

Tutorial for JADBio Regression analysis – Parkinson's disease Case study

INTRODUCTION	2
ACCESSING JADBio	3
Try for free (Sign up).....	3
To Log In (Sign in)	3
VIEWING DATA AND ANALYSES IN JADBio	4
JADBio header	4
Menu	4
Dashboard.....	5
Projects.....	8
Datasets	10
Analyses	13
Applied Models.....	14
WORKING IN JADBio	15
Creating a Project	15
Uploading a Dataset.....	16
Transformations.....	20
Background	23
Analysis.....	24
Analysis Results.....	26
Best Performing Model.....	29
Performance Overview	30
Feature Selection	31
Analysis Visualization	34
Best interpretable model.....	35
Apply Model.....	37

Introduction

JADBio is a platform designed specifically to extract value and insight from multi-omics data sets, typically thousands of measurements in a small number of samples. JADBio's uniquely tuned Automated Machine Learning (AutoML) is guided by Artificial Intelligence (AI) to provide accurate and efficient predictive models for Classification, Regression and Survival (Time to Event) analysis.

JADBio is very straightforward to use:

1. Upload your data
2. Transform your data (optional)
3. Analyze your data
4. View your results (optional)
5. Share your results
6. Test your data (optional)

Accessing JADBio

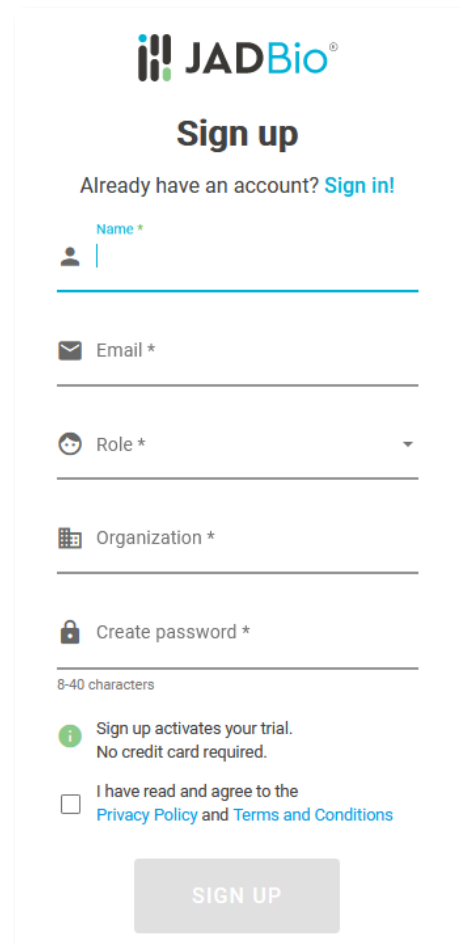
- In your browser of choice, navigate to JADBIO.com.
- Click on the **LOG IN** button in top right corner if you already have an account or the **TRY FOR FREE** button if you don't.

Try for free (Sign up)

- Complete the Register Dialog.
- Click on **SIGN UP**.

After you register, you will receive an e-mail from no-reply@gnosisda.gr with the subject, JADBio Email Verification.

- To verify your email and activate your account, please open the e-mail, and click on the provided link.



The screenshot shows the JADBio registration dialog box. At the top is the JADBio logo and the text "Sign up". Below this is a link: "Already have an account? [Sign in!](#)". The form contains several fields: "Name *" with a person icon, "Email *" with an envelope icon, "Role *" with a person icon and a dropdown arrow, "Organization *" with a calendar icon, and "Create password *" with a lock icon. Below the password field is the text "8-40 characters". There is an information icon (i) next to the text "Sign up activates your trial. No credit card required." Below that is a checkbox with the text "I have read and agree to the [Privacy Policy](#) and [Terms and Conditions](#)". At the bottom of the form is a grey button labeled "SIGN UP".

Figure 1 Registration dialog

To Log In (Sign in)

- Type in your username and password and click on **SIGN IN**.

Viewing data and analyses in JADBio

JADBio header

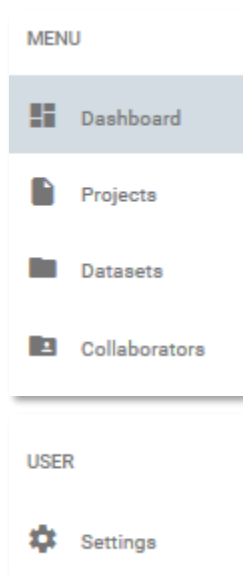
The **JADBio header** contains:

- Notifications under the bell icon on the right.
- Sign on and off function under your username.



Figure 2 JADBio header

Menu



MENU sidebar provides navigation to the **Dashboard**, to your **Projects**, to your **Datasets** and to your **Collaborators** as well as to the **USER Settings**.

Figure 3 MENU and USER sidebars

Dashboard

The **Dashboard** provides an overview of your Account. This includes:

1. **Projects:** A graphic representation of shared and exclusive projects. To start, you will have one shared project.

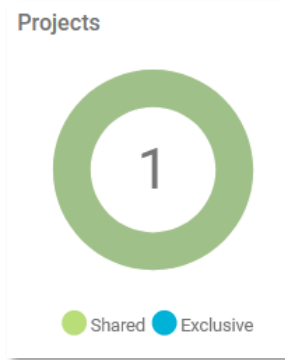


Figure 4 Projects

2. **Infrastructure Statistics:** A summary of your current Storage and Compute. Standard subscriptions include 2 or 6 cores.

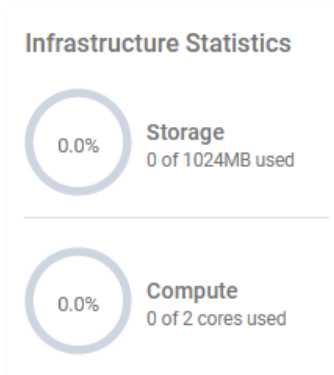


Figure 5 Infrastructure Statistics – the number of available cores is dependent on your subscription

3. **Datasets:** The total number of your Datasets.

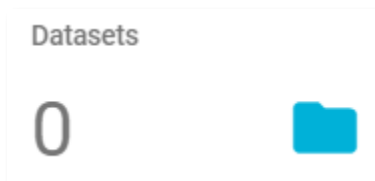


Figure 6 Datasets

4. **Analyses:** The total number of your Analyses.



Figure 7 Analyses

5. **Recently Shared:** A list of your current Collaborators. When you share a project with another subscriber, they become a collaborator, and your project is defined as shared.

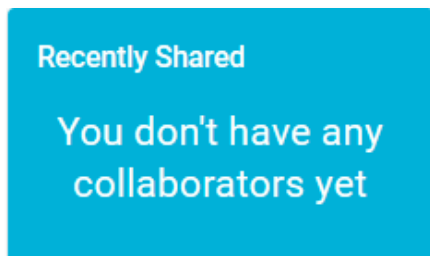


Figure 8 Recently Shared

6. **Active Subscription:** Your Subscription Plan and its expiration date.



Figure 9 Subscription Plan

7. **Collaborators:** The total number of your Collaborators, other subscribers with whom you have shared projects.

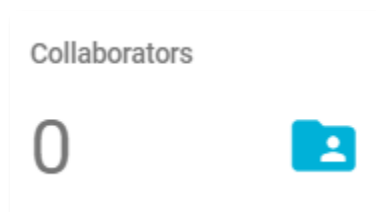


Figure 10 Collaborators

8. **Recent Analyses:** A list of all currently running Analyses.

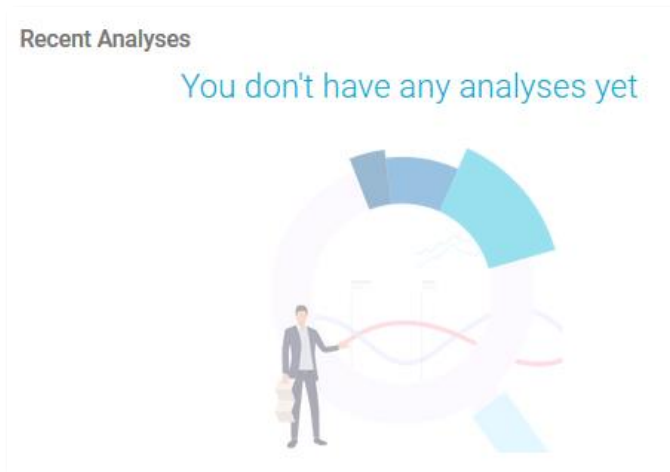


Figure 11 Recent analyses

9. **Recent projects:** A list with links to your current Projects.

The image shows a screenshot of the 'Recent Projects' section. It features a table with the following columns: Name, Updated, Owner, Dataset, and Actions. There is one row of data.





Name	Updated	Owner	Dataset	Actions
 JAD Use Cases - Demo Use Cases f	2020-09-...	vlagani		...

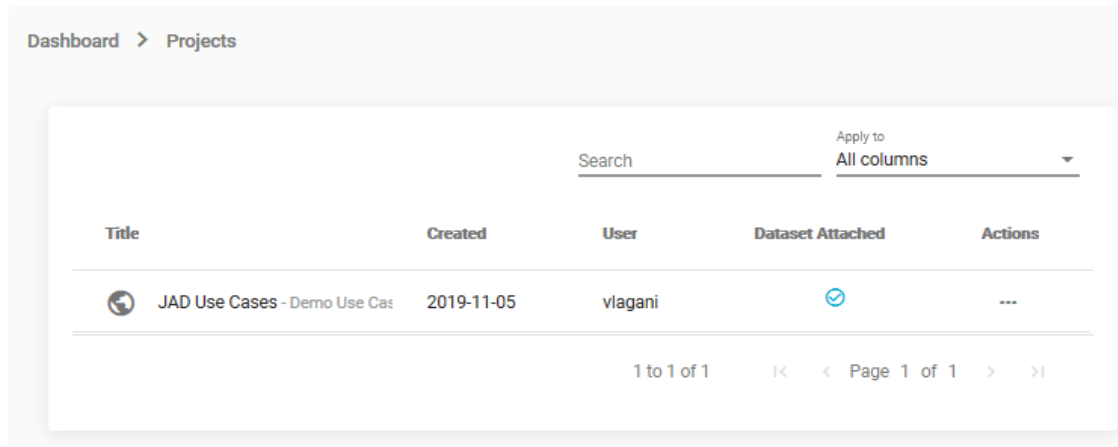
Figure 12 Recent Projects

Also, there is a  button in the top right of your dashboard that allows you to create a project and a  button that that open your **Projects** window.

Projects

- Click on the **LIST PROJECTS** button.

Note: The **Projects** includes a shared project, **JAD Use Cases**, which includes several datasets in order for you to have some examples to work with in JADBio. You cannot change the data in this public project, but, as you can see later, you will be able to import this data into another project.





Title	Created	User	Dataset Attached	Actions
 JAD Use Cases - Demo Use Cas	2019-11-05	vlagani		...

