Artificial intelligence-assisted analysis of CT abnormalities during COVID-19 recovery

Rebuttal Letter 2

Department of Radiology, Medical University of Innsbruck

2025-03-15

*Dear Anna, Dear Gerlig, I’ve already adjusted the nummeration of Supplementary figures and Supplementary Tables in the second rebuttal letter.*

# Reviewer 1

## Concern 1

### Reviewer, first round

The paper lacks details about hyper-parameters for machine learning models. For example, what was the C and class weight value for the SVM model?

### Response, first round

Thanks for this point. We agree that the information on hyper-parameters in the initial submission was too specific and hard to understand for many clinical readers. In the revised manuscript, we briefly describe the hyper-parameters in Supplementary Methods/Multi-parameter modeling of lung function. These descriptions were also added as footnotes of **Supplementary Table S7**.

### Reviewer, second round

Providing only good parameters is not sufficient. There is a need to demonstrate the effect of different parameters so that it can be concluded that your selected value is good.

### Response, second round

Thanks for this point. In the revised manuscript, we provide the complete tuning grids of hyper-parameters with the corresponding values of the cost functions (Youden’s J for the classification models and mean absolute error [MAE] for the regression models) as **Supplementary Table S6**. Additionally, we show visualizations of the J and MAE as a function of combinations of the hyper-parameters for classification and regression models of DLCO < 80% and DLCO in **Supplementary Figures S9** and **S10**. Plots of the tuning process for FVC < 80%, FEV1 < 80%, FVC, and FEV1 are presented in **Reviewer Figures 1** - **4** attached to the rebuttal letter.

## Concern 2

### Reviewer, first round

Have you considered the GridSearchCV for model hyper-parameters or set it manually?

### Response, first round

Thanks for this question. The hyper-parameters were found by searching grids of values with the maximum Youden’s J statistic for the classification models, and minimum mean absolute error (MAE) for the regression models in 10-repeats 10-fold patient-blocked cross-validation as the selection criteria. This tuning procedure was executed by calling train() function from caret package (1), which, together with our open-source package [*caretExtra*](https://github.com/PiotrTymoszuk/caretExtra), belongs to our standard toolbox for explainable machine learning. This information is provided in Supplementary Methods/Multi-parameter modeling of lung function.

### Reviewer, second round

Have you specifically considered the GridSeachCV (<https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>) to find the best parameters? If yes, then which range of values (parameters grid) have you provided to find the best values?

### Response, second round

Thanks for the suggestion - no, we have not considered *GridSearchCV* for tuning of hyper-parameters in the current paper, but we may consider it for future projects. Please note, that our approach to tuning via repeated cross-validation, which uses a well established and highly cited R package *caret*, does not differ from other similar works in the field of respiratory medicine (2–4) (references included in **Supplementary Methods**). While, we appreciate your suggestion for *GridSearchCV*, we do not believe that the tuning process based on patient-blocked repeated cross-validation would yield substantially different or better machine learning models if we used *GridSearchCV* instead of *caret*.

