

# Tutorial for JADBio Multi-Classification analysis – MultiClass cancer Case study

<b>INTRODUCTION .....</b>	<b>2</b>
<b>ACCESSING JADBio .....</b>	<b>3</b>
Try for free (Sign up).....	3
To Log In (Sign in) .....	3
<b>VIEWING DATA AND ANALYSES IN JADBio .....</b>	<b>4</b>
JADBio header .....	4
Menu .....	4
Dashboard.....	5
Projects.....	8
Datasets .....	10
Analyses.....	13
Applied Models .....	14
<b>WORKING IN JADBio .....</b>	<b>15</b>
<b>Creating a Project .....</b>	<b>15</b>
Uploading a Dataset.....	16
Background .....	20
<b>Analysis.....</b>	<b>21</b>
Analysis Results.....	24
Best Performing Model.....	26
Performance Overview .....	28
Feature Selection .....	30
Analysis Visualization .....	34

## 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

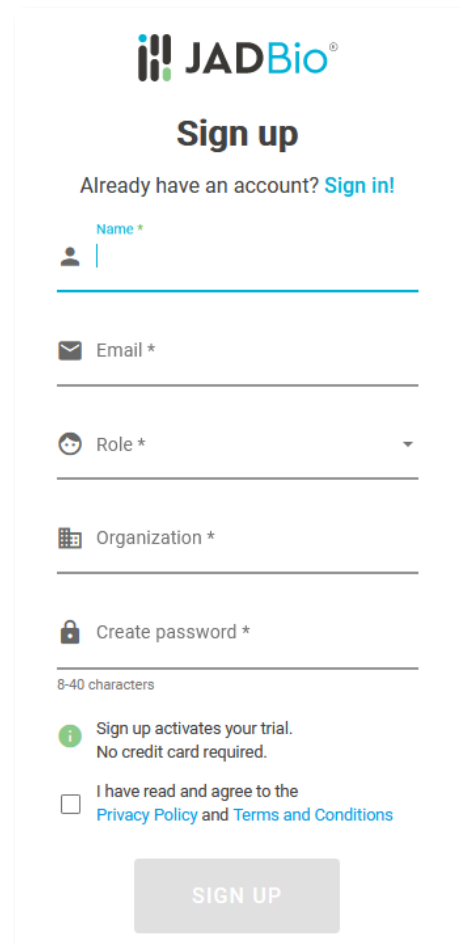
- In your browser of choice, navigate to [JADBio.com](https://JADBio.com).
- Click on the **LOG IN** button in top right corner if you already have an account or the **TRY FOR FREE** button if you don't.

### 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](mailto: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 input fields: "Name \*" with a person icon, "Email \*" with an envelope icon, "Role \*" with a person icon and a dropdown arrow, and "Organization \*" with a calendar icon. Below these is a "Create password \*" field with a lock icon and a note "8-40 characters". There is an information icon (i) next to the text "Sign up activates your trial. No credit card required." At the bottom, there is a checkbox next to the text "I have read and agree to the [Privacy Policy](#) and [Terms and Conditions](#)". A "SIGN UP" button is located at the bottom right of the form.

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

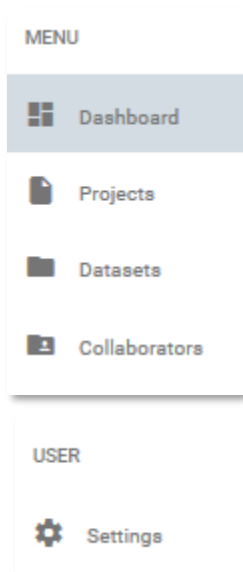
The **JADBio header** contains:

- 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 sidebar

## 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.

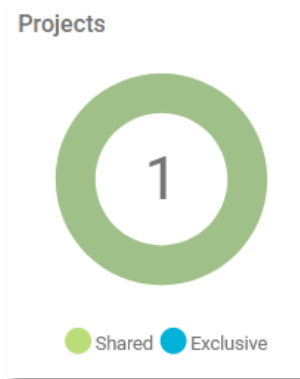


Figure 4 Projects

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

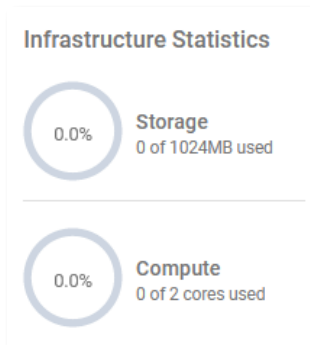


Figure 5 Infrastructure Statistics – the number of available cores is dependent on your subscription

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.

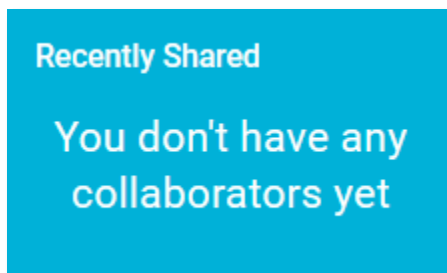


Figure 8 Recently Shared

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.