Figure 13 Projects list

Each project is described by the following characteristics:

1. An icon that describes the number of collaborators and sharing status. The **JAD Use Cases**, has global icon, because this data is shared publicly.
2. The name of the project, and a short description, both of which can be used as values for a project search in the **Filter** tool at the top of the **Projects** window.
3. The date the project was **Created**.
4. The initial creator of the Project.
5. A flag to indicate the presence of a dataset.
6. **Actions** that will allow you to view or remove the project.

- For **JAD Use Cases**, under **Actions**, Click on the **View project** icon.

Within the **JAD Use Cases** Project you will have several layers of information:

1. Below the JADBio header, you will now see that you are in **Dashboard > Projects > JAD Use Cases**.

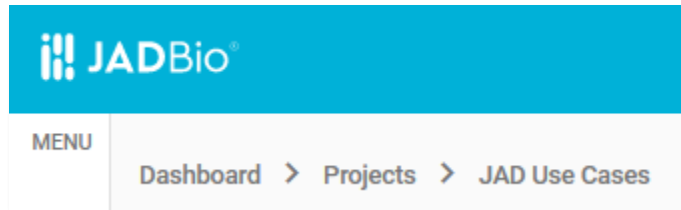


Figure 14 Breadcrumbs, JAD Use cases

2. In the top left, the **ACTIONS** sidebar, you will have five buttons that will allow you to perform a variety of functions with the Project:
 - a. **Add Data**
 - b. **Transform Data**
 - c. **Analyze Data**
 - d. **Apply Model**
 - e. **Delete Project**
3. In the same sidebar, JADBio provides the option to view/add **Collaborators** to your Project.
4. In the **PROJECT DETAILS** sidebar, JADBio provides information about your Project: **Name**, **Owner**, and a short **Description** of your Project, if one was created.
5. JADBio provides navigation to three different views of your Project window: **Datasets**, **Analyses**, and **Applied Models**.

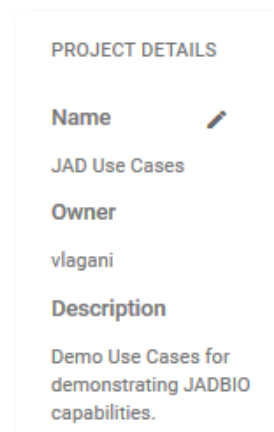


Figure 15 Project info

Datasets

- Click on the **Datasets** label, to view the available Datasets in the **JAD Use Cases** project.

Datasets includes a description of all of the datasets in a tabular format:

- i. Name
- ii. Created
- iii. Features
- iv. Samples
- v. Size
- vi. Under the **Actions** column, JADBio provides buttons to:
 - Preview Dataset
 - Perform Analysis
 - Detach Dataset

Datasets		Analyses	Applied Models			
					Search	Apply to All columns
Name	Created	Features	Samples	Size	Actions	
COVID-19_GSE152075 - RNA	2020-08-13	35787	484	38.11MB	...	
BrCa_paired_methylation - T	2020-07-24	485580	116	871.70MB	...	
BrCa_paired_transcriptomic	2020-07-01	48703	114	45.21MB	...	
MultiClassExpressionCance	2020-06-04	16064	190	11.60MB	...	
Psoriasis_SNP_test - The ori	2020-05-13	1082	2250	12.20MB	👁️ ▶️ ✕	
Psoriasis_SNP_train - The ori	2020-05-13	1082	5103	27.65MB	...	
P_falciparum_CNV - This dat	2020-04-27	11374	183	15.46MB	...	
Oyster_transcriptomics - Thi	2020-04-26	31921	177	40.56MB	...	
Multiple_sclerosis_proteomi	2020-04-24	9304	80	5.80MB	...	

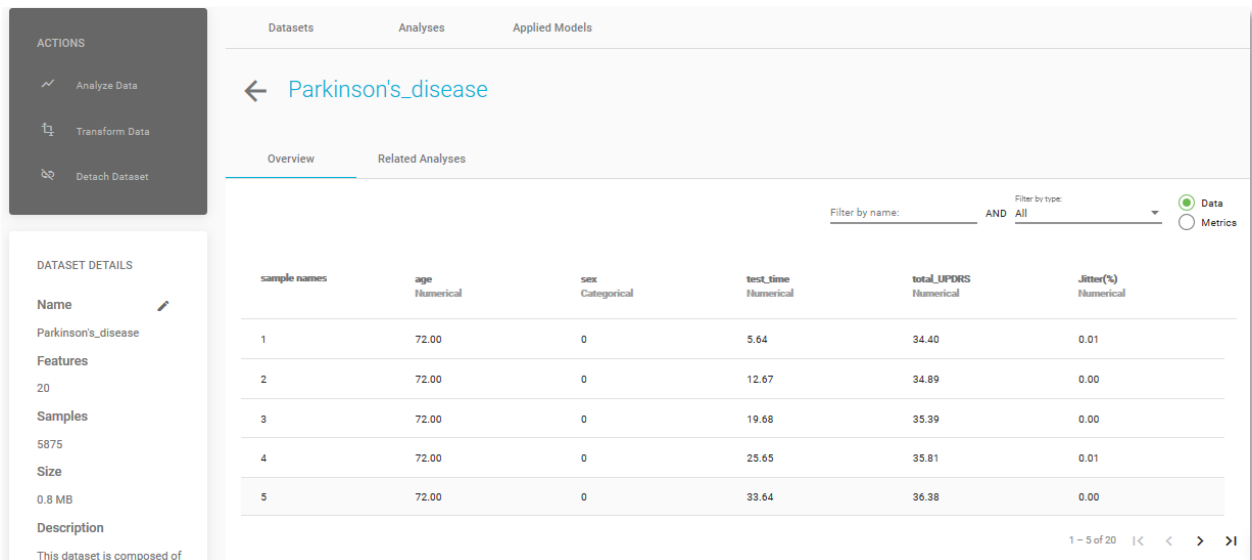
Figure 16 JAD Use Cases, Datasets

- Find the **Parkinson's disease** dataset.
- Click on the three dots, to view the **Actions** menu associated with **Parkinson's disease** dataset.

- Click on the **Preview dataset**  icon.

In Preview dataset:

- **ACTIONS** sidebar, provides 3 possible actions to perform with the dataset: to **Analyze Data**, to **Transform Data** or to **Detach Dataset**.
- **DATASET DETAILS** sidebar, includes: **Name**, number of **Features**, number of **Samples**, file **Size** and a short **Description** of the dataset, if one was created.
- **Overview** displays the dataset's column and row labels for the first five samples and first five features with their assigned data types. Tools to navigate or Filter, by name or by type, are embedded in the preview.
- **Related Analyses** displays the previous run analyses of the dataset, when the data have been analyzed previously, as in this Project.



sample names	age Numerical	sex Categorical	test_time Numerical	total_LUPORS Numerical	Jitter(%) Numerical
1	72.00	0	5.64	34.40	0.01
2	72.00	0	12.67	34.89	0.00
3	72.00	0	19.68	35.39	0.00
4	72.00	0	25.65	35.81	0.01
5	72.00	0	33.64	36.38	0.00

Figure 17 Preview dataset, Overview

- In **Overview** tab, Select the **Metrics** radio button to view graphical and numerical summaries of each feature in the dataset (i.e. histograms).

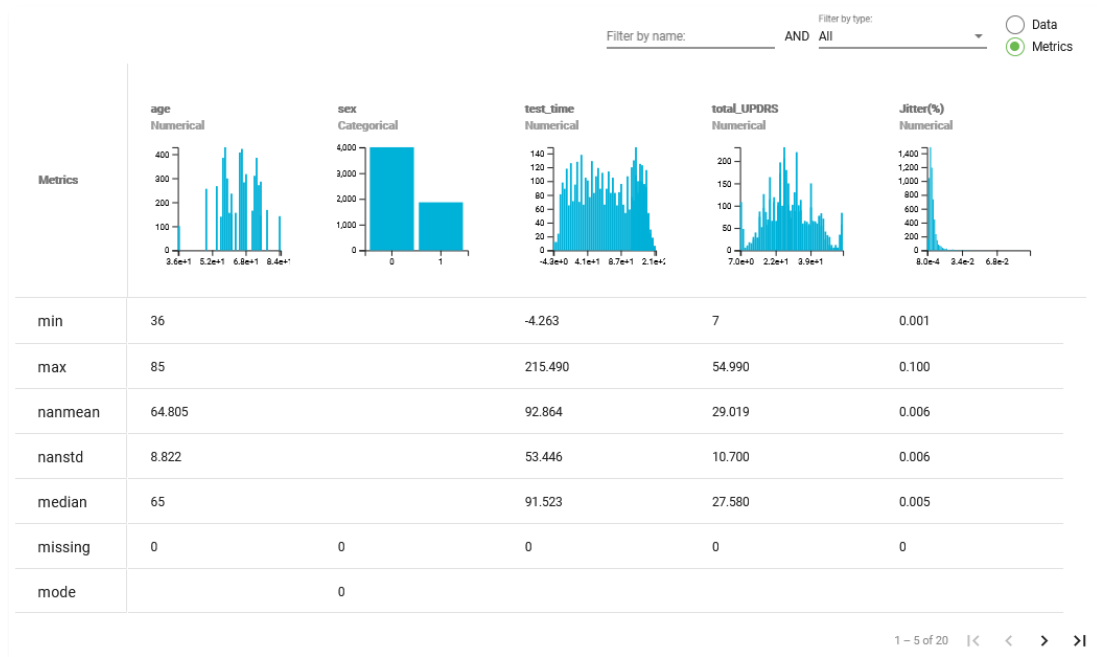


Figure 18 Metrics view

In the **Metrics** view, for the total_UPDRS feature, one can see that the nanmean value is 29.019 with minimum 7 and maximum 54.990. Note: nanmean is the mean excluding missing values.

Under **Actions**, JADBio gives you the option to **Perform Analysis** of the dataset. We will perform an analysis with this data later in the tutorial, but in a new Project.

Analyses

- Click on the **Analyses** label, to view the previously run Analyses in the **JAD Use Cases** project.





