## Concern 3

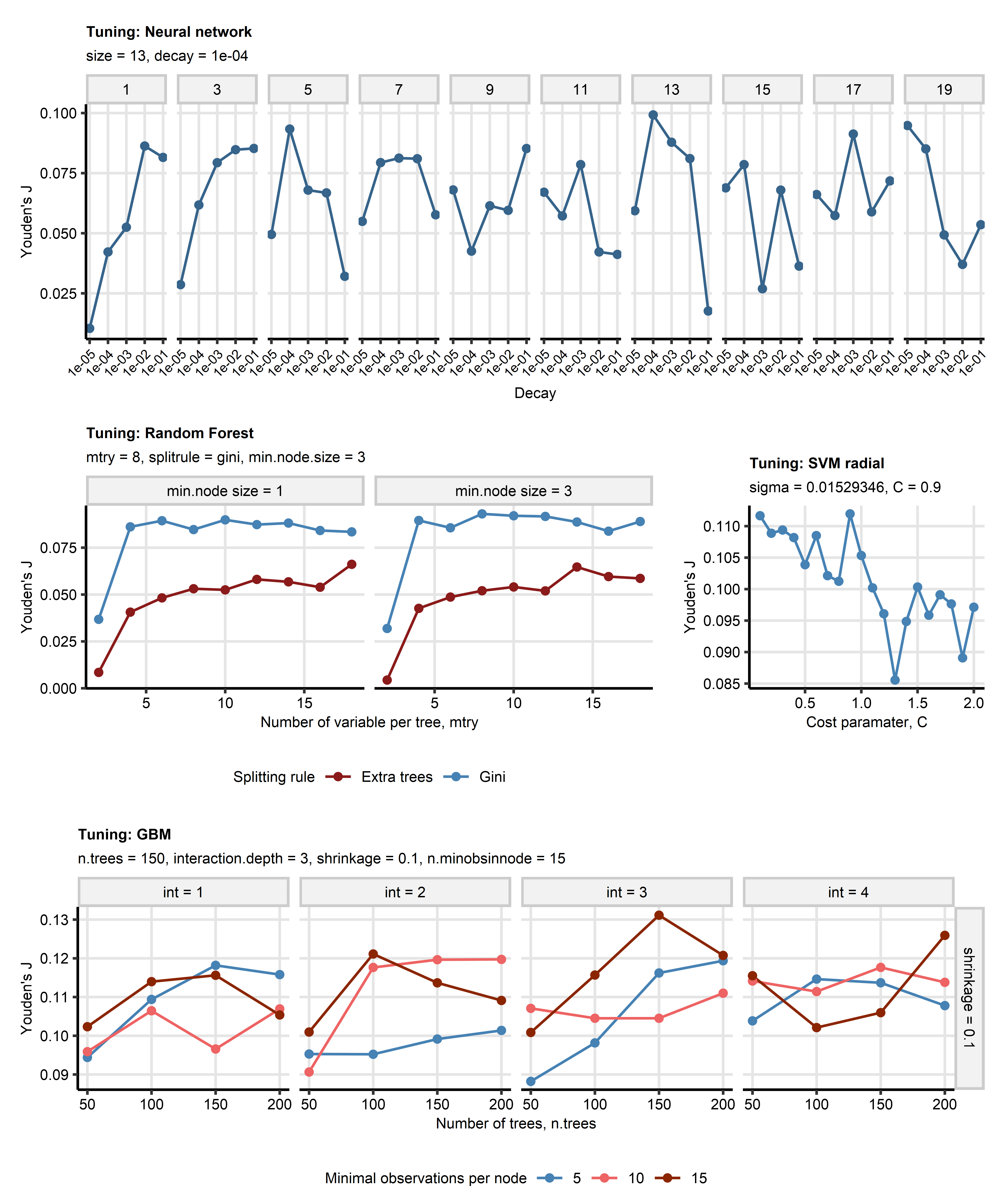
### Reviewer, second round

Furthermore, the details of practical implication and also conclusion are not well written. Please thoroughly improve the paper in terms of practicality and write-up.

### Response, third round

The Conclusion section has been re-written to improve structure, readability, and impact.

