted to identify a high-performing model for predicting response to bupropion. The random search is an unbiased neural architecture search (NAS) that was chosen because it has been shown to outperform grid search [1] and when properly configured can provide performance competitive with leading NAS methods such as ENAS [14].

200 architectures were sampled randomly from a uniform distribution over a defined hyperparameter space (**Table 1**) and then used to construct models that were trained in parallel on 4 NVIDIA P100 GPUs. All models contained a single neuron output layer to predict  $\Delta$ HAMD and were trained with the *Nadam* optimizer, 1000 maximum epochs, and early stopping after 50 epochs without decrease in validation root mean squared error (RMSE).

The combination of 200 model architectures with 4 different parcellations resulted in a total of 800 distinct model configurations that were tested. To ensure robust model selection and to accurately estimate generalization performance, these 800 model configurations were tested with a nested K-fold cross-validation scheme with 3 outer and 3 inner folds. Although a single random split is commonly used in place of the outer validation loop, a nested cross-validation ensures that no test data is used during training or model evaluation and provides an unbiased estimate of final model performance [24]. Within each outer fold, the best-performing model was selected based on mean root mean squared error (RMSE) over the inner folds. The model was then retrained on all training and validation data from the inner folds and final generalization performance was evaluated on the held-out test data of the outer fold. Repeating this process for each outer fold yielded 3 best-performing models, and the mean test performance of these models is reported here.

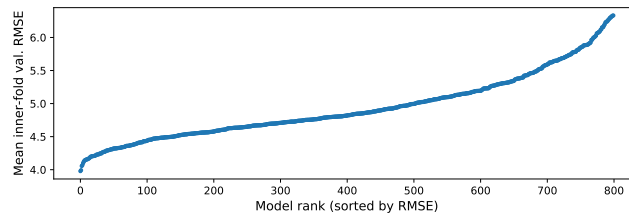
**Table 1.** Hyperparameter space defined for the random neural architecture search. For each model, one value was randomly selected from each of the first set of hyperparameters; for each layer in each model, one value was randomly selected from the second set of hyperparameters.

Hyperparameter	Possible values
Per-model hyperparameters	
Number of dense hidden layers	1, 2, 3, 4, 5
Number of neurons in 1 <sup>st</sup> hidden layer	$32N$ for $N \in [1, \dots, 16]$
Activation for all layers	Leaky ReLU, ReLU, ELU, PReLU
Learning rate	$0.0001n$ for $n \in [1, \dots, 50]$
Per-layer hyperparameters	
% decrease in neurons from previous layer	None, 0.25, 0.5, 0.75
Weight regularization	$L_1$ , $L_2$ , $L_1$ and $L_2$
Activity regularization	$L_1$ , $L_2$ , $L_1$ and $L_2$
Batch normalization	Yes, No
Dropout rate	0, 0.3, 0.5, 0.7

### 3 Results and Discussion

#### 3.1 Neural Architecture Search (NAS)

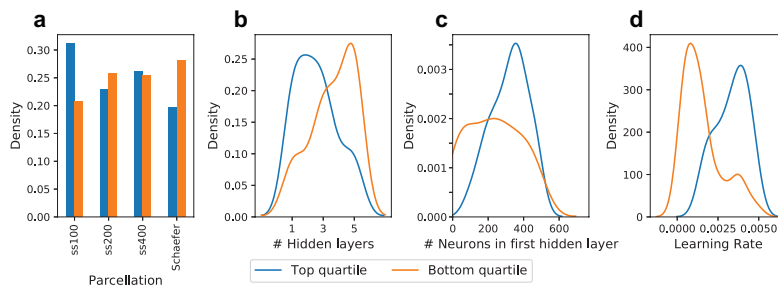
Results indicate that the NAS is beneficial. In particular, a wide range of validation RMSE was observed across the 800 tested model configurations (**Fig. 1**). Certain models performed particularly well achieving RMSE approaching 4.0, while other model architectures were less suitable. NAS helped identify high-performing configurations expediently.



