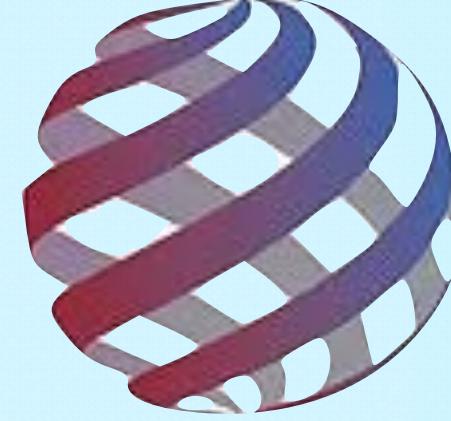




MONASH  
University



CAR

Centre for Air pollution, energy and health Research

# DEEPER workshop 2022

## Graphical User Interface of deeper

Liam(Yanning) Liu

Climate, Air Quality Research Unit

Monash University

8th September 2022

# Typing the codes or clicking the buttons?



VS

RUN



For those who do not like typing the codes,  
we have got you covered!

# General steps

## 1. Collecting

- Prepare the data in one .csv file for training

## 2. Training

- Load the data file
- Choose the hyper-parameters
- Define the variable types
- Choose model depth
- Construct the model
- Choose the test size
- Train the model
- Visualize the results and save the model

## 3. Predicting

- Load the trained model and new data
- Predict and save the results



# Collecting

Prepare the .CSV files contains all your variables and predictive target

Variable name



no2_annual	year	lon_metres	lat_metres	pos_elevation	pos_distocean	pop_dens_10000	em_rds_dist	em_majrd_invdist	em_minrd_invdist	em_majrd_length_200
4.177731799	2018	1723117.173	-3882593.071	-0.544294473	-0.727166305	-0.85584571	-0.18758751	-0.359988814	0.075111775	-0.484918123
2.126906714	2018	1720246.656	-3885162.733	-0.435240007	-0.689640936	-0.835740569	-0.186216924	-0.360363679	-0.176187475	-0.484918123
7.382675624	2018	1723137.19	-3812193.262	-0.328907473	-0.089544131	-0.419274515	-0.352977145	-0.265585772	0.077285297	0.21700719
7.449784085	2018	1748316.873	-3836652.973	-0.360790207	-0.563637169	1.133860457	-0.492881214	1.617709623	-0.302817454	0.241042137
4.667770885	2018	1746639.224	-3838431.334	-0.460185782	-0.568335574	0.831609122	0.772619258	-0.425732554	-0.249150839	-0.484918123
4.142557015	2018	1717309.145	-3843952.778	0.11789035	-0.372473368	-0.702683265	0.479277258	-0.419236284	-0.258922561	-0.484918123
9.765305923	2018	1721123.38	-3846161.261	-0.091687077	-0.431976075	-0.731729068	-0.494935391	1.987768919	0.195466588	1.486380856
3.584640508	2018	1758606.621	-3811543.029	0.241110514	-0.568013976	-0.236868647	0.110143804	-0.402277818	-0.226950894	-0.484918123
6.29356366	2018	1776867.523	-3790949.588	-0.579438671	-0.663423538	-0.996569046	0.378067159	-0.415995059	0.153178584	-0.484918123
1.709336368	2018	1782499.854	-3779185.954	-0.488179432	-0.637105814	-0.886778319	0.270928627	-0.411644339	-0.222676644	-0.484918123
11.06242354	2018	1739308.241	-3820156.315	-0.33307101	-0.346835293	0.854695976	-0.495648744	2.149360809	-0.318004011	0.967605914
4.981112448	2018	1751966.426	-3830567.762	-0.395348058	-0.571560596	2.113248062	-0.184767007	-0.360756765	-0.214878949	-0.484918123
3.785114346	2018	1741213.682	-3807586.514	-0.122718269	-0.311593493	-0.066011492	0.124535333	-0.403310722	-0.241370968	-0.484918123
6.748023393	2018	1744247.517	-3811313.369	0.131546677	-0.369547291	0.429519754	-0.476553814	0.48788452	-0.247031435	0.512975175