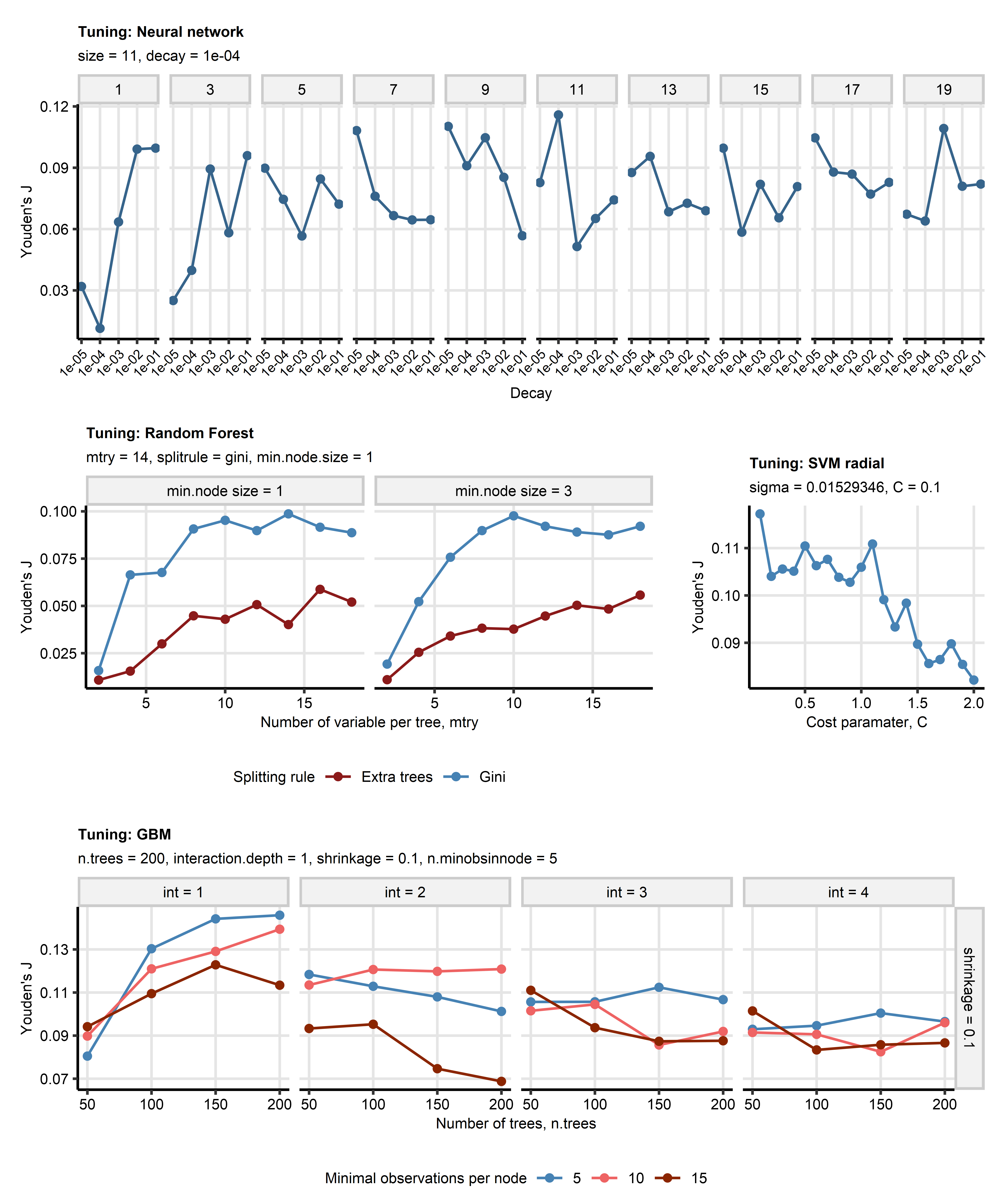
# Review Figures



**Review Figure 1. Choice of hyper-parameters for machine learning classification models of reduced forced vital capacity.**

*Classification machine learning models of abnormalities of lung function testing (diffusion capacity of carbon monoxide [DLCO] < 80%, forced vital capacity [FVC] < 80%, forced expiratory volume in one second [FEV1] < 80% of the patient’s reference) were constructed with the neural network, random forest, support vector machines with radial kernel (SVM), and gradient boosted machines (GBM) algorithms. The explanatory variables included demographic information, information on course of acute COVID-19, longitudinally recorded persistent symptoms, and longitudinally recorded readouts of structural lung damage via computed tomography of the chest.* *Choice of the algorithms’ hyper-parameters (tuning) was accomplished by maximizing Youden’s J statistic () for out-of-fold predictions in patient-blocked 10-repeats 10-fold cross-validation.* *Trajectories of J values for combinations of the hyper-parameters of the models of FVC < 80% are presented in line plots. The optimal combinations of the hyper-parameters used for training of the final models are displayed in the plot captions.*