**Fig. 1.** Mean inner validation fold RMSE of the 800 model architecture & parcellation combinations evaluated in the unbiased neural architecture search. Results from one outer cross-validation fold are illustrated here, and findings for the other two folds were similar.

The information from the NAS can be examined for insight into what configurations constitute high versus low performing models and whether the ranges of hyperparameters searched were sufficiently broad. Towards this end, the hyperparameter distributions of the top and bottom quartiles of these 800 model

configurations, sorted by RMSE, were compared. Substantial differences in the hyperparameter values that yielded high and low predictive accuracy are observed (**Fig. 2**). Notably, the custom, study-specific parcellation with 100 ROIs (*ss100*) provided significantly better RMSE than the “off-the-shelf” Schaefer parcellation ( $p = 0.023$ ). Additionally, the top quartile of models using *ss100* used fewer layers (1–2), but more neurons (384–416) in the first hidden layer, compared to the bottom quartile of models. Note that unlike in a parameter sensitivity analysis, where ideal results exhibit a uniform model performance over a wide range of model *parameters*, in a neural architecture search, an objective is to demonstrate adequate coverage over a range of *hyperparameters*. This objective is met when local performance maxima are observed. This is shown in (**Fig. 2b,c,d**) where peaks in the top quartile (blue curve) of model architectures are evident.



**Fig. 2.** Hyperparameter patterns for the top (blue) and bottom (orange) quartiles of the 800 model configurations evaluated in the unbiased neural architecture search. Representative results for one of the outer cross-validation folds are presented. **a:** Top quartile models tended to use the *ss100* parcellation, while bottom quartile models tended to use the Schaefer parcellation. **b-d:** Distributions of three selected hyperparameters compared for the top and bottom quartiles of model configurations, revealing the distinct patterns of hyperparameters for high-performing models. The top quartile of model architectures have fewer layers (peaking at 1–2) but more neurons in the first hidden layer (peaking at 384–416 neurons).

The best performing model configuration used an architecture with two hidden layers and the 100-ROI study-specific parcellation (*ss100*). Regression accuracy in predicting  $\Delta$ HAMD in response to bupropion treatment was RMSE 4.71 and  $R^2$  0.26. This  $R^2$  value (95% confidence interval 0.12–0.40 for  $n = 37$ ) constitutes a highly significant effect size for a neuroimaging study where effect sizes are commonly much lower, e.g. 0.01–0.10 in [3] and 0.09–0.15 in [21]. Furthermore, this predictor identifies individuals who will experience clinical remission ( $\text{HAMD}(\text{week } 8) \leq 7$ ) with number of subjects needed-to-treat (NNT) of 3.2 subjects and AUC of 0.71. This NNT indicates that, on average, one additional remitter will be identified for every 3 individuals screened by this predictor.

**Table 2.** Performance of the best model configuration from the neural architecture search. To obtain classifications of remission, the model’s regression outputs were thresholded post-hoc using the clinical criteria for MDD remission (HAMD(week 8) < 7). *RMSE*: root mean squared error, *NNS*: number needed to screen, *PPV*: positive predictive value, *AUC*: area under the receiver operating characteristic curve.

Target	Performance
$\Delta$ HAMD	$R^2$ 0.26 (95% CI 0.12–0.40), RMSE 4.45
Remission	NNS 3.2, PPV 0.64, NPV 0.81, AUC 0.71

In comparison, clinically-adopted pharmacological and psychotherapeutic treatments for MDD have NNTs ranging from 2–25 [18], and other proposed predictors for antidepressants besides bupropion have reported NNTs of 3–5 [9, 10]. Therefore, this NNT of 3.2 has high potential for clinical benefit in identifying individuals mostly likely to respond to bupropion.