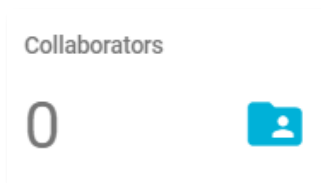


Figure 10 Collaborators

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

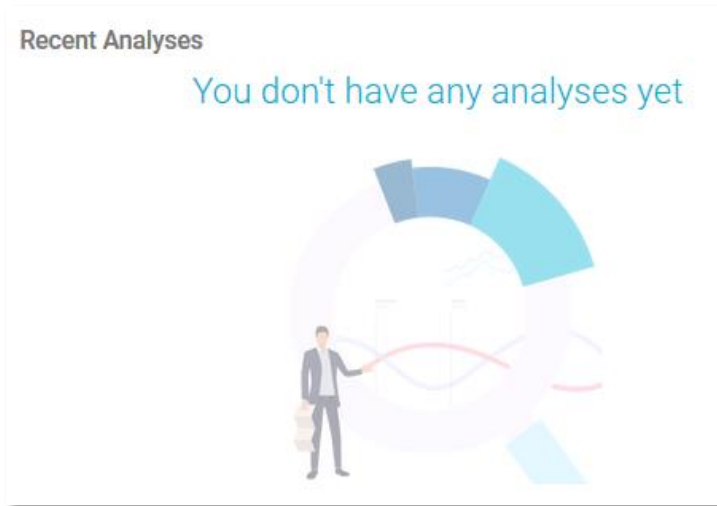


Figure 11 Recent Analyses

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



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

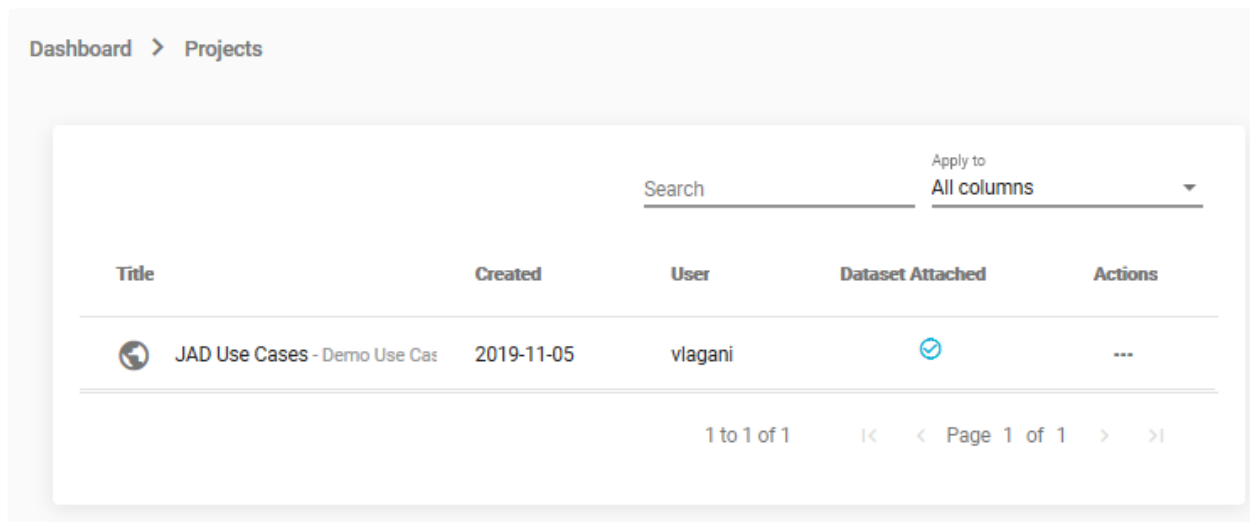
Figure 12 Recent Projects

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 contains several datasets in order for you to have some examples to work with in JADBio. You cannot change the data in this 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. **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**.