Datasets		Analyses			Applied Models			
		Search			Apply to		Columns to display	
					All columns		Name, Started, Dataset, M...	
Name	Started	Dataset	Metric	Metric Value	Outcome	Type	Progress	Actions
 Alzheimer's_transc	2020-09-07 11:34	Alzheimer's_transc	AUC	0.98	diagnosis		FINISHED (100%)	...
 COVID-19_GSE152	2020-09-02 23:53	COVID-19_GSE1520	AUC	0.96	covid		FINISHED (100%)	...
 Tuberculosis_prote	2020-07-15 12:39	Tuberculosis_protec	AUC	0.77	Status		FINISHED (100%)	...
 Parkinson's_disease	2020-07-02 23:16	Parkinson's_disease	R2	0.84	total_UPDRS		FINISHED (100%)	...
 Chemosensitivity_I	2020-07-02 16:56	Chemosensitivity_P	AUC	0.78	target		FINISHED (100%)	...
 Chemosensitivity_I	2020-07-02 16:46	Chemosensitivity_P	AUC	0.79	target		FINISHED (100%)	...
 Parkinson's_disease	2020-07-02 08:37	Parkinson's_disease	R2	0.84	total_UPDRS		FINISHED (100%)	...
 MultiClassExpress	2020-06-26 17:39	MultiClassExpressic	AUC	0.90	class		FINISHED (100%)	...
 Tuberculosis_prote	2020-06-25 18:53	Tuberculosis_protec	AUC	0.77	Status		FINISHED (100%)	...
 Oyster_transcriptoi	2020-06-25 18:17	Oyster_transcriptorr	AUC	0.92	Population		FINISHED (100%)	...

Figure 19 JAD Use Cases, Analyses

Within the **Analyses** page, JADBio describes all analyses, including those currently running. Furthermore, the analysis type is described by an icon. Here, we have **Regression**, **Classification**, and **Survival (Time to Event)** analyses.

Analyses also includes a description in tabular format:

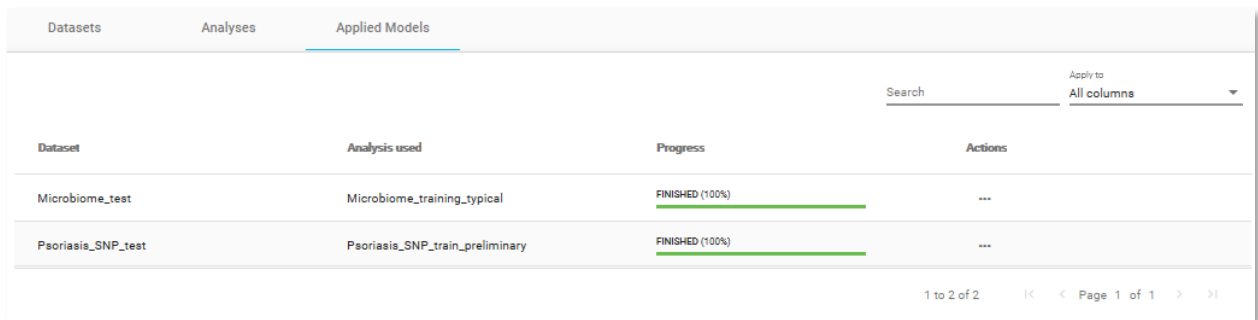
- i. Name
- ii. Started (date)
- iii. Dataset
- iv. Metric
- v. Metric Value
- vi. Outcome
- vii. Type
- viii. Progress
- ix. Actions, includes: View Results, Apply Model and Remove Analysis buttons

Applied Models

- Click on the **Applied Models** label, to view the results from the reapplication of a model on a novel or test dataset.

Applied Models description includes:

- i. Dataset
- ii. Analysis used
- iii. Progress
- x. Actions, includes:
 - View Results
 - Remove Applied Model



Dataset	Analysis used	Progress	Actions
Microbiome_test	Microbiome_training_typical	FINISHED (100%)	...
Psoriasis_SNP_test	Psoriasis_SNP_train_preliminary	FINISHED (100%)	...

Figure 20 JAD Use Cases, Applied Models

Working in JADBio

Creating a Project

All of the datasets and analyses within the **JAD Use Cases** project are read-only. In order to start working with data, you must create a new project. For this example, you can create a new project and upload data from the **JAD Use Cases** project.

- On the **MENU** sidebar at the left of the JADBio window, click on **Projects**.

This will bring you to the **Projects** window, which includes the **Create a new Project** function.

- Click on **CREATE** button.
- Within the **Name your Project** dialog, enter the following **Project title** and **Project Description**.



Figure 21 Create a new project

- Click on the **CREATE** button.

Your new project will now appear in the **Projects** window.

Uploading a Dataset

- Under **ACTIONS**, in the top left corner, click **Add Data** icon to add a dataset to this project.

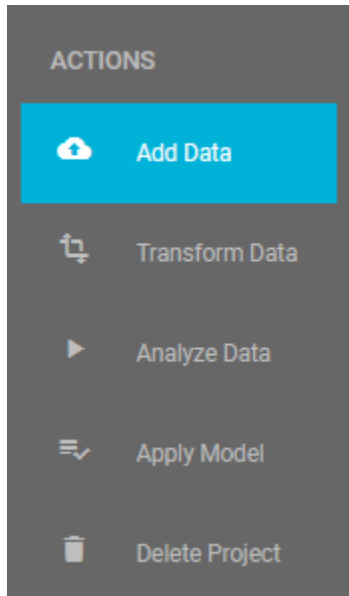


Figure 22 ACTIONS options in project window, prior to the upload of any datasets.

You will have two options. If you were working with an external file, you would select the **Upload a file** option, but for this demonstration, you will use the existing **Parkinson's disease** dataset from the **JAD Use cases** project.

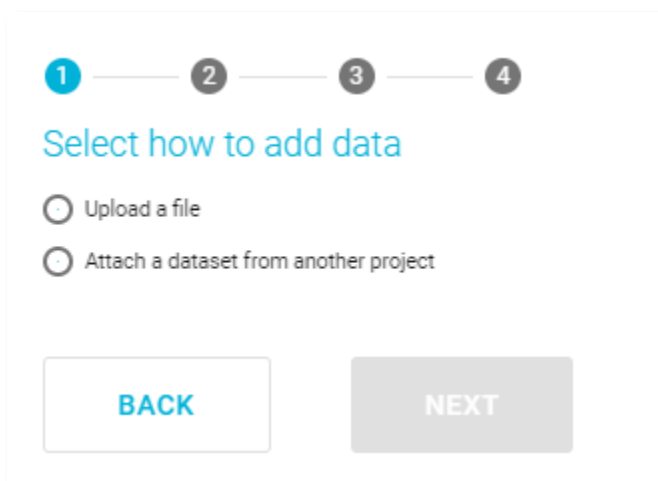


Figure 23 Add Data, Select how to add data

- Click on the radio button, **Attach a dataset from another project.**

JADBIO will now present you with a list of datasets in a tabular format.

- Select the **Parkinson's disease** dataset. (Note: You may have to search for it or scroll down to see it.)

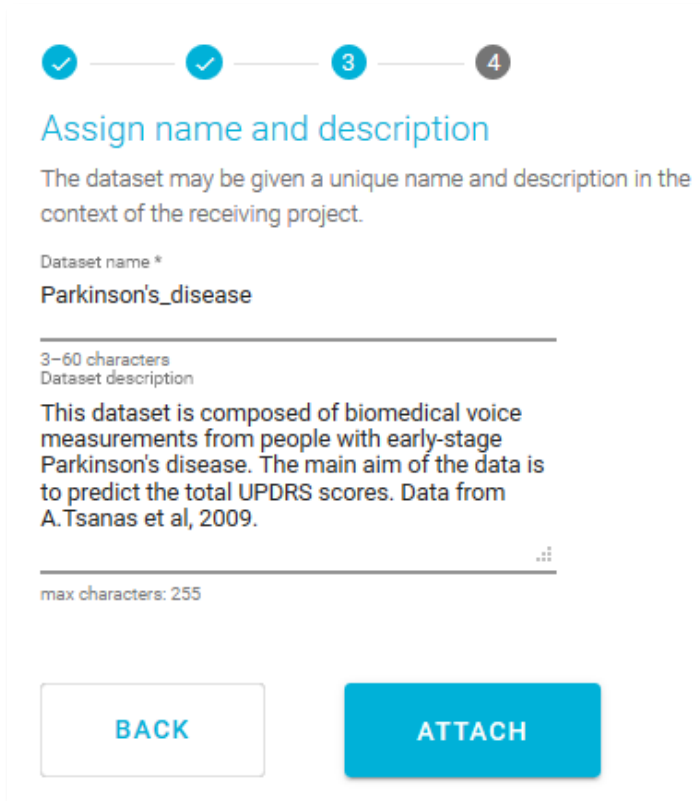


Figure 24 Select a dataset to attach

- Click the **NEXT** button.

This opens the **Assign name and a description** window. Here, you are able to change your dataset's name and add a description for your dataset.

- Click the **ATTACH** button, to connect this dataset with your project.



✓ — ✓ — 3 — 4

Assign name and description

The dataset may be given a unique name and description in the context of the receiving project.

Dataset name *

Parkinson's_disease

3-60 characters

Dataset description

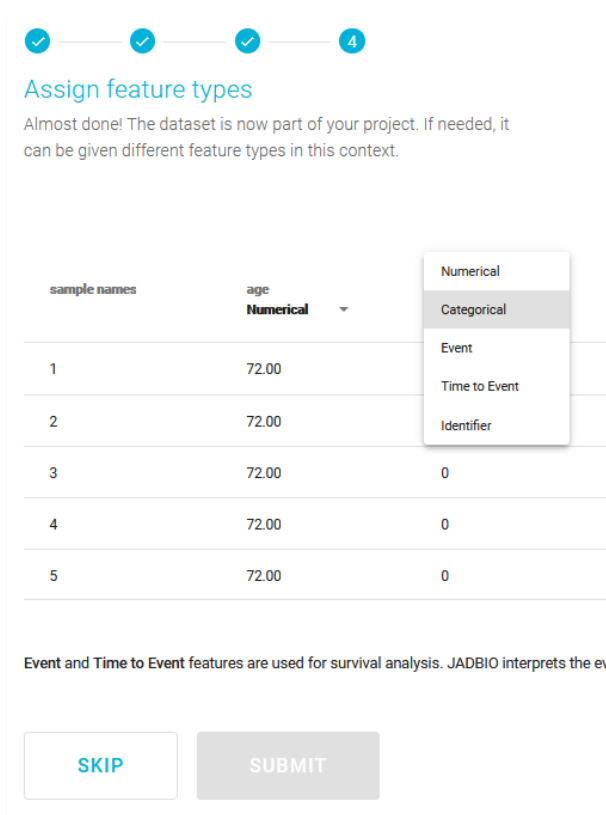
This dataset is composed of biomedical voice measurements from people with early-stage Parkinson's disease. The main aim of the data is to predict the total UPDRS scores. Data from A.Tsanas et al, 2009.

max characters: 255

Figure 25 Assign name and description

This opens the **Assign feature types** window. JADBio automatically assigns feature types based on the content of each column, but you can change the feature types here or in a transformation step to suit your understanding of your data. Here, you can change the type of features i.e. to assign an event or a time-to-event feature type. Also, as in the **Preview dataset** window, you are able to navigate across the dataset, search and filter the dataset, and view the **Metrics** of each feature in the **Metrics** view.

Note: Feature types are critical to the process of using the JADBio platform. If you are not familiar with feature types, please reach out to technical support for additional information.