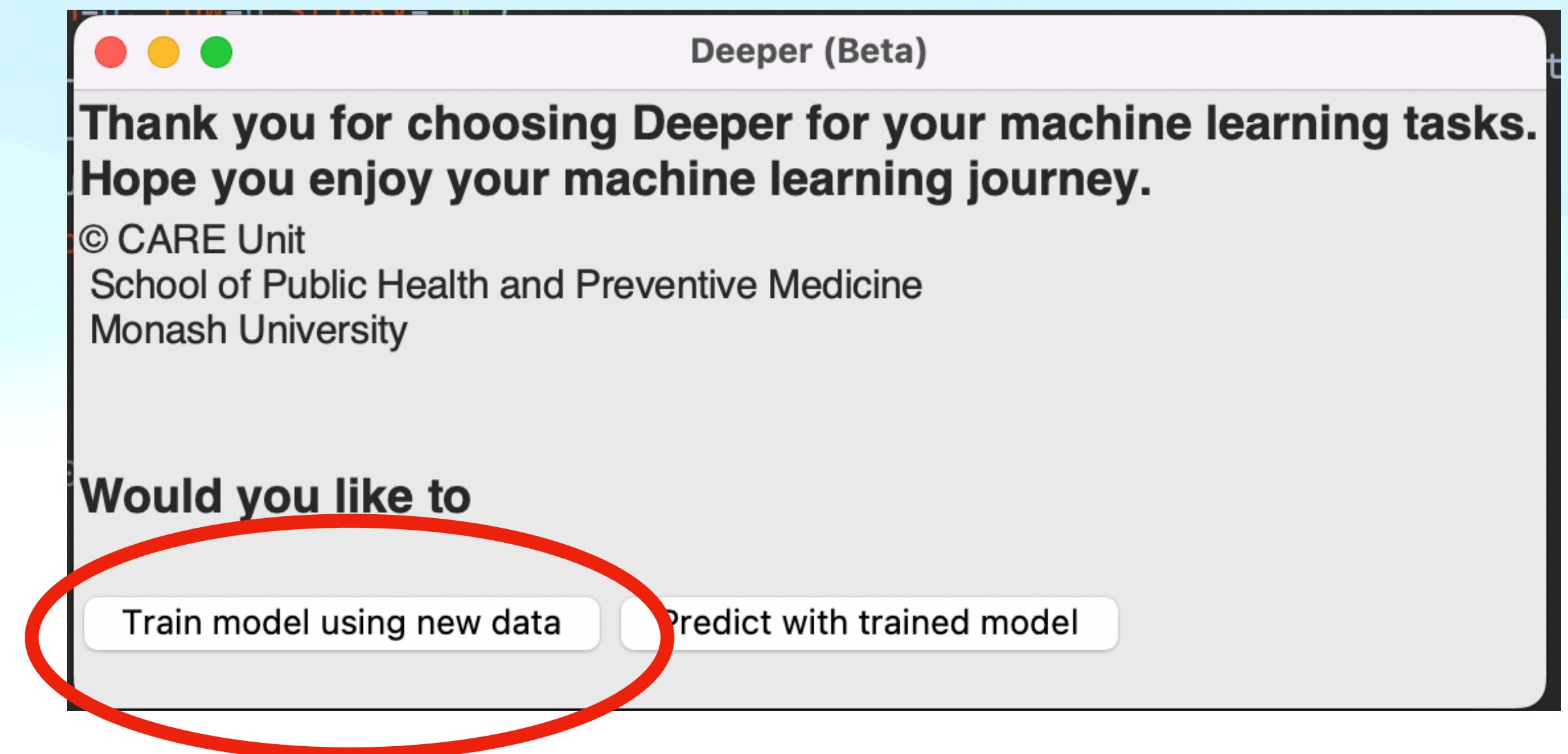
Predictive target

Variables

# Training

Click the “Train model using new data” button for training

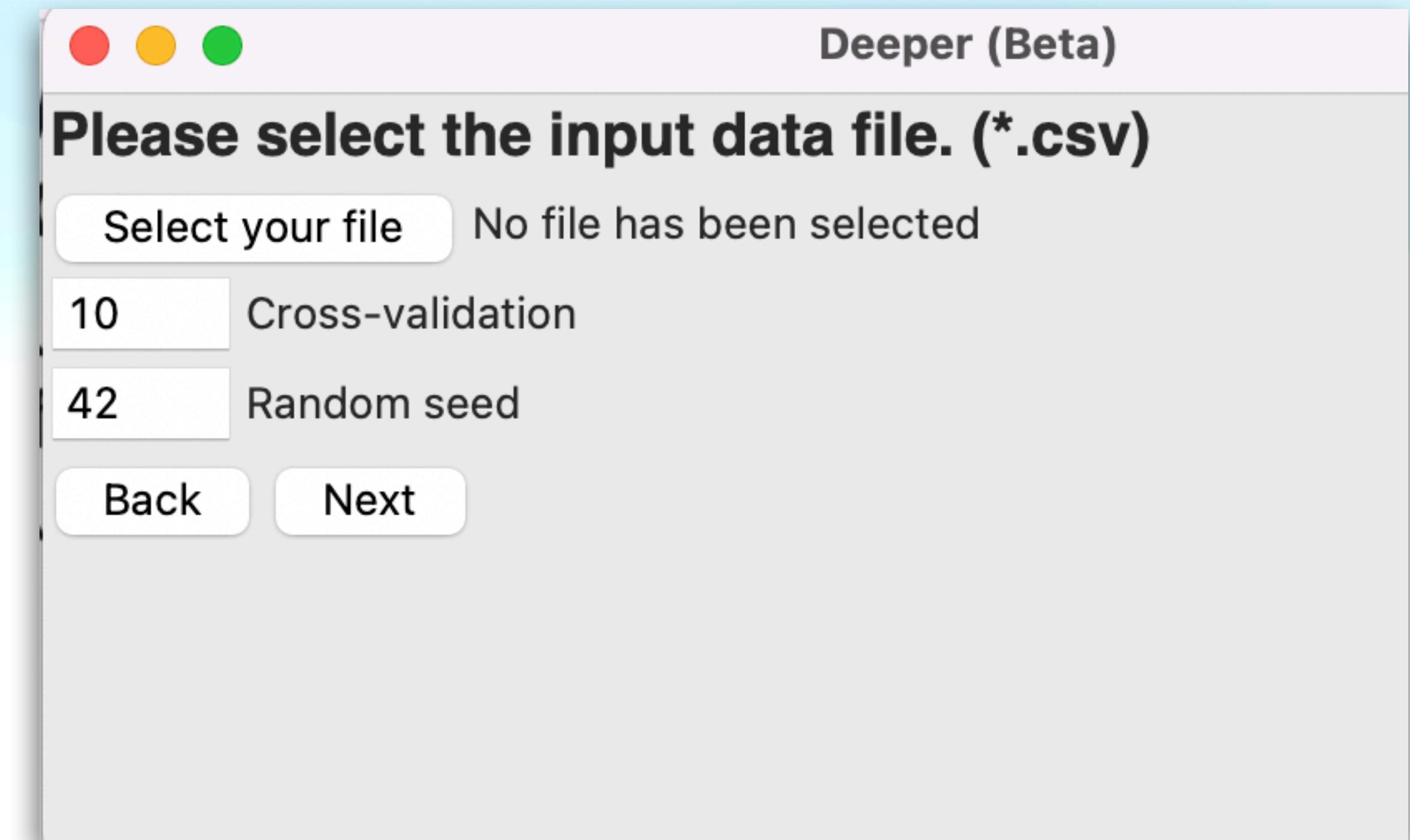
## The DEEPER user interface



# Training

Load the data file & choose the hyper-parameters