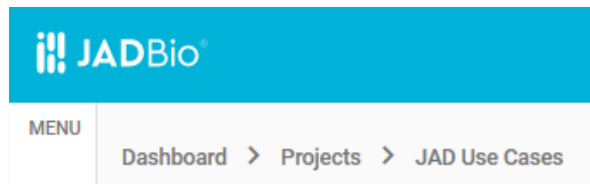


Figure 14 Breadcrumbs, JAD Use cases

2. In the top left, in 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**.

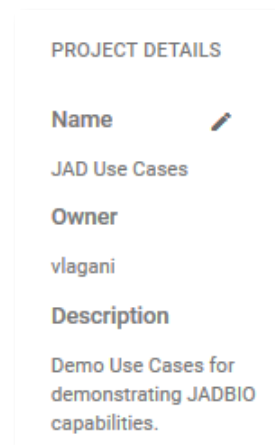


Figure 15 Project info

## 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	...		

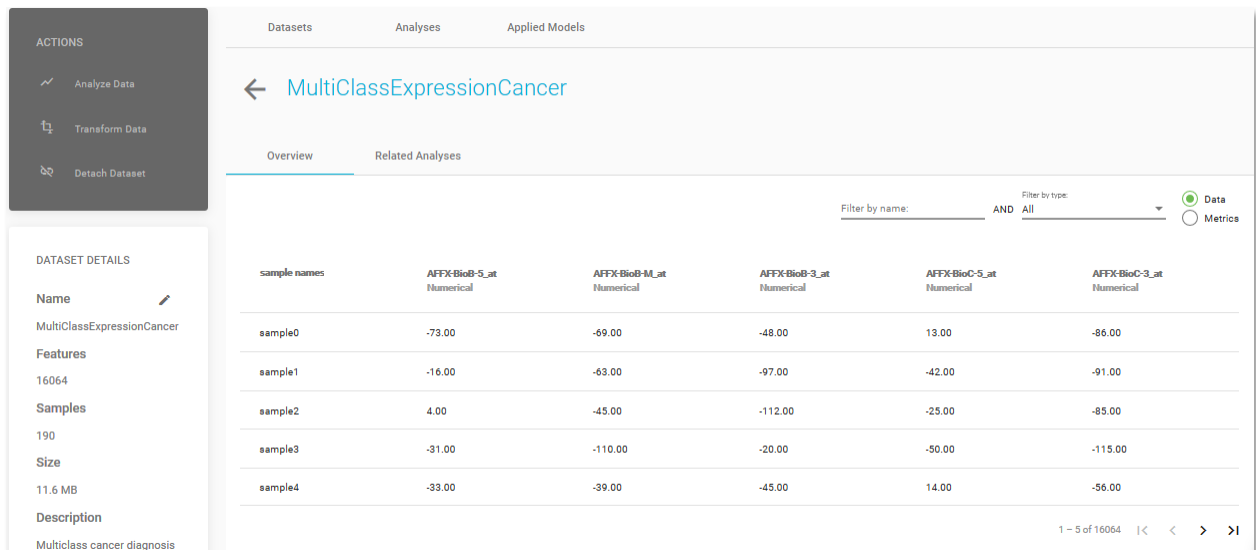
Figure 16 JAD Use Cases, Datasets

- Find the **MultiClassExpressionCancer** dataset.
- Click on the three dots, to open the **Actions** menu associated with the **MultiClassExpressionCancer** 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	AFFX-BioB-5_at Numerical	AFFX-BioB-M_at Numerical	AFFX-BioB-3_at Numerical	AFFX-BioC-5_at Numerical	AFFX-BioC-3_at Numerical
sample0	-73.00	-69.00	-48.00	13.00	-86.00
sample1	-16.00	-63.00	-97.00	-42.00	-91.00
sample2	4.00	-45.00	-112.00	-25.00	-85.00
sample3	-31.00	-110.00	-20.00	-50.00	-115.00
sample4	-33.00	-39.00	-45.00	14.00	-56.00

Figure 17 Preview dataset, Overview

- Select the **Metrics** radio button to view graphical and numerical summaries of each feature in the dataset (i.e. histograms).
- Click on **Expand** on class **feature**.

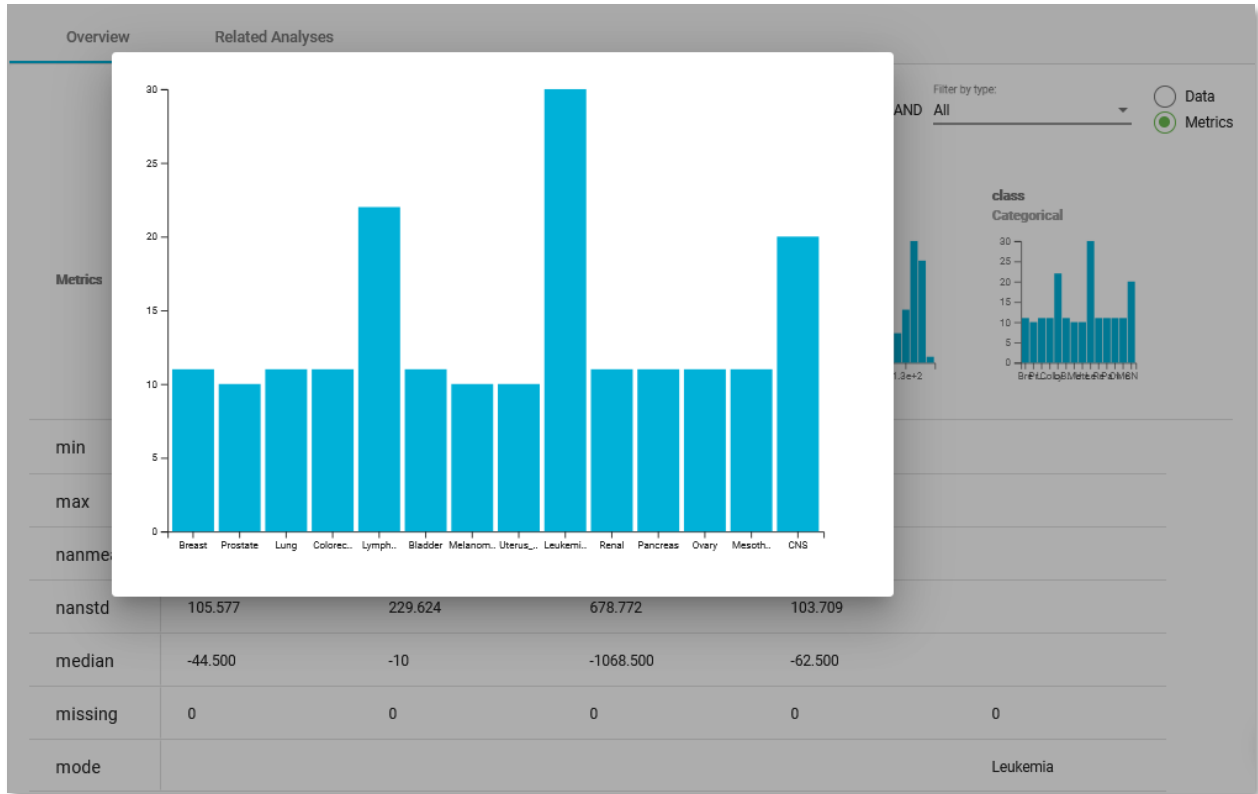


Figure 18 Metrics view

In the **Metrics** view, for the class feature, one can see that this feature includes 14 categories/cancer types.