Assign feature types

Almost done! The dataset is now part of your project. If needed, it can be given different feature types in this context.

sample names	age	
1	72.00	
2	72.00	
3	72.00	0
4	72.00	0
5	72.00	0

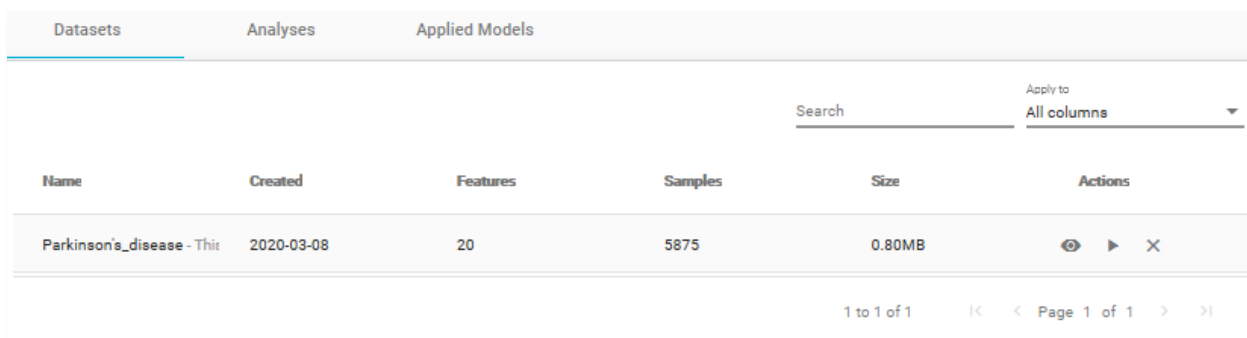
Event and Time to Event features are used for survival analysis. JADBIO interprets the ev

SKIP **SUBMIT**

Figure 26 Assign feature types

- Click **SKIP** to move the data, without changing any feature type, into your project.

You are now in the **Datasets** window of the **Parkinson's_disease_demo** Project. You can see that your new dataset has 20 Features and 5875 Samples. Under the **Actions** header on the three dots, you have three options: **Preview Dataset**, **Perform Analysis** and **Detach Dataset** from project.



Datasets		Analyses	Applied Models			
Name	Created	Features	Samples	Size	Actions	
Parkinson's_disease - This	2020-03-08	20	5875	0.80MB	Search <input type="text"/> Apply to All columns	

1 to 1 of 1 << < Page 1 of 1 > >>

Figure 27 Uploaded dataset.

Transformations

JADBio provides numerous transformation tools today and we will continue to augment this functionality to support your needs. For this demonstration, you will split the data into two datasets, one for training and one for testing.

- Click on **Transform Data** in the **ACTIONS** sidebar.

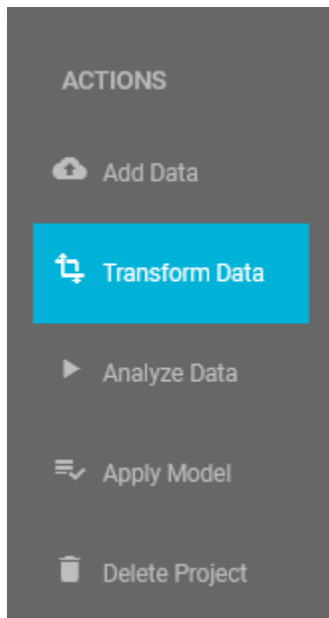
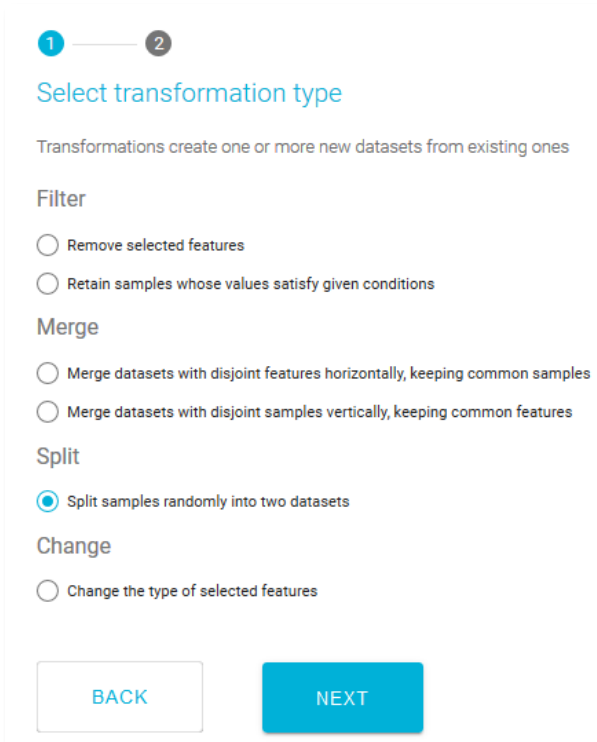


Figure 28 ACTIONS sidebar, Transform Data

Transformation takes place in two steps: **1. Select transformation type. 2. Transformation configuration**, and JADBio currently provides four different transformation types: **Filter, Merge, Split and Change** dataset. You are going to create two datasets, **test** and **training**, from the imported dataset, so you will use the **Split** function.

- Select the **Split** option – **Split samples randomly into two datasets.**
- Click **NEXT.**



1 — 2

Select transformation type

Transformations create one or more new datasets from existing ones

Filter

Remove selected features

Retain samples whose values satisfy given conditions

Merge

Merge datasets with disjoint features horizontally, keeping common samples

Merge datasets with disjoint samples vertically, keeping common features

Split

Split samples randomly into two datasets

Change

Change the type of selected features

BACK NEXT

Figure 29 Transform data

This opens the wizard, **Configure sample splitting**.

- From the **Select input dataset** pulldown, select the **Parkinson's_disease** dataset.
- Name the first output dataset **Parkinson's_training** and name the second, **Parkinson's_test**.
- Leave the Split percentage at **70%**.

Configure sample splitting

This transformation splits samples randomly into two new datasets

Select input dataset: **Parkinson's_disease**

Name of first output dataset*: **Parkinson's_training**
At least 3 characters

Name of second output dataset*: **Parkinson's_test**
At least 3 characters

Split percentage (70%): 70%70%

Advanced options

Preview dataset?

Click Transform to split the samples of **Parkinson's_disease** into **Parkinson's_training** (70%) and **Parkinson's_test** (30%).

Figure 30 Transformation configuration

Note: Advanced options will allow you to select a feature for stratified splitting and/or a feature that defines how samples are grouped. This is not necessary for this demonstration.

- Click on the **TRANSFORM** button.

You will now have three datasets in your **Parkinson's_disease_demo** project.

- Note the sample split in the two new datasets.

Datasets		Analyses	Applied Models			
Name	Created	Features	Samples	Size	Actions	
Parkinson's_test	2020-09-26	20	1762	0.24MB	---	
Parkinson's_training	2020-09-26	20	4113	0.56MB	---	
Parkinson's_disease - This dataset is c	2020-03-08	20	5875	0.80MB	---	

1 to 3 of 3 | Page 1 of 1

Figure 31 Project Datasets

Background

Parkinson's disease (PD) is the second most common neurodegenerative disorder that affects movement. Symptoms start gradually, sometimes starting with a barely noticeable tremor in just one hand. Management of PD typically involves the administration of physical examinations applying various empirical tests, including speech and voice tests. The Unified Parkinson Disease Rating Scale (UPDRS) is a rating tool used to gauge the course of Parkinson's disease in patients. Tracking Parkinson's disease symptom progression based on UPDRS assessment requires the patient's presence in clinic, and time-consuming physical examinations by trained medical staff. Thus, symptom monitoring is costly and logistically inconvenient for patient and clinical staff alike, also hindering recruitment for future large-scale clinical trials.

An accurate remote monitoring (telemonitoring) tool of PD progression that predicts UPDRS scores could be derived from the analysis of simple, noninvasive vocal measurements. In this demonstration, you will create, train and test a model based on 5875 recordings of speech signals and clinical information including age and gender from PD patients (see below original publication). You will see that with JADBio you will produce accurate predictive models regardless of your expertise in advanced machine learning techniques.

Original Publication: Tsanas et al. "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests." *IEEE Trans Biomed Eng* (2010) Apr;57(4):884-93

Analysis

With both a training dataset and a test dataset, within JADBio you are able to train and test a predictive signature for predicting the UPDRS score based on speech signals collected at the PD patient's home using a telemonitoring system.

- From the table of **Datasets**, in the **Parkinson's_training** dataset click on **Perform Analysis**.

Name	Created	Features	Samples	Size	Actions
Parkinson's_test	2020-09-26	20	1762	0.24MB	...
Parkinson's_training	2020-09-26	20	4113	0.56MB	👁️ ▶️ ✕
Parkinson's_disease - This data	2020-03-08	20	5875	0.80MB	Perform Analysis

Figure 32 Parkinson's_training dataset, Perform Analysis

- Select outcome: total_UPDRS**, by placing a checking the box next to the name of the feature and click **NEXT**.

1
2

Select outcome

For regression analysis, select a Numerical feature. i
 For classification analysis, select a Categorical feature.
 For survival analysis, select both an Event and a Time to Event.

Filter by name: _____ Filter
 AND All

sample names	age <input type="checkbox"/> <small>Numerical</small>	sex <input type="checkbox"/> <small>Categorical</small>	test_time <input type="checkbox"/> <small>Numerical</small>	total_UPDRS <input checked="" type="checkbox"/> <small>Numerical</small>
1	72.00	0	5.64	34.40
3	72.00	0	19.68	35.39
4	72.00	0	25.65	35.81
6	72.00	0	40.65	36.87
7	72.00	0	47.65	37.36

Everything looks good.
 Click Next to perform a regression analysis, finding a model that predicts total_UPDRS.

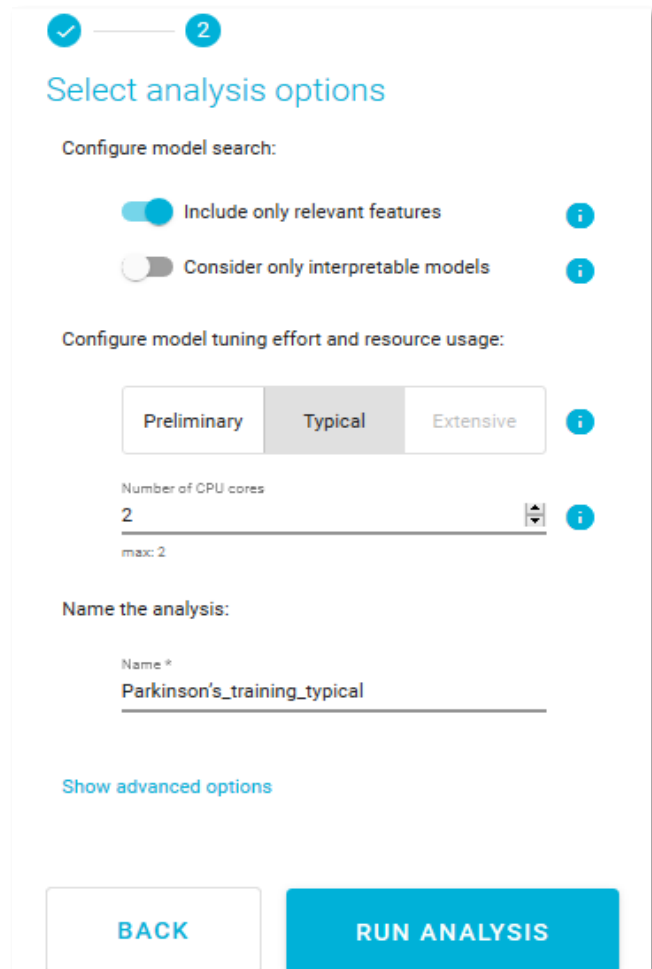
Figure 33 Select outcome - predicted feature

Note: Because you have chosen an outcome in which the values are number, a continuous feature, JADBio will create a model based on a regression analysis.