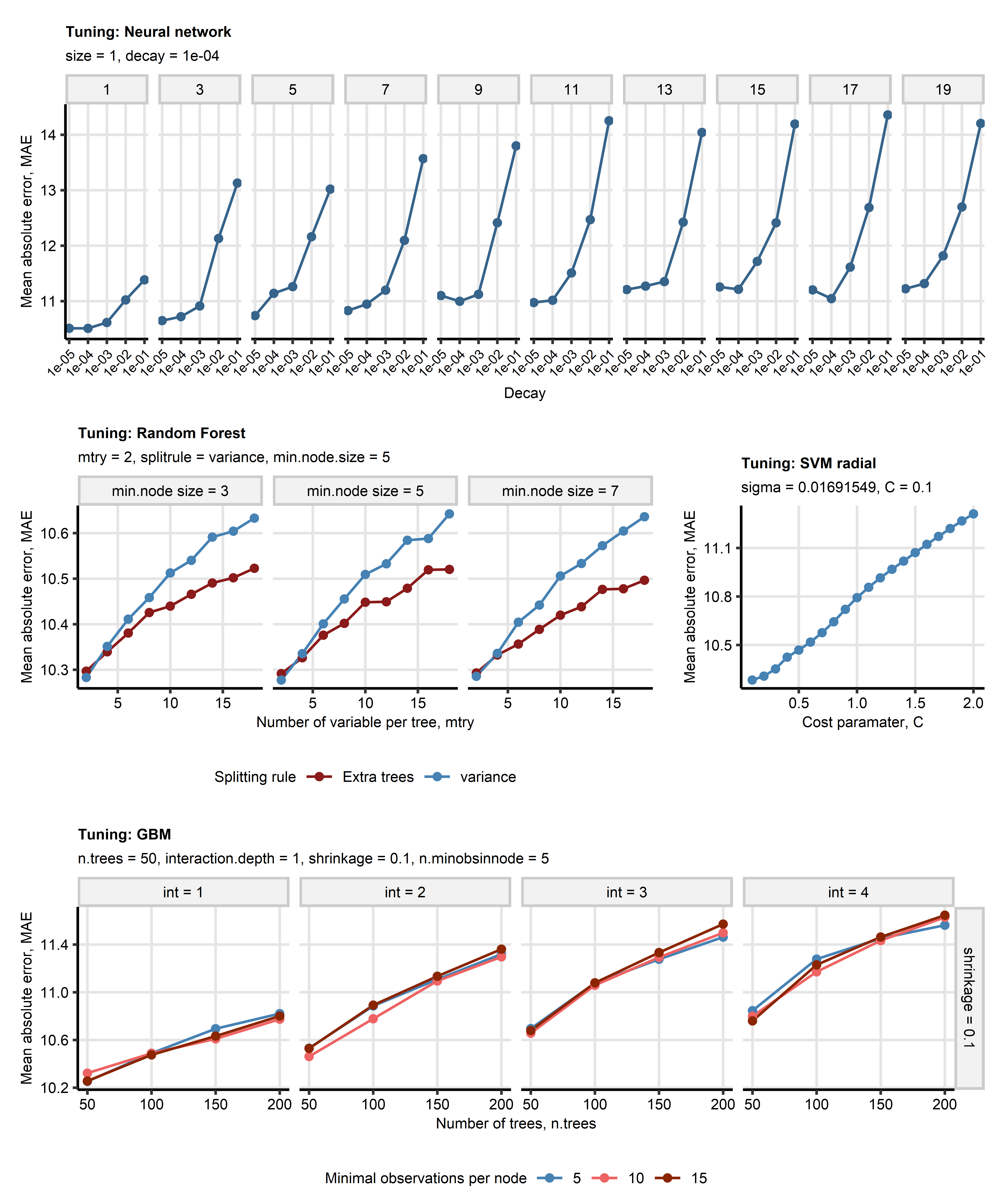
*size: number of neurons in the hidden layer of the neuronal model; decay: decay of weights in the neuronal network; mtry: number of randomly chosen explanatory variables for construction of decision trees; splitrule: rule used for splitting of explanatory variable at branches of decision trees; min.node.size: minimal size of the decision tree node; sigma: sigma parameter that defines width of the radial kernel; C: cost/penalty applied to misclassified observations; n.tree: number of decision trees; interaction.depth/int: number of splits in the decision tree; shrinkage: regularization penalty applied to predictions by single decision trees; n.minobsinnode: minimal size of the decision tree node.*



**Review Figure 2. Choice of hyper-parameters for machine learning classification models of reduced forced expiratory volume in one second.**

*Classification machine learning models of abnormalities of lung function testing (diffusion capacity of carbon monoxide [DLCO] < 80%, forced vital capacity [FVC] < 80%, forced expiratory volume in one second [FEV1] < 80% of the patient’s reference) were constructed with the neural network, random forest, support vector machines with radial kernel (SVM), and gradient boosted machines (GBM) algorithms. The explanatory variables included demographic information, information on course of acute COVID-19, longitudinally recorded persistent symptoms, and longitudinally recorded readouts of structural lung damage via computed tomography of the chest.* *Choice of the algorithms’ hyper-parameters (tuning) was accomplished by maximizing Youden’s J statistic () for out-of-fold predictions in patient-blocked 10-repeats 10-fold cross-validation.* *Trajectories of J values for combinations of the hyper-parameters of the models of FEV1 < 80% are presented in line plots. The optimal combinations of the hyper-parameters used for training of the final models are displayed in the plot captions.*