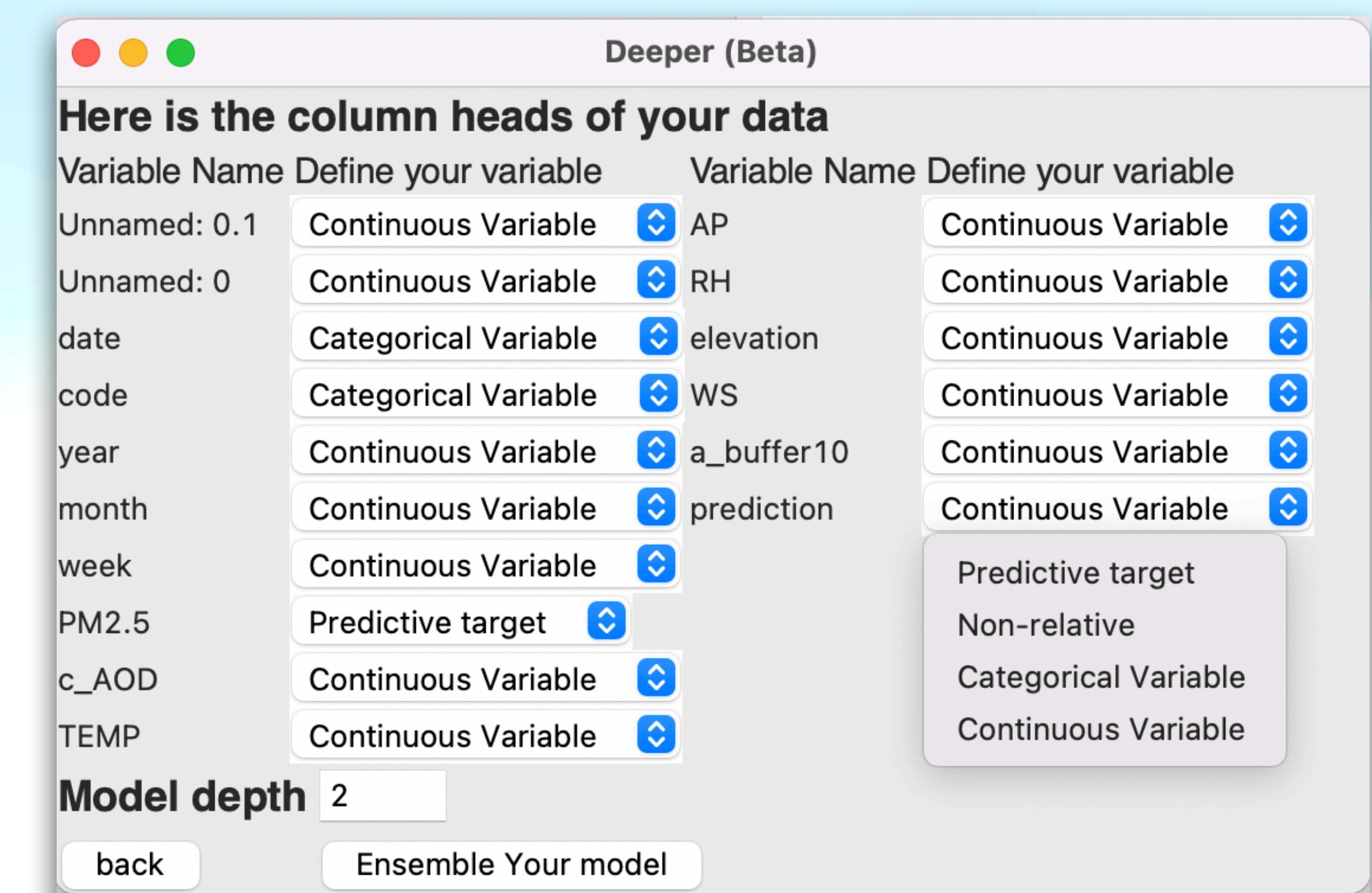
- The .csv file containing all the variables and predictive targeting should be selected.
- Cross-validation determines how many sub-set the training set is divided into to prevent overfitting
- A fixed random seed makes sure the outcome consistent