- **Select analysis options:** Set the Predictive analysis options as follows:
 - Include only relevant features: (This ensures that the models will only include the most relevant features.)
 - Tuning Effort: **Typical**, for an intermediate assessment.
 - Number of CPU cores: **2** (In order your analysis to be completed faster.)

JADBio will automatically create a descriptive name for your analysis based on your selections, which you can change.

- Name the Analysis:
Parkinson's_training_typical



✓ — 2

Select analysis options

Configure model search:

Include only relevant features ⓘ

Consider only interpretable models ⓘ

Configure model tuning effort and resource usage:

Preliminary Typical Extensive ⓘ

Number of CPU cores

2 ⓘ

max: 2

Name the analysis:

Name *

Parkinson's_training_typical

[Show advanced options](#)



BACK RUN ANALYSIS

Figure 34 Select analysis options

While there are many other options, the default values for the advanced settings are the following:

- ✓ There is no requirement for an **Identifier feature**, to group the samples of the dataset.
 - ✓ The analysis will optimize performance based on the R^2 .
 - ✓ Maximum signature size is 25 features.
 - ✓ Maximum multiple signatures to visualize is 5.
 - ✓ JADBio will create two plots, Residuals and ICE.
- Click **RUN ANALYSIS** to start the analysis.

As soon as you begin the analysis, JADBio reports **Progress** in the **Analyses** table.

Datasets		Analyses			Applied Models			
Name	Started	Dataset	Metric	Metric Value	Outcome	Type	Progress	Actions
 Parkinson's_t	2020-09-26 18:20	Parkinson's_train	R2		total_UPDRS		<div style="width: 5%;"><div style="width: 5%;"></div></div> RUNNING (5%)	...

1 to 1 of 1 < < Page 1 of 1 > >

Figure 35 Running analysis

- Click on the blue progress bar to view learning progress, you can watch each step in the analysis progression.

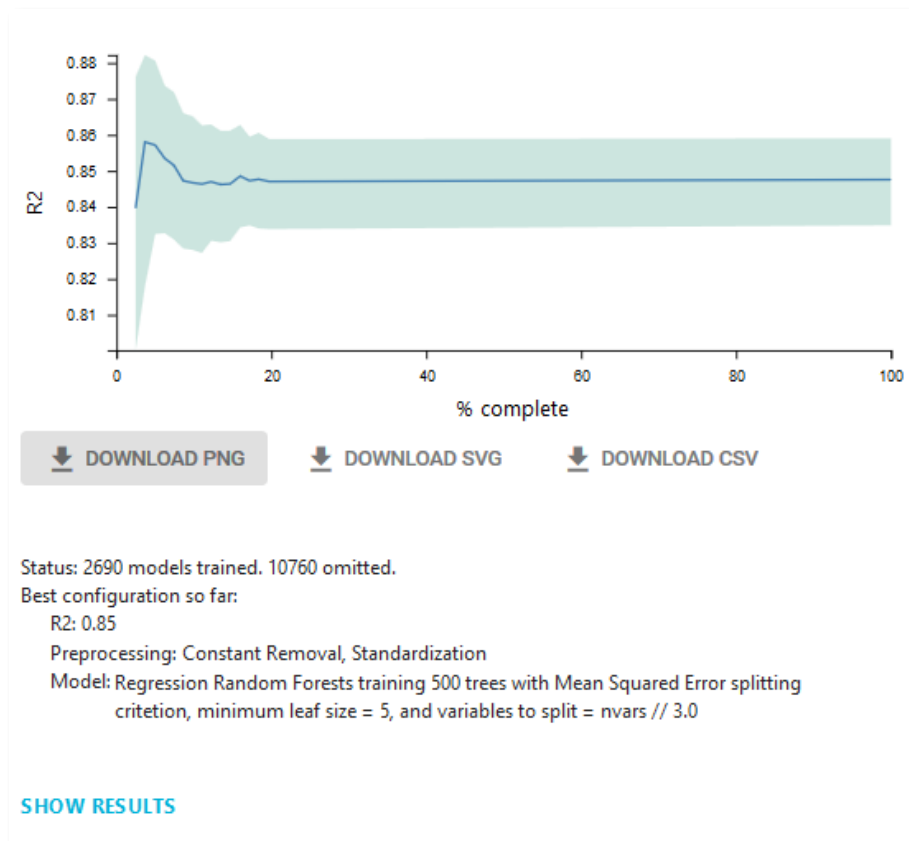


Figure 36 Learning progress

- As soon as the analysis is completed, click on **SHOW RESULTS** to view the results of the analysis

Analysis Results

Note: The results you achieve may be different than the results reported here, because we are constantly updating and improving JADBio.

In the **ANALYSIS ACTIONS** sidebar, JADBio provides options to:

1. Summarize the analysis (as an **Analysis Report**),
2. **Show Model** (if it is an interpretable one)
3. **Download Predictions** of the samples
4. **Apply Model** to new samples.

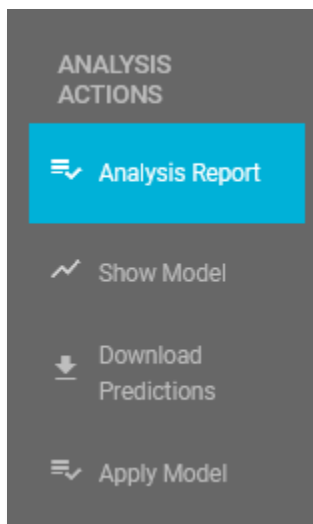


Figure 37 ANALYSIS ACTIONS sidebar


- Click on **Analysis Report**.


Here, JADBio provides a summary audit of the analysis including: **Dataset, Outcome to Predict, Analysis Type, etc.** and a description of the selected options as well as the **JADBio version**. Version 1.1.152 was used for this tutorial.

Summary	
Dataset	Parkinson's_training
Outcome to Predict	total_UPDRS
Analysis Type	Regression
Metric	R2
Grouping Feature	None
Relevant Features Only	Yes
Interpretable Models Only	No
Tuning Effort	Typical
CPU Cores	2
Configurations Tried	20
Models Trained	470 models trained.
Execution Time	00:15:24
JADBio Version	1.1.152

Figure 38 Analysis info

JADBio also presents a list of all of the configurations that were tested in order to produce the model and selected features. In this example, 20 configurations were tested which you can

download by clicking on  button.

- Click on  return button to return to the main analysis results page.

Note: In the top right corner, there is  link you can use to share your results.

Best Performing Model

The **Analysis** page provides an overview of the analysis process and a description of the **Best Performing Model** and the **Best Interpretable Model**.

The methods include the optimal configurations for: **Preprocessing**, **Feature selection**, and **Predictive algorithm**.

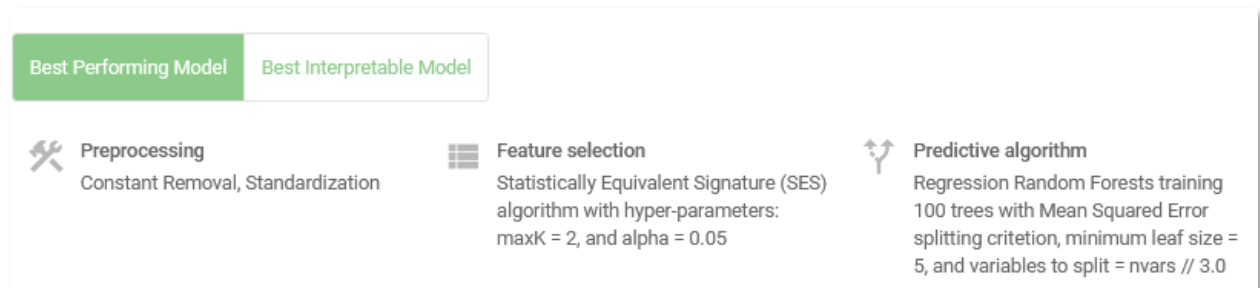


Figure 39 Best performing model

Note: When you selected the analysis parameters, you did not force JADBio to only report interpretable models. With this setting, if the Best Performing Model is not interpretable, JADBio will provide two models, **Best Performing Model** and **Best Interpretable Model**. In this case, the best performing model is Regression Random Forests and the best interpretable is Ridge Linear Regression.

- Click on the **Download Predictions**.

The downloaded **analysis_predictions.txt** table is best viewed in a spreadsheet. In it you will see values from each of the analyzed samples and two new columns, based on the cross validation of the best configuration. The new columns are **Prediction** and **Difficulty to Predict**. The **Difficulty to Predict** column is rated **TRUE** or **FALSE** based on the likelihood that the predicted outcome is correct. The **Prediction** column lists the calculated numerical value for a sample given the model and associated features. The **Label** column is the actual values from the dataset.

	A	B	C	D
1	Sample name	Prediction	Difficult to Predict	Label
2	1	37,90	FALSE	34,40
3	3	35,62	FALSE	35,39
4	4	34,31	FALSE	35,81
5	6	30,45	FALSE	36,87
6	7	36,54	FALSE	37,36
7	9	32,55	FALSE	38,35
8	12	35,81	FALSE	39,83
9	14	33,64	FALSE	41,03
10	15	31,93	FALSE	41,66
11	16	30,66	FALSE	42,28
12	17	27,79	FALSE	42,90
13	18	30,62	FALSE	43,52
14	20	27,23	FALSE	44,86
15	22	27,12	FALSE	46,01
16	23	39,13	FALSE	46,64
17	24	34,36	FALSE	47,97
18	25	34,65	FALSE	34,40
19	26	32,04	FALSE	34,89
20	27	36,70	FALSE	35,39
21	28	36,33	FALSE	35,81

Figure 40 Downloaded predictions viewed in spreadsheet

Performance Overview

For regression analysis, the performance of the classifier is described by seven different metrics. In this case, the R^2 (R-squared) is 0.843 [0.799, 0.881], indicating that the speech signals carry useful information in order to estimate the UPDRS score. Also, the Mean Absolute Error is 3.003 [2.639, 3.366] and the Mean Squared Error is 17.858 [13.295, 22.822].