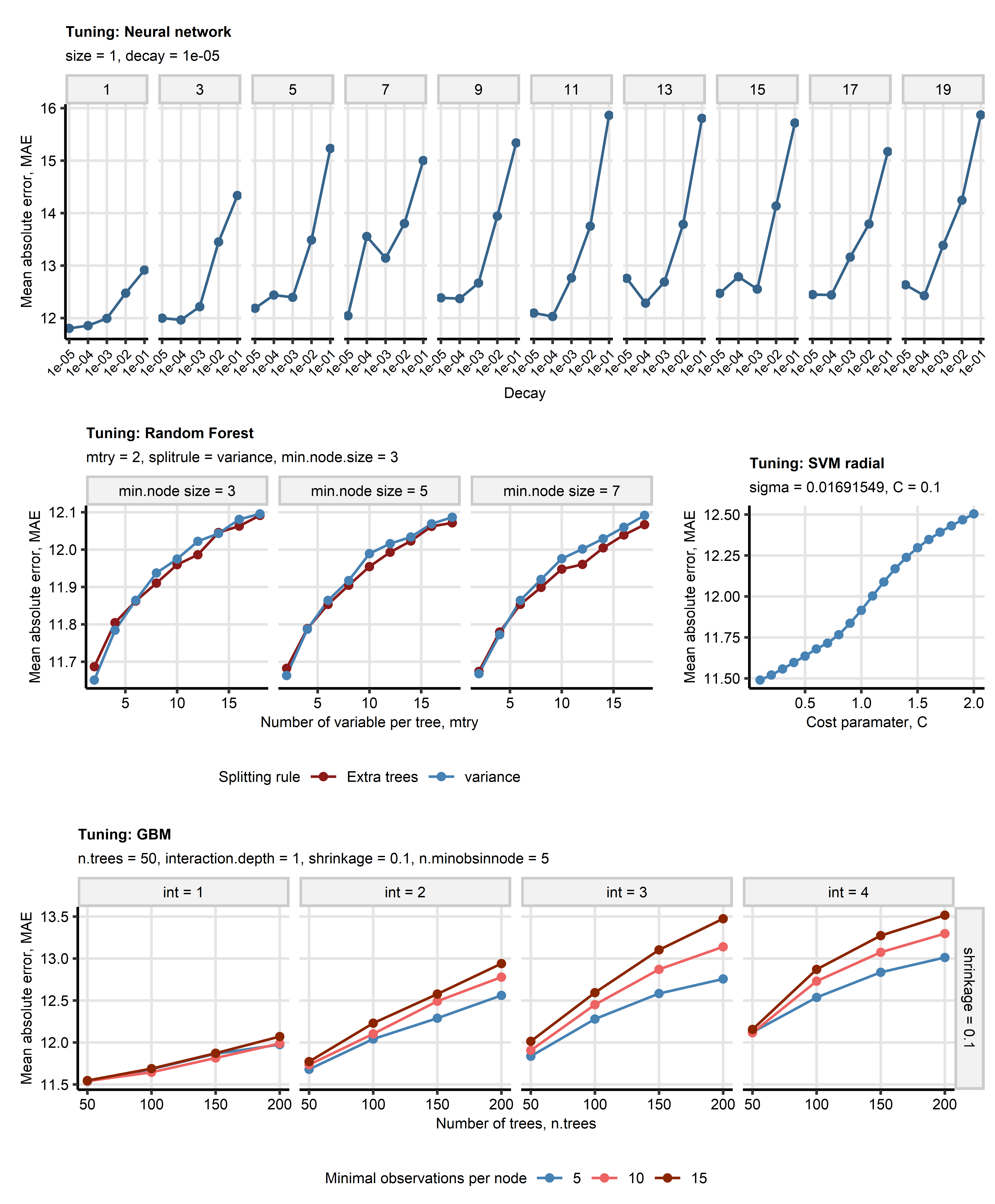
*size: number of neurons in the hidden layer of the neuronal model; decay: decay of weights in the neuronal network; mtry: number of randomly chosen explanatory variables for construction of decision trees; splitrule: rule used for splitting of explanatory variable at branches of decision trees; min.node.size: minimal size of the decision tree node; sigma: sigma parameter that defines width of the radial kernel; C: cost/penalty applied to misclassified observations; n.tree: number of decision trees; interaction.depth/int: number of splits in the decision tree; shrinkage: regularization penalty applied to predictions by single decision trees; n.minobsinnode: minimal size of the decision tree node.*