Under **Actions**, JADBio gives you the option to **Perform Analysis** with this 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_transcr	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_diseas	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_diseas	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**, 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:

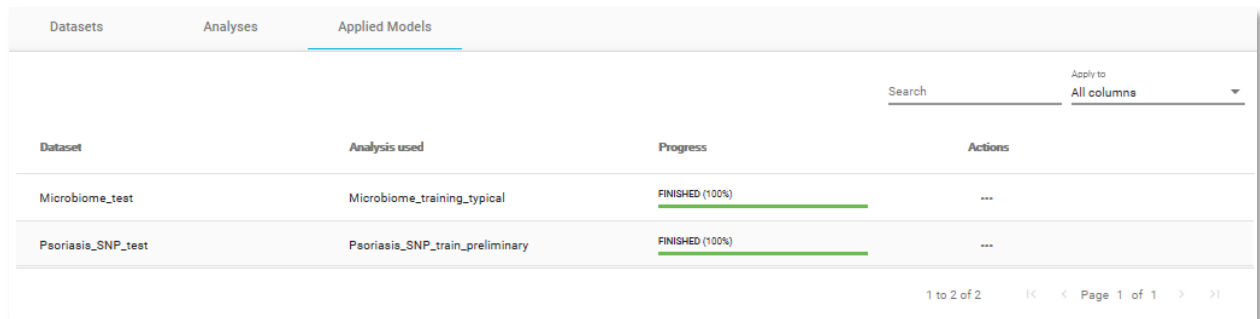
- i. Name
- ii. Started (date)
- iii. Dataset
- iv. Metric
- v. Metric Value
- vi. Outcome
- vii. Type
- viii. Progress
- ix. Actions, includes: View Results, Apply Model and Remove Analysis buttons.

## 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 and Remove Applied Model buttons.



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

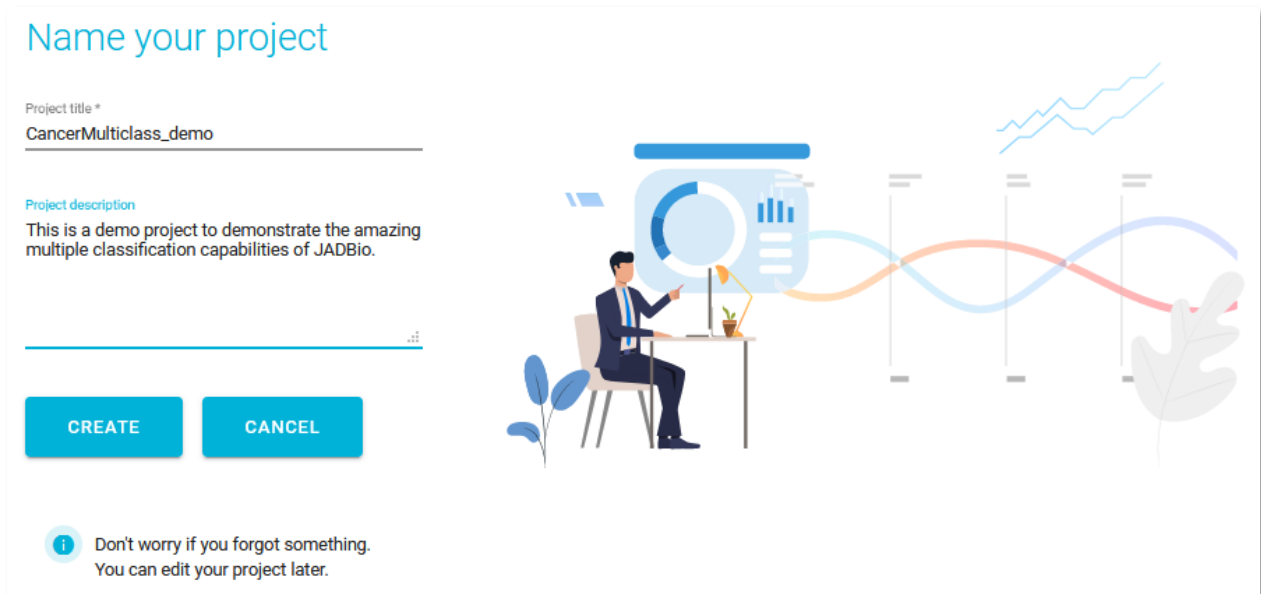
### Creating a Project

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** and then click on **CREATE** button.



**Name your project**

Project title \*  
CancerMulticlass\_demo

Project description  
This is a demo project to demonstrate the amazing multiple classification capabilities of JADBio.

**CREATE** **CANCEL**

**i** Don't worry if you forgot something.  
You can edit your project later.

Figure 21 Create a new project

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.

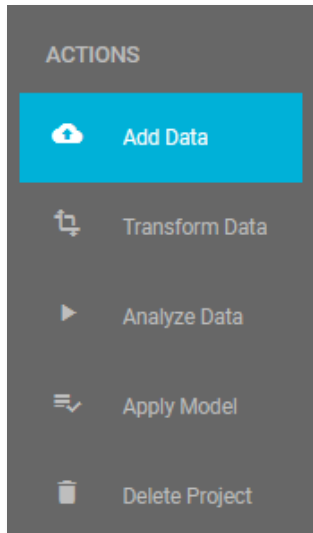


Figure 22 ACTIONS options in project window, prior to the upload of any datasets.