- Click on the R^2 metric to display its distribution.

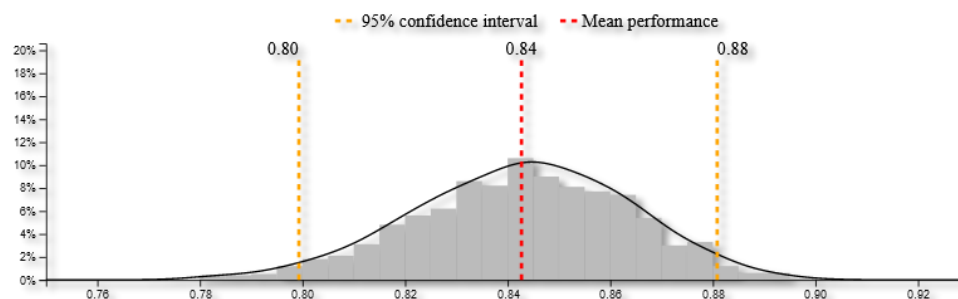


Figure 41 R^2 metric distribution

Below, JADBio reports 3 other metrics and their confidence intervals based on your **Best Performing Model**.

- Hover over any “i” adjacent to a metric for an explanation of the score.

Note: As you move your cursor in the JADBio windows, JADBio will provide contextual information or links to relevant locations within the application.

Feature Selection

Feature Selection is a process that identifies a minimal-size subset of features that is maximally predictive of the outcome of interest, the selected target feature.

- Select the **Feature Selection** tab.

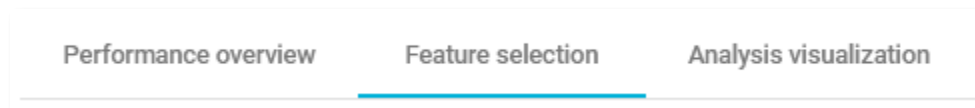


Figure 42 Feature selection

Selected Signatures

A signature is a minimal subset of predictive features that, when considered jointly, are maximally informative for an outcome of interest. As a product of each analysis, JADBio produces all signatures that perform equally well, up to the maximum limit defined in parameters. In this example the column represents one signature with a total of 6 features. For this example, only one signature is produced but for other analyses this list can be large.

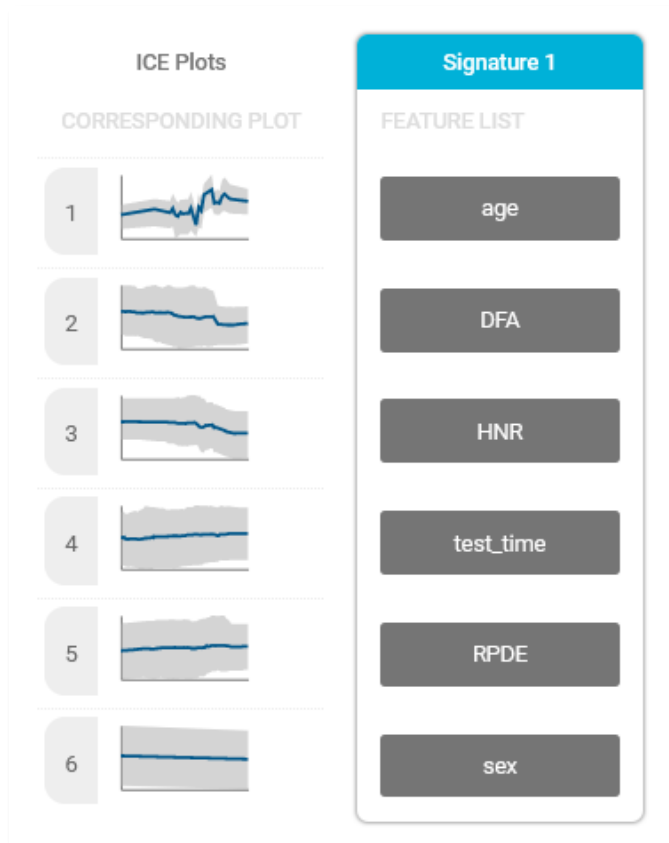


Figure 43 Selected signature

The **Individual Conditional Expectation** (ICE) plots further reveal the nature of the contribution of each feature to the model.

- Click on the thumbprint ICE plot adjacent to HNR feature to enlarge the ICE plot.

This opens the ICE plot for the prediction of HNR.

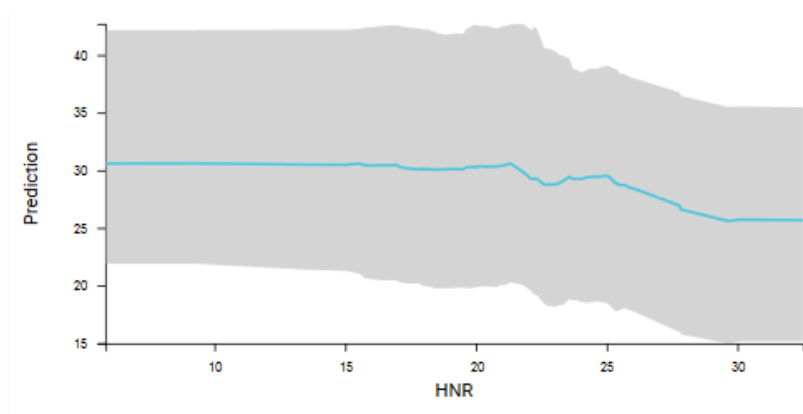


Figure 44 Individual Contribution Expectation plot for HNR predictor

- Scroll down to view the histogram of HNR predictor.

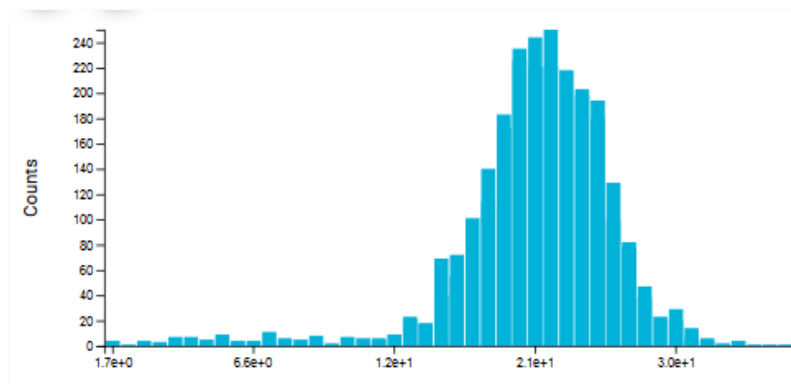


Figure 45 Histogram of HNR predictor

Feature Importance plots

The practical use of **Feature Importance plots** is evident in the case of selecting biomarkers. For instance, the purpose of this analysis is to identify the optimal list of biomarkers that predict UPDRS score. However, in order to satisfy economical or technical constraints on an assay, JADBio also reports the cost to performance that occurs when one chooses to further reduce the total number of predictive biomarkers from those included in the Best Performing Model. In this way, you, can evaluate the trade-off between reducing the number of biomarkers and achieving optimal performance.

Both in the **Progressive Feature Inclusion** and in the **Feature Importance** plots, JADBio displays the six features of the selected signature and their relative performance.

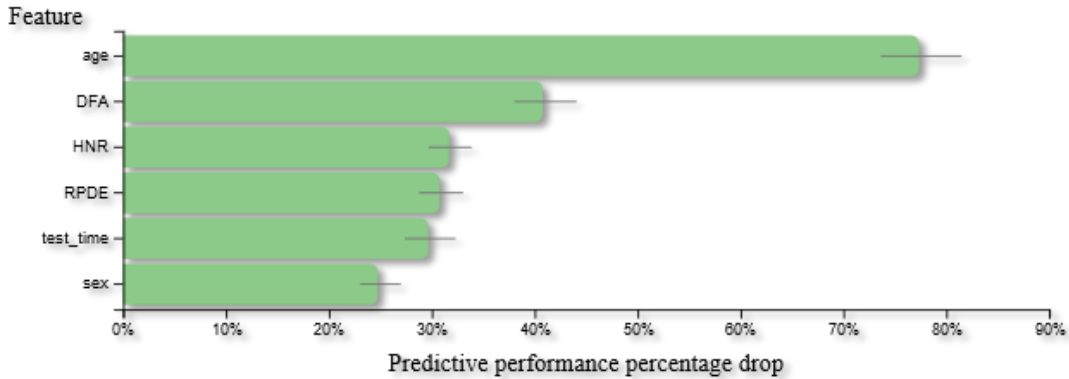


Figure 46 Feature Importance plot (The image used here is from the downloaded PNG, rather than a screen shot).

Analysis Visualization

- Select the **Analysis Visualization** tab.

Residuals plot: Residuals are the differences between observed values and those predicted by the model. Any pattern in the residuals is evidence of an inadequate model or of irregularities in the data, such as outliers, and suggests how the model may be improved. This plot shows Residuals (Actual – Predicted values) vs the Predicted values. Note that this plot preserves all outliers. For a perfect model, residuals are randomly and symmetrically distributed around zero.

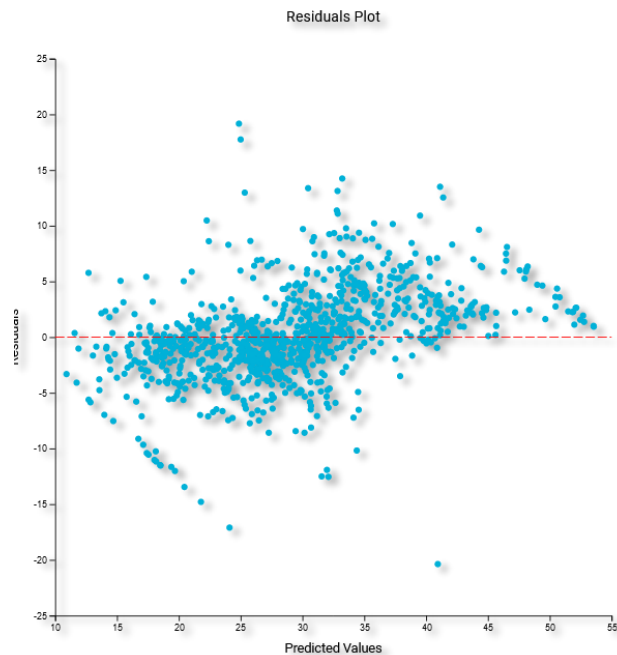


Figure 47 Residuals plot (The image used here is from the downloaded SVG, rather than a screen shot).

Best interpretable model

Now, Let's have a quick look at the **Best Interpretable Model**.

- Click on the **Best Interpretable Model** button.

Now, the predictive algorithm changed to an interpretable one (Ridge Linear Regression algorithm).