# Training

## Define the variable types & choose the model depth

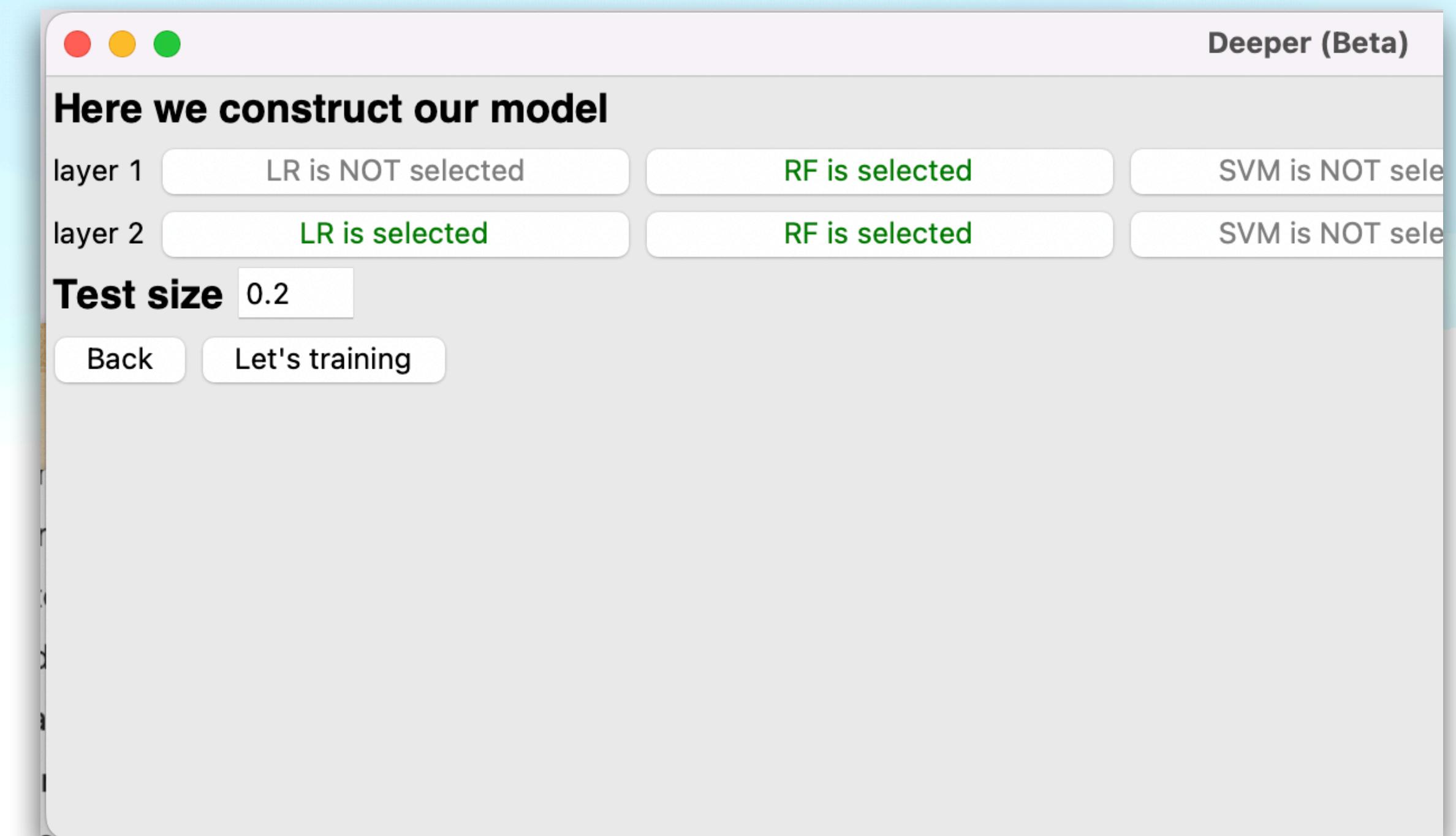
- The column heads of the variables in the loaded data will be listed
- For each variable, a data type must be defined
- The numerical data will be defaultly taken as a continuous variable and other data will be default taken as a categorical variable.
- You can manually change the data type into one of the four types: Non-relative, categorical variable, continuous variable
- ONE and ONLY ONE predictive target must be defined
- Model depth is the number of layers you want to ensemble