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 **MultiClassExpressionCancer** dataset from the **JAD Use cases** project.

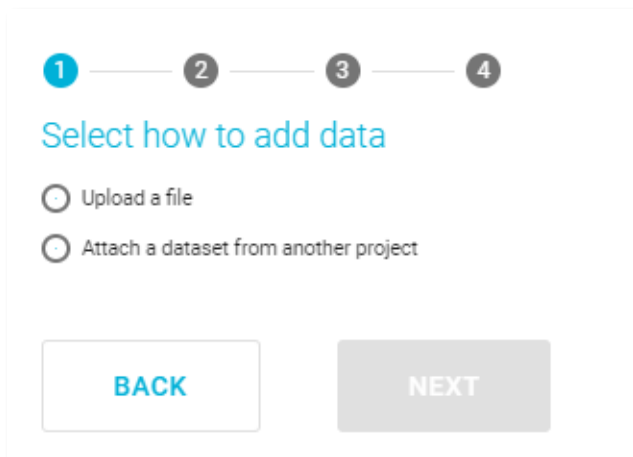


Figure 23 Add Data, select how to add data

- 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 **MultiClassExpressionCancer** dataset. (Note: You may have to search for it or scroll down to see it.)

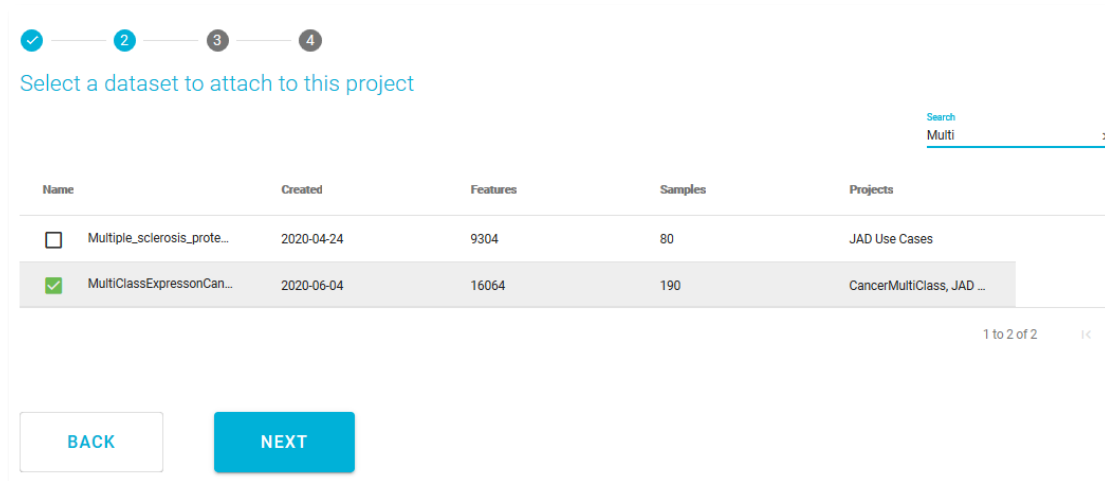
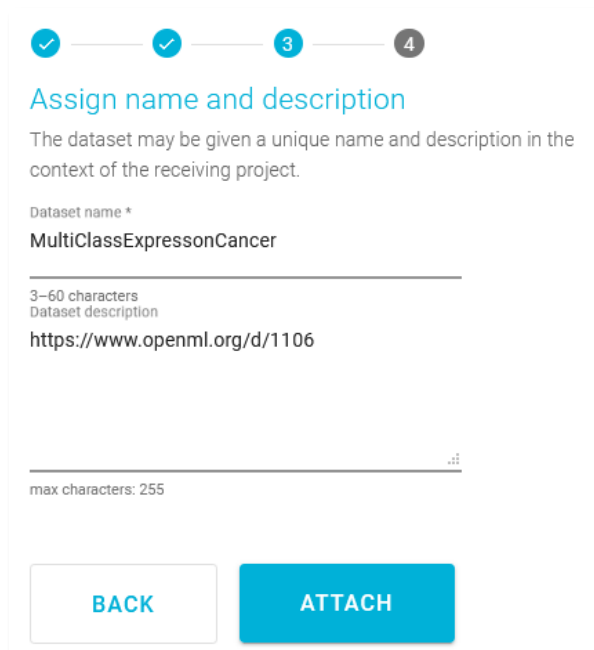


Figure 24 Select a dataset to attach

- 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.



Assign name and description

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

Dataset name \*

MultiClassExpressonCancer

3-60 characters

Dataset description

https://www.openml.org/d/1106

max characters: 255

BACK ATTACH

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.