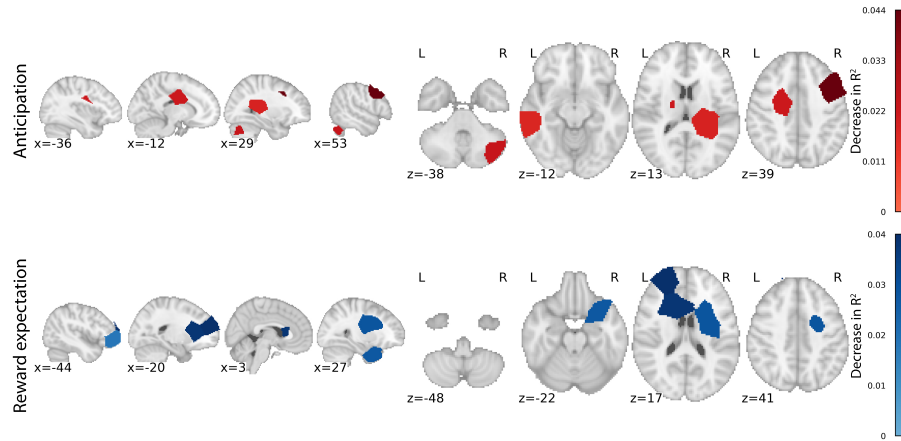
When evaluated on sertraline and placebo-treated subjects from the this dataset, the model demonstrated poor accuracy (negative  $R^2$ ), which is desirable because it indicates the model learned features specific to bupropion response. Additionally, clinical covariates such as demographics, disease duration, and baseline clinical scores were added to the data in another NAS, but this did not increase predictive power. Lastly, less statistically complex models, including multiple linear regression and a support vector machine, performed poorly with negative  $R^2$ , even after hyperparameter optimization with a comparable random search of 800 configurations. This finding suggests that a model with a higher statistical capability such as a neural network was needed to learn the association between the data and treatment outcome.

### 3.2 Learned Neuroimaging Biomarker

Permutation feature importance was measured on the best-performing model configuration to extract a composite neuroimaging biomarker of bupropion response. Specifically, for each feature, the change in  $R^2$  was measured after randomly permuting the feature’s values among the subjects. This was repeated 100 times per feature, and the mean change in  $R^2$  provided an estimate of the importance of each feature in accurate predicting bupropion response. The 10 most important regions for bupropion response prediction are visualized in **Fig. 3** and include the medial frontal cortex, amygdala, cingulate cortex, and striatum. The regions this model has learned to use agree with the regions neurobiologists have identified as key regions in the reward processing neural circuitry [4]. This circuit is the putative target of bupropion and the circuit largely measured by the reward expectancy task in this task-based fMRI study.

## 4 Conclusions

In this work, deep learning and an extensive, unbiased NAS were used to construct predictors of bupropion response from pretreatment task-based fMRI.



**Fig. 3.** The 10 most important ROIs for bupropion response prediction, as measured by permutation feature importance. These included 5 regions in the anticipation contrast map ( $C_{antic}$ , top row) and 5 regions in the reward expectation contrast map ( $C_{re}$ , bottom row). Darker hues indicate greater importance in predicting  $\Delta$ HAMD.

These methods produced a novel, accurate predictive tool to screen for MDD patients likely to respond to bupropion, to estimate the degree of long-term symptomatic improvement after treatment, and to identify patients who will not respond appreciably to the antidepressant. Predictors such as the one presented are an important step to help narrow down the set of candidate antidepressants to be tested for each patient and to address the urgent need for individualized treatment planning in MDD. The results presented also underscore the value of fMRI and in MDD treatment prediction, and future work will target extension to additional treatments.

## References

1. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. *J. M.L. Research* **13**, 281–305 (2012)
2. Calhoun, V.D., et al.: The impact of t1 versus epi spatial normalization templates for fmri data analyses. *Human Brain Mapping* **38**, 5331–5342 (2017). <https://doi.org/10.1002/hbm.23737>
3. Chan, M.Y., et al.: Socioeconomic status moderates age-related differences in the brains functional network organization and anatomy across the adult lifespan. *PNAS* **115**, E5144–E5153 (2018). <https://doi.org/10.1073/pnas.1714021115>
4. Chau, D.T., et al.: The neural circuitry of reward and its relevance to psychiatric disorders. *Curr. Psych. Reports* **6**(5), 391–399 (2004). <https://doi.org/10.1007/s11920-004-0026-8>