**Review Figure 3. Choice of hyper-parameters for machine learning regression models of forced vital capacity.**

*Regression machine learning models of parameters of lung function testing (diffusion capacity of carbon monoxide [DLCO], forced vital capacity [FVC], forced expiratory volume in one second [FEV1]) were constructed with the neural network, random forest, support vector machines with radial kernel (SVM), and gradient boosted machines (GBM) algorithms. The explanatory variables included demographic information, information on course of acute COVID-19, longitudinally recorded persistent symptoms, and longitudinally recorded readouts of structural lung damage via computed tomography of the chest.* *Choice of the algorithms’ hyper-parameters (tuning) was accomplished by minimizing mean absolute errors (MAE) for out-of-fold predictions in patient-blocked 10-repeats 10-fold cross-validation.* *Trajectories of MAE values for combinations of the hyper-parameters of the models of FVC are presented in line plots. The optimal combinations of the hyper-parameters used for training of the final models are displayed in the plot captions.*

*size: number of neurons in the hidden layer of the neuronal model; decay: decay of weights in the neuronal network; mtry: number of randomly chosen explanatory variables for construction of decision trees; splitrule: rule used for splitting of explanatory variable at branches of decision trees; min.node.size: minimal size of the decision tree node; sigma: sigma parameter that defines width of the radial kernel; C: cost/penalty applied to misclassified observations; n.tree: number of decision trees; interaction.depth/int: number of splits in the decision tree; shrinkage: regularization penalty applied to predictions by single decision trees; n.minobsinnode: minimal size of the decision tree node.*



**Figure 4. Choice of hyper-parameters for machine learning regression models of forced expiratory volume in one second.**

*Regression machine learning models of parameters of lung function testing (diffusion capacity of carbon monoxide [DLCO], forced vital capacity [FVC], forced expiratory volume in one second [FEV1]) were constructed with the neural network, random forest, support vector machines with radial kernel (SVM), and gradient boosted machines (GBM) algorithms. The explanatory variables included demographic information, information on course of acute COVID-19, longitudinally recorded persistent symptoms, and longitudinally recorded readouts of structural lung damage via computed tomography of the chest.* *Choice of the algorithms’ hyper-parameters (tuning) was accomplished by minimizing mean absolute errors (MAE) for out-of-fold predictions in patient-blocked 10-repeats 10-fold cross-validation.* *Trajectories of MAE values for combinations of the hyper-parameters of the models of FEV1 are presented in line plots. The optimal combinations of the hyper-parameters used for training of the final models are displayed in the plot captions.*

*size: number of neurons in the hidden layer of the neuronal model; decay: decay of weights in the neuronal network; mtry: number of randomly chosen explanatory variables for construction of decision trees; splitrule: rule used for splitting of explanatory variable at branches of decision trees; min.node.size: minimal size of the decision tree node; sigma: sigma parameter that defines width of the radial kernel; C: cost/penalty applied to misclassified observations; n.tree: number of decision trees; interaction.depth/int: number of splits in the decision tree; shrinkage: regularization penalty applied to predictions by single decision trees; n.minobsinnode: minimal size of the decision tree node.*

# References

1. Kuhn M. Building predictive models in R using the caret package. *Journal of Statistical Software* (2008) 28:1–26. doi: [10.18637/jss.v028.i05](https://doi.org/10.18637/jss.v028.i05)

2. Sharifi H, Lai YK, Guo H, Hoppenfeld M, Guenther ZD, Johnston L, Brondstetter T, Chhatwani L, Nicolls MR, Hsu JL. Machine Learning Algorithms to Differentiate Among Pulmonary Complications After Hematopoietic Cell Transplant. *Chest* (2020) 158:1090. doi: [10.1016/J.CHEST.2020.02.076](https://doi.org/10.1016/J.CHEST.2020.02.076)

3. Murdaca G, Caprioli S, Tonacci A, Billeci L, Greco M, Negrini S, Cittadini G, Zentilin P, Spagnolo EV, Gangemi S. A Machine Learning Application to Predict Early Lung Involvement in Scleroderma: A Feasibility Evaluation. *Diagnostics* (2021) 11:1880. doi: [10.3390/DIAGNOSTICS11101880](https://doi.org/10.3390/DIAGNOSTICS11101880)

4. Nikolaou V, Massaro S, Garn W, Fakhimi M, Stergioulas L, Price DB. Fast decliner phenotype of chronic obstructive pulmonary disease (COPD): applying machine learning for predicting lung function loss. *BMJ Open Respiratory Research* (2021) 8: doi: [10.1136/BMJRESP-2021-000980](https://doi.org/10.1136/BMJRESP-2021-000980)