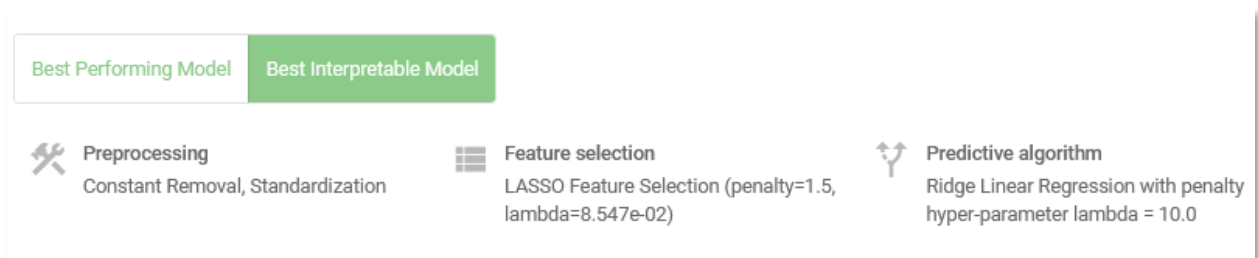


Figure 48 Best Interpretable Model

- Click on **Show Model** on the **ANALYSIS ACTIONS** sidebar.

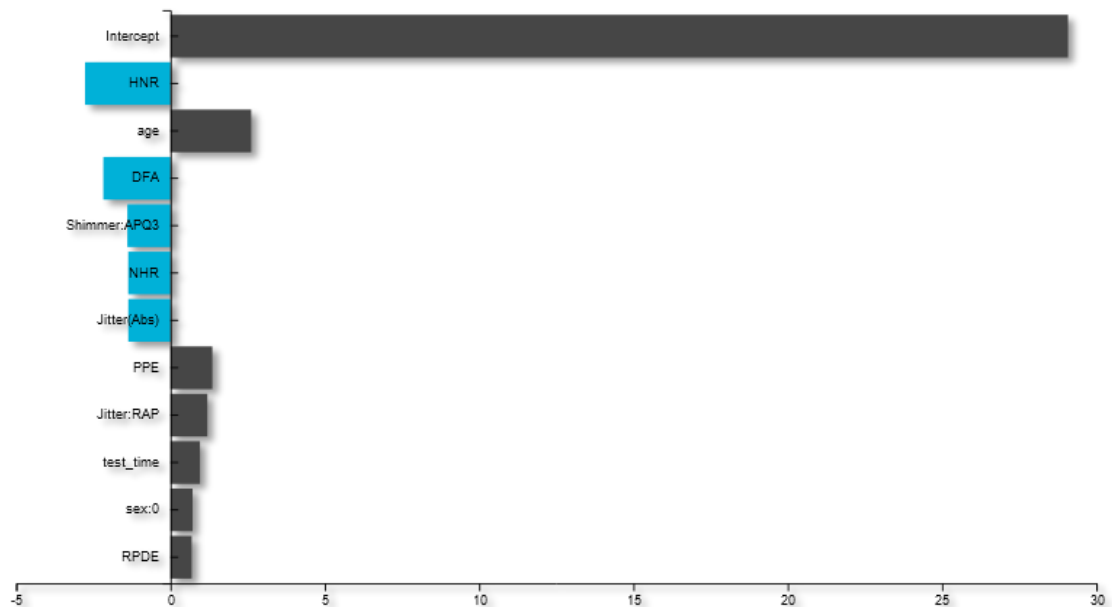


Figure 49 Show model

JADBio displays 11, out of the total 20 features, and the intercept that provide the most accurate prediction of the UPDRS score.

- Hover over the features bar to visualize the values for the optimized Ridge Linear Regression model, which can be described as:

The numbers provided describe the relative strength of the predictors based on the logistic model. The larger the absolute value of the feature's value in the model, the greater the impact that feature is in the analysis of any one individual's outcome. In this Parkinson's example, HNR and age are the stronger level of evidence for the prediction of the UPDRS score.

Note: It is possible to download both the image and the numbers supporting the image from the



buttons.

Performance Overview, **Feature selection** and **Analysis visualization** tabs contain the same type of information as that presented in the **Best Performing Model** analysis.

If you compare the **Visualizations** from the **Best Performing Model** to that of the **Best Interpretable Model**, you can easily see that the relationship between the predicted values and the actual values is much tighter in the **Best Performing Model**.

Apply Model

All of the above analysis was performed on the training dataset. Now that the analysis of the trained model is complete, you can run the test dataset.

- Scroll to the top of the page, and click on the **Parkinson's_disease_demo** project label.
- In the **ACTIONS** sidebar, click on **Apply Model**.

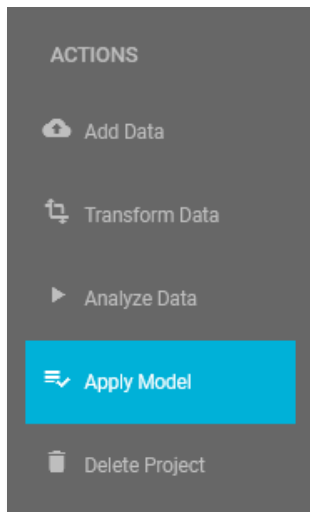
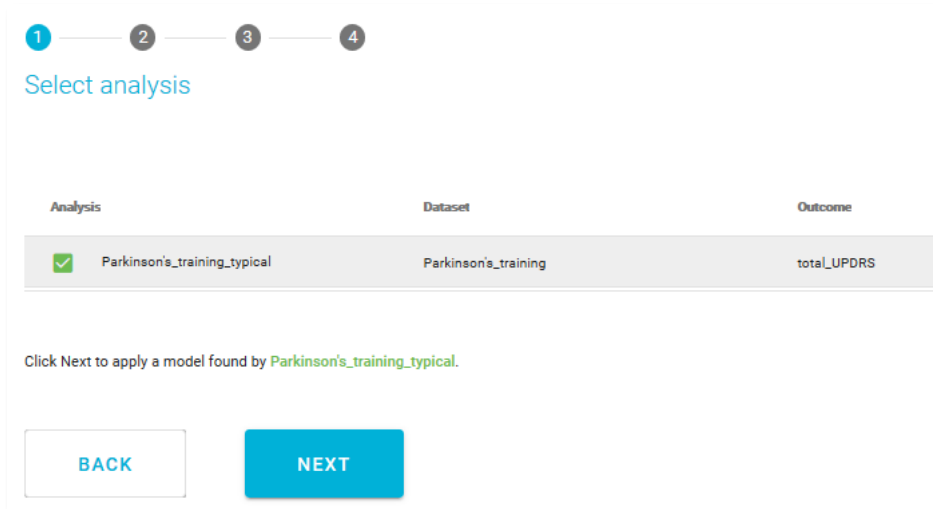


Figure 50 ANALYSIS ACTIONS, Apply Model

There are four steps to the validation process:

1. Select Analysis:

- Check the only analysis option available in the Project, **Parkinson's_training_typical**, which you just created and reviewed.
- Click **NEXT**.



1 — 2 — 3 — 4

Select analysis

Analysis	Dataset	Outcome
<input checked="" type="checkbox"/> Parkinson's_training_typical	Parkinson's_training	total_UPDRS

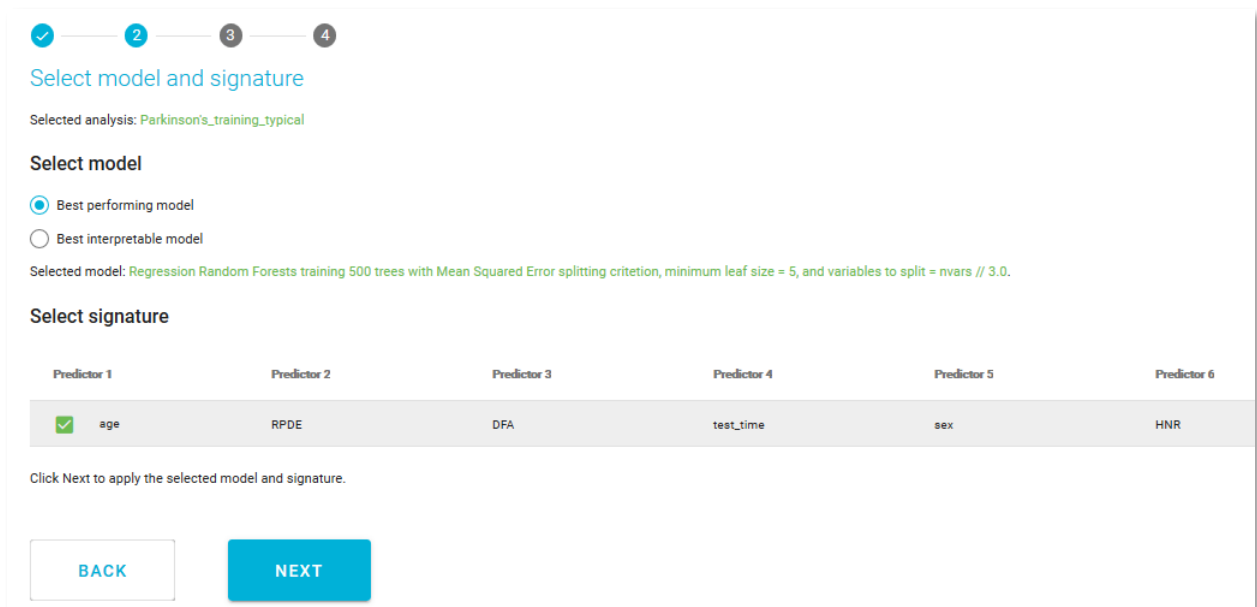
Click Next to apply a model found by [Parkinson's_training_typical](#).

[BACK](#) [NEXT](#)

Figure 51 Select analysis

2. Select model and signature:

- **Select model:** Select the best performing model
- **Select signature:** Place a check in the available signature.
- Click **NEXT**.