5. Craddock, R.C., et al.: A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping* **33**(8), 1914–1928 (2012). <https://doi.org/10.1002/hbm.21333>
6. Dohmatob, E., et al.: Inter-subject registration of functional images: Do we need anatomical images? *Front. Neurosci.* **12**, 64 (2018). <https://doi.org/10.3389/fnins.2018.00064>
7. Dupuy, J.M., et al.: A critical review of pharmacotherapy for major depressive disorder. *Int. J. Neuropsychopharmacology* **14**(10), 1417–1431 (2011). <https://doi.org/10.1017/S1461145711000083>
8. Etkin, A., et al.: Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* **51**(6), 871–882 (2006). <https://doi.org/10.1016/j.neuron.2006.07.029>
9. Etkin, A., et al.: A cognitive-emotional biomarker for predicting remission with antidepressant medications: a report from the ispot-d trial. *Neuropsychopharmacology* : official publication of the American College of Neuropsychopharmacology **40**(6), 1332–1342 (2015). <https://doi.org/10.1038/npp.2014.333>
10. Gordon, E., et al.: Toward an online cognitive and emotional battery to predict treatment remission in depression. *Neuropsychiatric disease and treatment* **11**, 517–531 (2015). <https://doi.org/10.2147/NDT.S75975>
11. Greenberg, P.E., et al.: The economic burden of adults with major depressive disorder in the united states (2005 and 2010). *J. Clin. Psych.* **76**(2), 155–162 (2015). <https://doi.org/10.4088/JCP.14m09298>
12. Kessler, R.C., Bromet, E.J.: The epidemiology of depression across cultures. *Ann. Rev. Public Health* **34**, 119–138 (2013). <https://doi.org/10.1146/annurev-publhealth-031912-114409>
13. Lener, M.S., Iosifescu, D.V.: In pursuit of neuroimaging biomarkers to guide treatment selection in major depressive disorder: a review of the literature. *Annals NYAS* **1344**, 50–65 (2015). <https://doi.org/10.1111/nyas.12759>
14. Li, L., Talwalkar, A.: Random search and reproducibility for neural architecture search (2019), arXiv:1902.07638
15. Patel, K., et al.: Bupropion: a systematic review and meta-analysis of effectiveness as an antidepressant. *Ther. Adv. Psychopharm.* **6**, 99–144 (2016). <https://doi.org/10.1177/2045125316629071>
16. Phillips, M.L., et al.: Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches. *AJP* **172**(2), 124–138 (2015). <https://doi.org/10.1176/appi.ajp.2014.14010076>
17. Pizzagalli, D.A.: Frontocingulate dysfunction in depression: toward biomarkers of treatment response. *Neuropsychopharmacology* **36**(1), 183–206 (2011). <https://doi.org/10.1038/npp.2010.166>
18. Roose, S.P., et al.: Practising evidence-based medicine in an era of high placebo response: number needed to treat reconsidered. *Brit. J. of Psych.* **208**(5), 416–420 (2016). <https://doi.org/10.1192/bjp.bp.115.163261>
19. Rush, A.J., et al.: Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR\*D report. *AJP* **163**(11), 1905–1917 (2006). <https://doi.org/10.1176/ajp.2006.163.11.1905>
20. Schaefer, A., et al.: Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cereb. Cortex* **28**(9), 3095–3114 (2018). <https://doi.org/10.1093/cercor/bhx179>
21. Somerville, L.H., et al.: Interactions between transient and sustained neural signals support the generation and regulation of anxious emotion. *Cereb. Cortex* **23**, 49–60 (2012). <https://doi.org/10.1093/cercor/bhr373>

22. Trivedi, M.H., et al.: Evaluation of outcomes with citalopram for depression using measurement-based care in STAR\*D: implications for clinical practice. *AJP* **163**(1), 28–40 (2006). <https://doi.org/10.1176/appi.ajp.163.1.28>
23. Trivedi, M.H., et al.: Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *J. Psych. Res.* **78**, 11–23 (2016). <https://doi.org/10.1016/j.jpsychires.2016.03.001>
24. Varma, S., Simon, R.: Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* **7**, 91–99 (2006). <https://doi.org/10.1186/1471-2105-7-91>