# Training

Construct the model and choose the test size

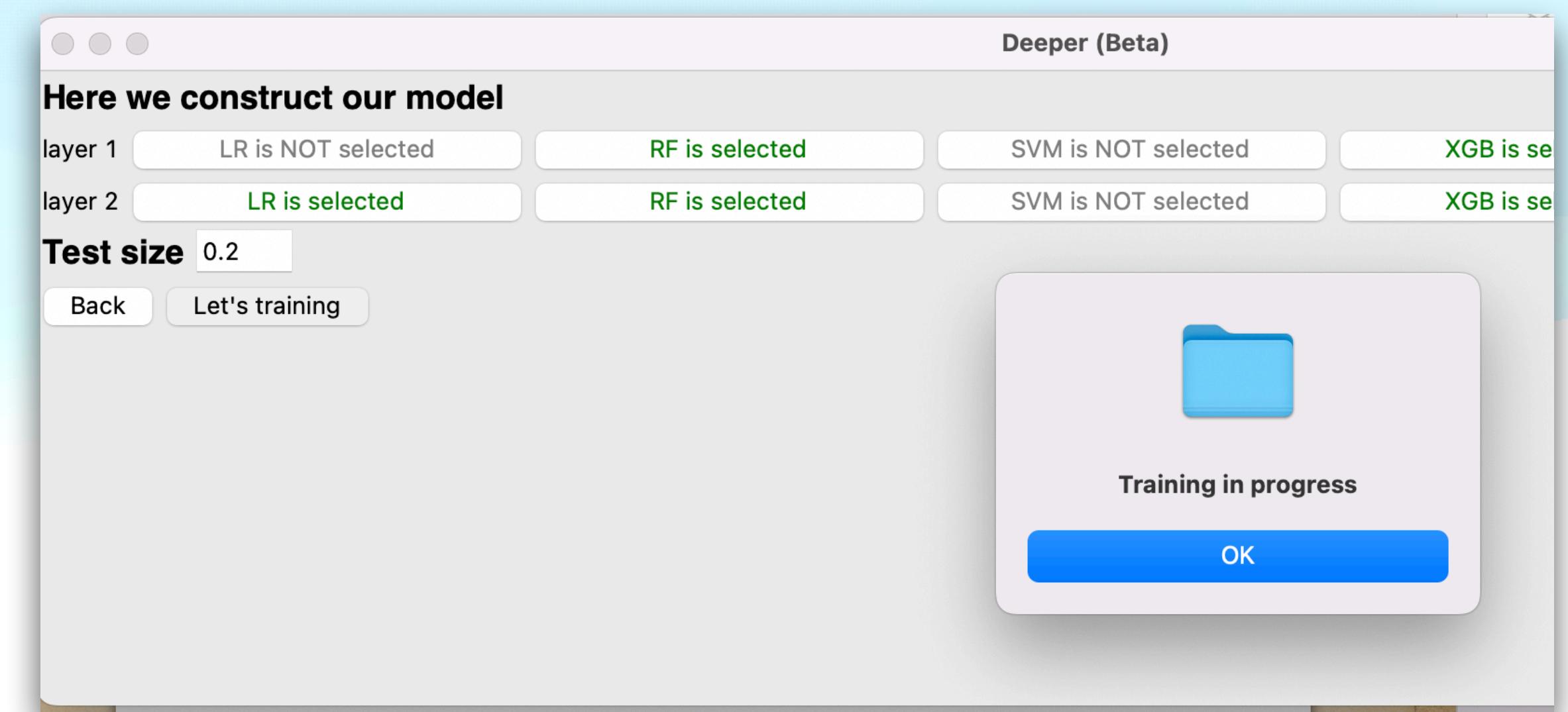
- Simply click the model you want to have in each layer.
- Test size is the portion of the training data you do not train so that it can be used to evaluate the model's performance.