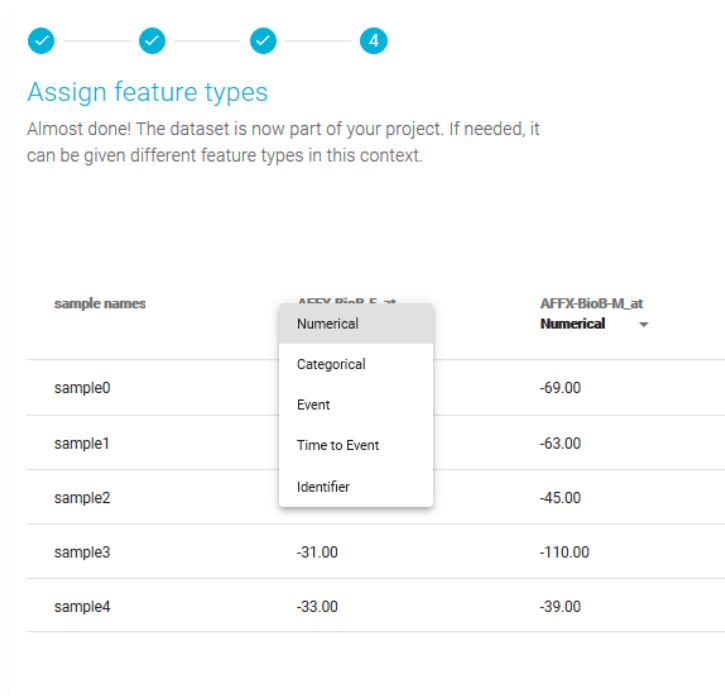


Figure 26 Assign feature types

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

You are now in the **Datasets** window of the **CancerMultiClass\_demo** Project. You can see that your new dataset has 16064 Features and 190 Samples. Under the **Actions** header on the three dots, you have three options: **Preview Dataset**, **Perform Analysis** and **Detach Dataset** from project.

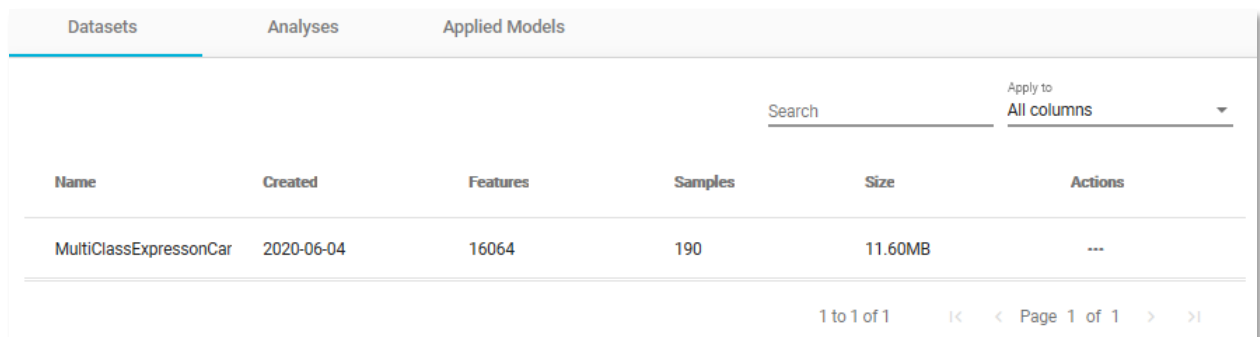


Figure 27 Uploaded dataset.

## Background

Nowadays, cancers are classified in two ways: by the type of tissue in which the cancer originates (histopathological information) and by primary site, or the location in the body where the cancer first developed. In addition, there is a wide spectrum in cancer morphology and many tumors are atypical or lack morphologic features that are useful for differential diagnosis. Molecular diagnostics could offer the promise of precise, objective, and systematic human cancer classification, but still these tests are not widely applied due to lack of characteristic molecular biomarkers for most solid tumors.


An accurate diagnostic test for multiclass cancer classification could be derived from the analysis of gene expression biomarkers. In this demonstration, you will train a model based on the expression levels of 190 tumor samples representing 14 common human cancer classes from Ramaswamy et al publication<sup>1</sup>. JADBio, applied on the same data, produces accurate diagnostic models that are on par with those presented in the original publication. You will see that with JADBio you will produce accurate predictive models that are on par with the ones presented in the original publication, regardless of your expertise in advanced machine learning techniques

<sup>1</sup>Original publication: Ramaswamy et al., PNAS December 18, 2001 98 (26) 15149-15154

## Analysis

With this dataset within JADBio you are able to train a predictive signature based on the gene expression profiles from tumor samples, spanning 14 common tumor types.

- From the table of **Datasets**, in the **MultiClassExpressonCancer** dataset click on **Perform Analysis**.

Name	Created	Features	Samples	Size	Actions
MultiClassExpressonCar	2020-06-04	16064	190	11.60MB	  



1 to 1 of 1   **Perform Analysis** 

Figure 28 MultiClassExpression dataset, Perform Analysis

- Select outcome: class**, 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:  x

sample names	class <input checked="" type="checkbox"/> Categorical
sample0	Breast
sample1	Breast
sample2	Breast
sample3	Breast
sample4	Breast

Some outcome classes are observed in less than 15 samples. Analysis results may not be reliable.  
Click Next to perform a classification analysis, finding a model that predicts class.

Figure 29 Select outcome - predicted feature

Note: Because you have chosen an outcome in which the values belong to different classes (categorical feature), JADBio will create a model based on a multi-class classification analysis.