1 — 2 — 3 — 4

Select model and signature

Selected analysis: [Parkinson's_training_typical](#)

Select model

Best performing model

Best interpretable model

Selected model: [Regression Random Forests training 500 trees with Mean Squared Error splitting criterion, minimum leaf size = 5, and variables to split = nvars // 3.0.](#)

Select signature

Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	Predictor 6
<input checked="" type="checkbox"/> age	RPDE	DFA	test_time	sex	HNR

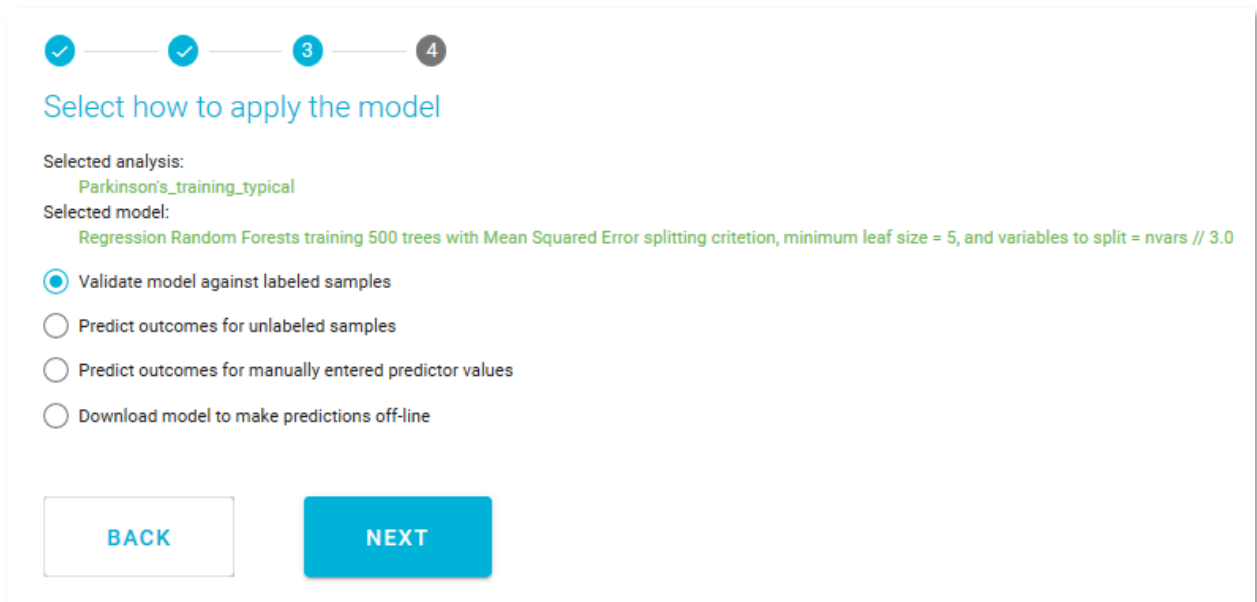
Click Next to apply the selected model and signature.

[BACK](#) [NEXT](#)

Figure 52 Model and signature selection

3. **Select how to apply the model:** Select the option, **Validate model against labeled samples**.
 - Click **NEXT**.

Note: This selection allows JADBio to use the Test data set you created when you split the original dataset into Training and Test datasets. If you choose to **Predict outcomes for Unlabeled samples**, JADBio will give you the option to upload another dataset. The only requirement you will have for this dataset is that it includes the Predictors in the selected signature with the same names as in the training data. If you select to **Predict outcomes for manually entered predictor values**, JADBio will provide a dialog for manually entering values for each of the Predictors in the Signature to generate a prediction. If you choose to **Download model to make predictions off-line** will give you the options to download a standalone version of the chosen model that can be applied on new data on a local machine.



3 4

Select how to apply the model

Selected analysis:
Parkinson's_training_typical

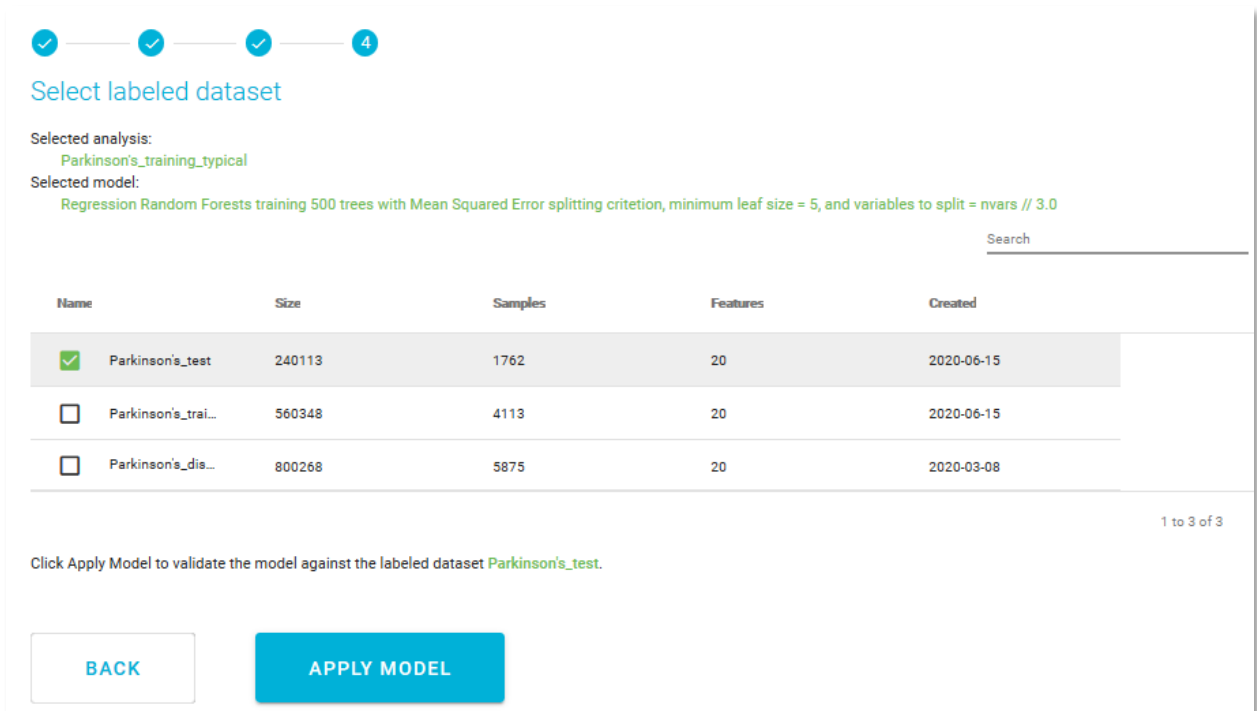
Selected model:
Regression Random Forests training 500 trees with Mean Squared Error splitting criterion, minimum leaf size = 5, and variables to split = nvars // 3.0

Validate model against labeled samples
 Predict outcomes for unlabeled samples
 Predict outcomes for manually entered predictor values
 Download model to make predictions off-line

Figure 53 Select how to apply the model

4. Select labeled dataset

- Select the Dataset, **Parkinson's_test**.
- Click **APPLY MODEL**.



Selected analysis:
Parkinson's_training_typical

Selected model:
Regression Random Forests training 500 trees with Mean Squared Error splitting criterion, minimum leaf size = 5, and variables to split = nvars // 3.0

Name	Size	Samples	Features	Created
<input checked="" type="checkbox"/> Parkinson's_test	240113	1762	20	2020-06-15
<input type="checkbox"/> Parkinson's_trai...	560348	4113	20	2020-06-15
<input type="checkbox"/> Parkinson's_dis...	800268	5875	20	2020-03-08

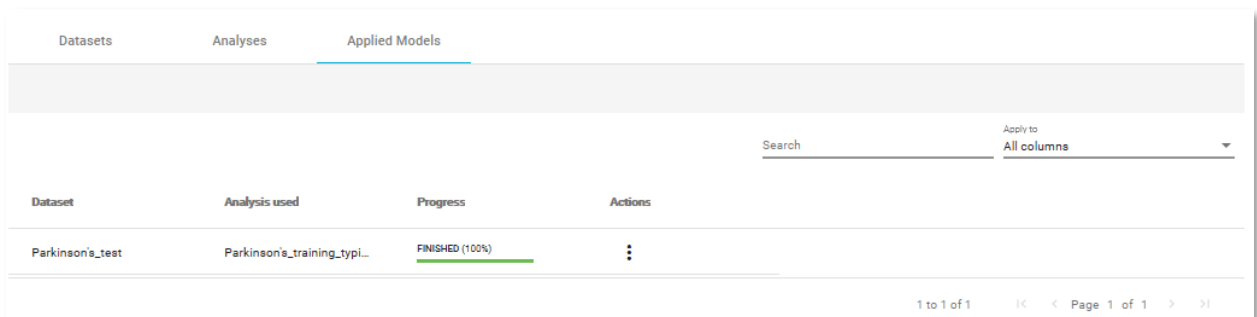
1 to 3 of 3

Click Apply Model to validate the model against the labeled dataset Parkinson's_test.

[BACK](#) [APPLY MODEL](#)

Figure 54 Apply model

JADBio will automatically bring you to the **Applied Models** window, where you can view the results of your applied model on the test data set.



Dataset	Analysis used	Progress	Actions
Parkinson's_test	Parkinson's_training_typi...	FINISHED (100%)	⋮

1 to 1 of 1 | Page 1 of 1

Figure 55 Applied Models

- Click on the **Actions** function, **View results** to open the results window.

Much of the results window is like the original training dataset window, but with some exceptions. In **VALIDATION ACTIONS** sidebar, there are buttons to **Download Predictions** of the samples and **Apply Model** to new samples.

In the main **Applied Model** window, JADBio displays the six predictors from the signature.

Now, the metrics include the results from the validation and the train dataset.

- Click on the **Analysis Visualization** tab.

Here, the **Residuals' plot** displays the segregation of the test data based on the model from the training data.

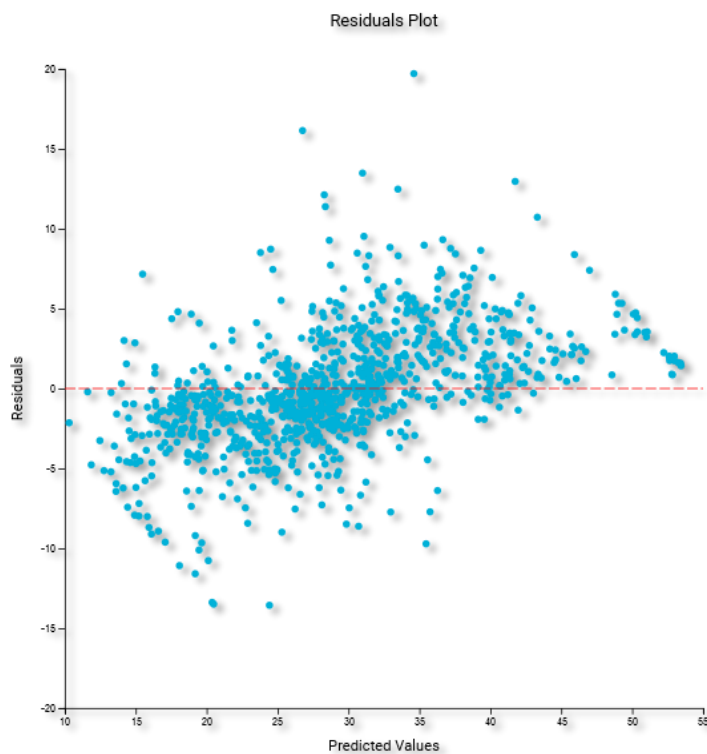


Figure 56 Residuals plot for applied model (The image used here is from the downloaded PNG, rather than a screen shot.)

Note of appreciation to JADBio users: We constantly make changes in the software and do our best to update these materials, but you may notice some differences. We welcome your feedback on how to make this more useful for you and requests for future tutorials.