# Training

## Train the model

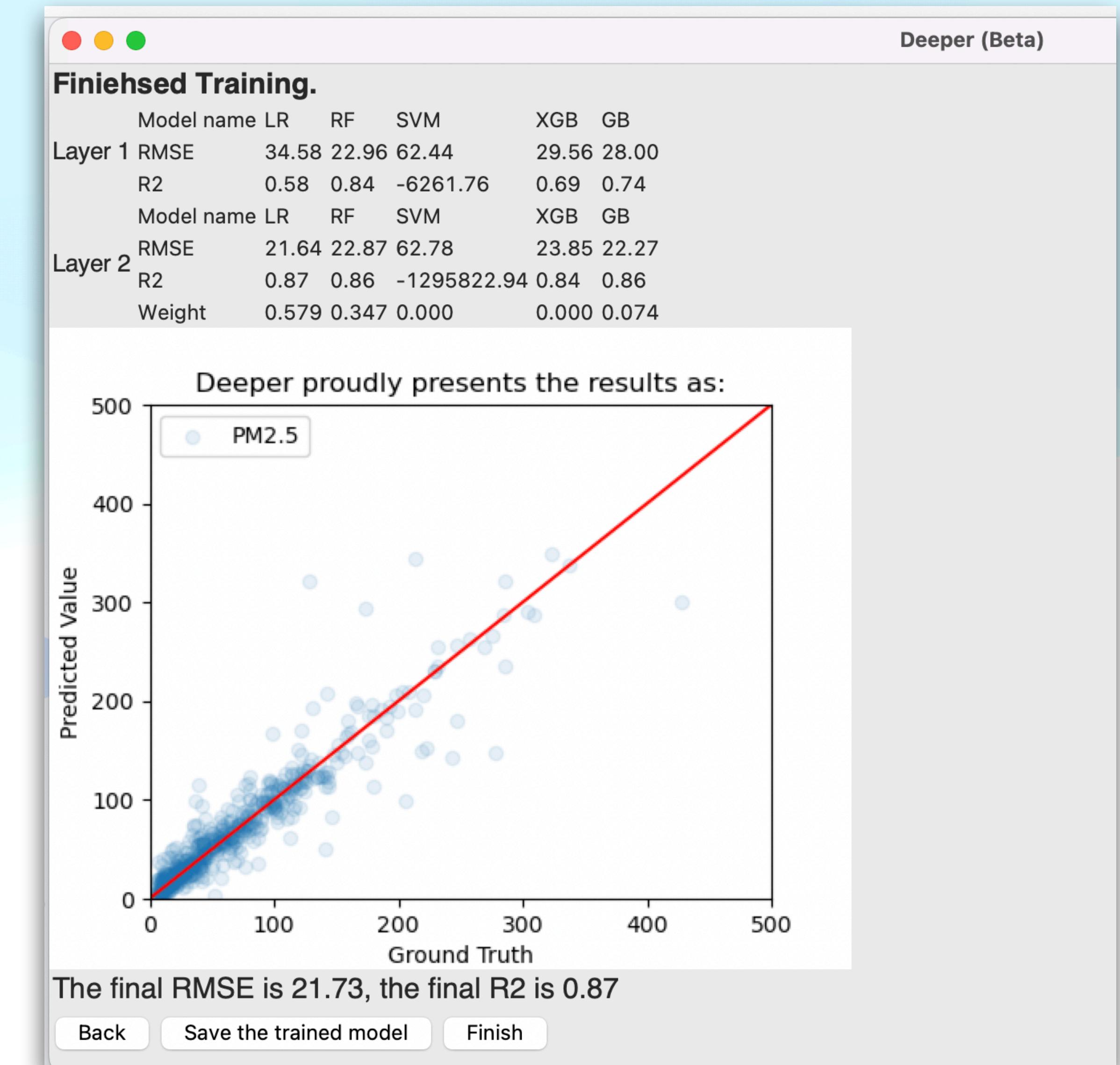
- Click “Let’s training” to start the training of the deeper model
- Please note this step is computation demanding, and the processing time may vary from few minutes to few hours depends on your data and model selection



# Training

Visualise the results and save the model

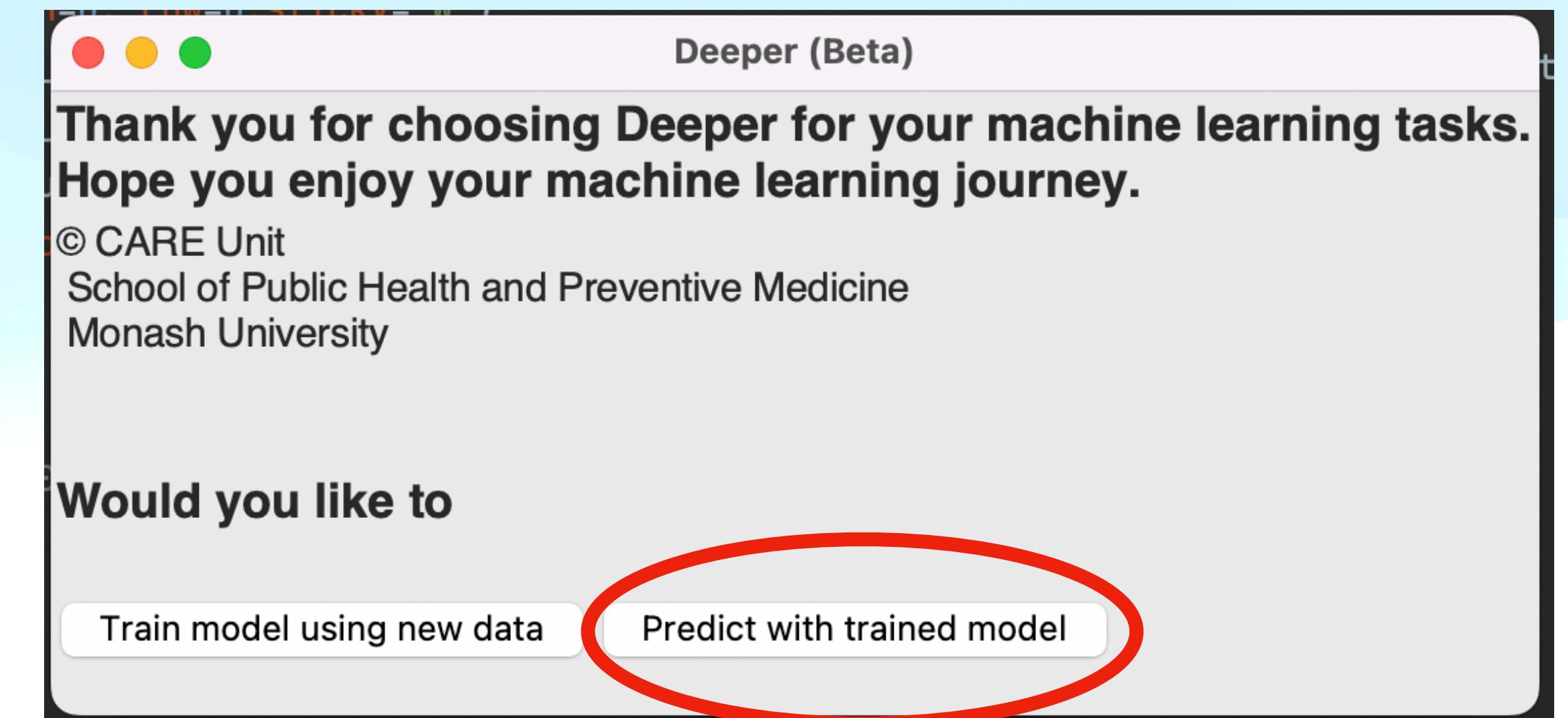
- The RMSE and R2 of each model are present for evaluation of the model performance
- A weight is shown in the final layer to describe the contribution of each model
- A scatter is plotted to visualise the final deeper model performance using the test set(from your test size)
- The final RMSE and R2 of the deeper are also present
- Now we can click the “save the model” button to save the trained model into a .pkl file