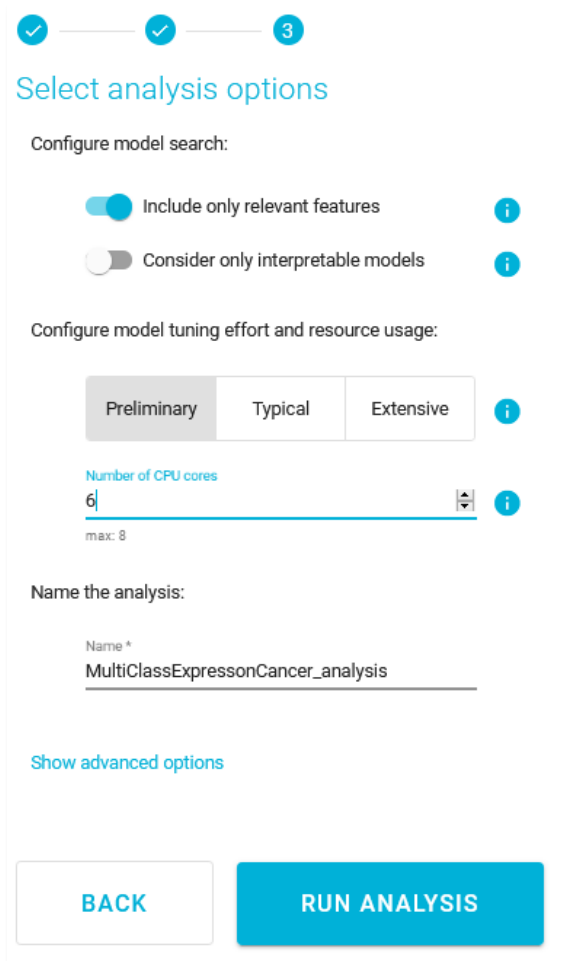


Figure 30 Select analysis options

- Set the predictive **analysis options** as follows:
  - Include only relevant features - **✓**
  - Tuning Effort: **Preliminary**, for a quick first assessment
  - Number of CPU cores: **6** (if you have a license with at least 6 cores)

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

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 AUC (area under the ROC curve).
- Maximum signature size is 25 features.
- Maximum multiple signatures to visualize is 5.
- JADBio will create four plots, PCA, UMAP, ICE, and Probabilities.

- 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			
		Search		Apply to		Columns to display		
				All columns		Name, Started, Dataset, M...		
Name	Started	Dataset	Metric	Metric Value	Outcome	Type	Progress	Actions
 MultiClassExpressionCar	2020-12-02 23:05	MultiClassExpressionC	AUC		class		<b>RUNNING (5%)</b>	...

Figure 31 Running analysis

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

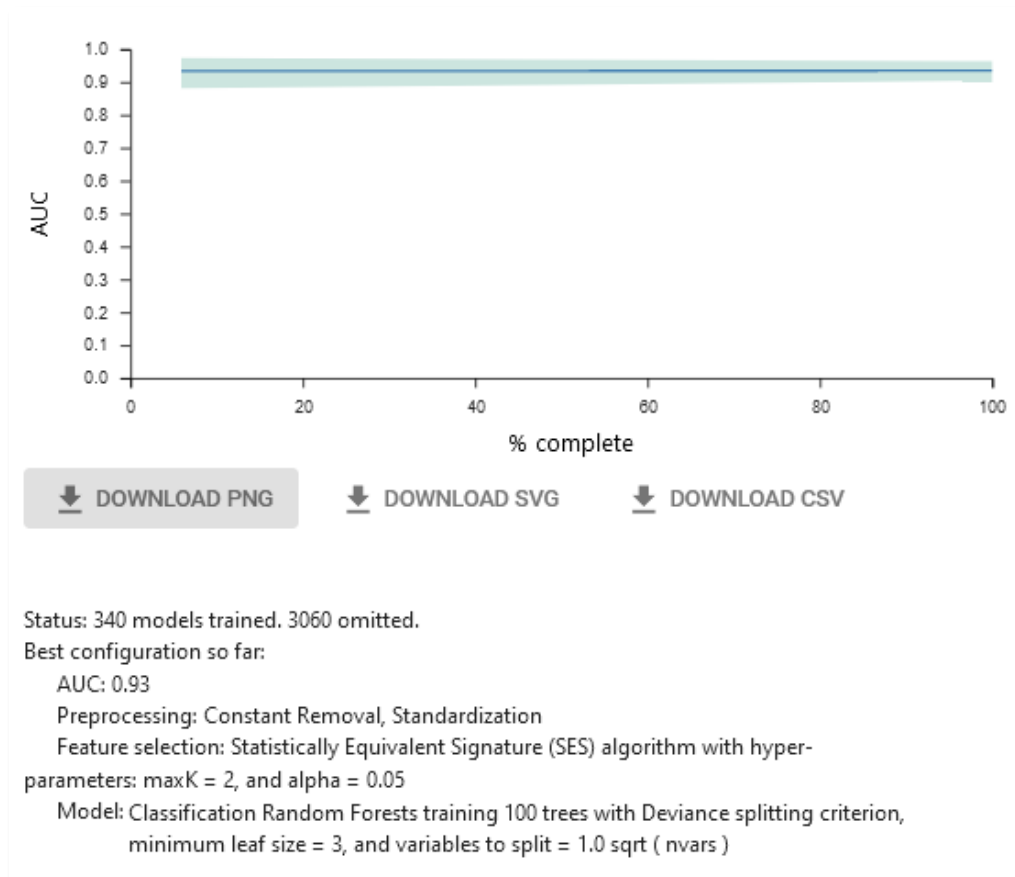


Figure 32 Learning progress

- As soon as your analysis is completed, in the **Actions** column, click on **View results**.

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.

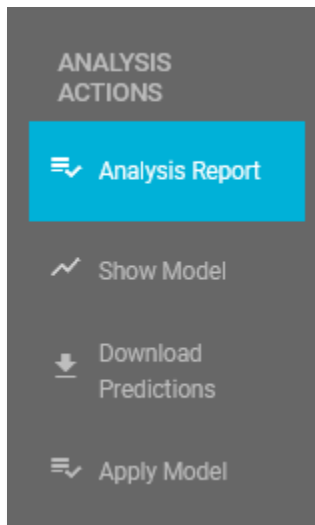


Figure 33 ANALYSIS ACTIONS sidebar

- Click on **Analysis Report**.