# Predicting

## The DEEPER user interface

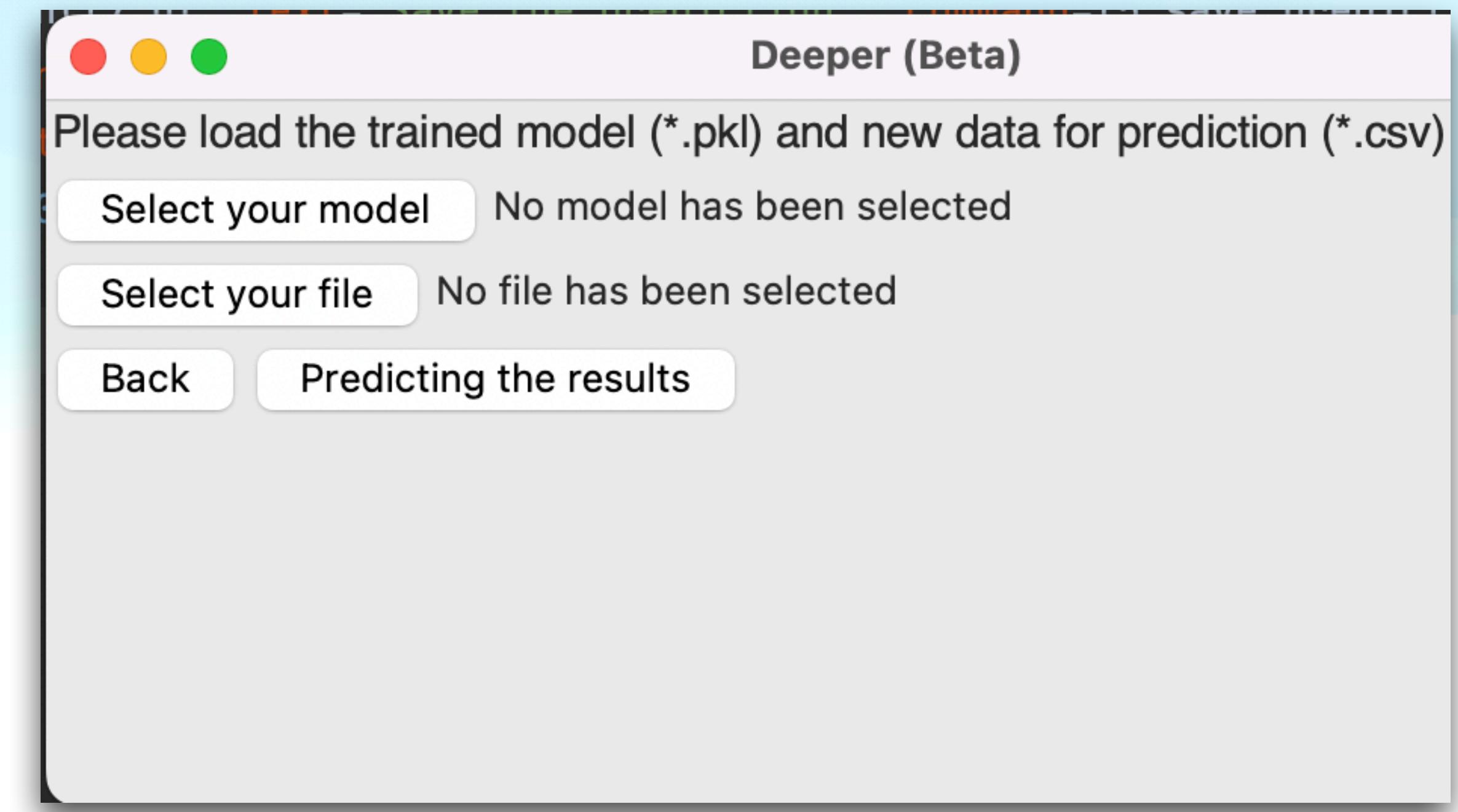
Click the “Predict with trained model” button for prediction



# Predicting

Load the trained model and new data

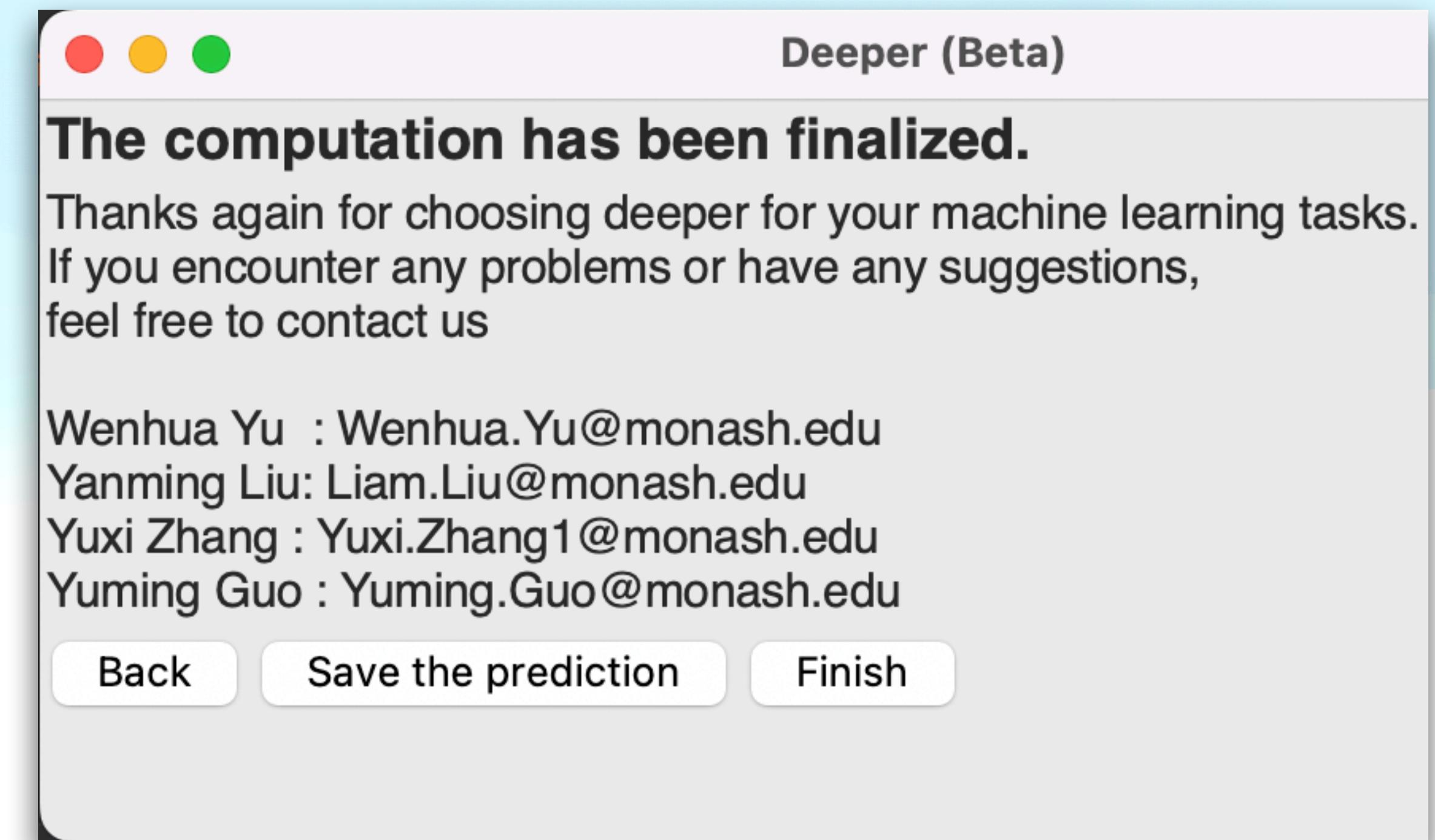
- Selected your trained model(.pkl file)
- Selected the new data(.CSV) for prediction
- It should be noted that the new data selected must have the same variable names when training the model
- Click “Predicting the results” button to run the model



# Predicting

## Predict and save the results

- Wait till the computation is finalised
- Save the prediction by clicking the button
- If you encounter any problems or have any suggestions, please contact us through the emails provided.
- The prediction results will be shown in the CSV file as an additional column.



# Limitation

Despite the ease of clicking,  
The GUI has its limitations:

- Hyper-parameter of each individual model
- Limited computer environment
- Parallel computing
- Fully control and modification



Thank you for listening