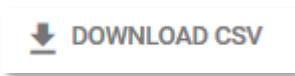
Here, JADBio provides a summary audit of the analysis including: **Dataset, Outcome to Predict, Analysis Type, etc.** and a description of the selected options as well as the **JADBio version**.




Summary	
Dataset	<a href="#">MultiClassExpressonCancer</a>
Outcome to Predict	class
Analysis Type	Classification
Metric	AUC
Grouping Feature	None
Relevant Features Only	Yes
Interpretable Models Only	No
Tuning Effort	Preliminary
CPU Cores	6
Configurations Tried	17
Models Trained	340 models trained. 3060 omitted.
Execution Time	19:30:08
JADBio Version	1.1.153

Figure 34 Analysis info

JADBIO also presents a list of all of the configurations that were tested in order to produce the Model and selected features. The list is not long in this example, but more complex analyses may test 100s of configurations. Specifically, for this analysis 17 configurations were tested which you

can download by clicking on  button.

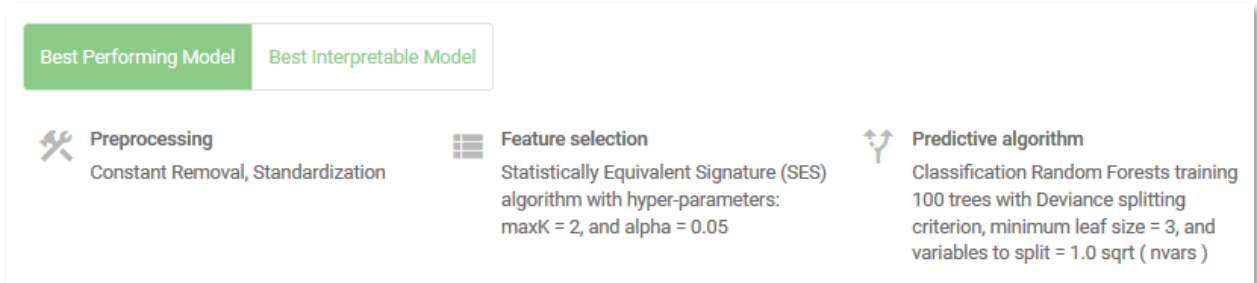
- Click on  return button to return to the main analysis results page.

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**.

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



Best Performing Model | Best Interpretable Model

- Preprocessing**  
Constant Removal, Standardization
- Feature selection**  
Statistically Equivalent Signature (SES) algorithm with hyper-parameters: maxK = 2, and alpha = 0.05
- Predictive algorithm**  
Classification Random Forests training 100 trees with Deviance splitting criterion, minimum leaf size = 3, and variables to split = 1.0 sqrt ( nvars )

Figure 35 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**.

- Click on **Best Interpretable Model**.



Best Performing Model | Best Interpretable Model

- Preprocessing**  
Constant Removal, Standardization
- Feature selection**  
Statistically Equivalent Signature (SES) algorithm with hyper-parameters: maxK = 2, and alpha = 0.05
- Predictive algorithm**  
Classification Decision Tree with Deviance splitting criterion and hyper-parameters: minimum leaf size = 3, and pruning parameter alpha = 0.05

Figure 36 Best Interpretable Model

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

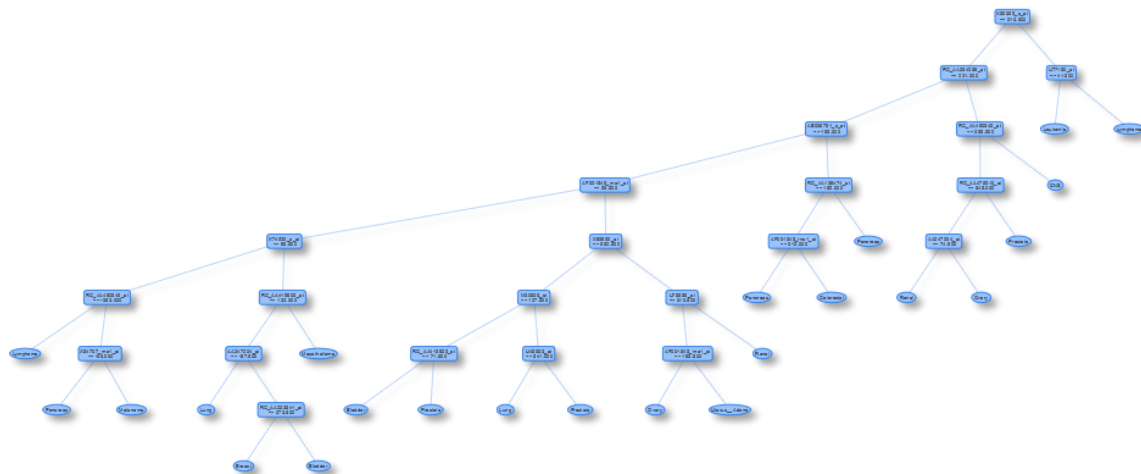


Figure 37 Show model (The image used here is from the downloaded SVG, rather than a screen shot).

JADBio displays a classification decision tree with multiple nodes and multiple arcs. In this decision tree, each arc ends at one of the 14 cancer types, but there are some cancer types, such as bladder, that are represented at the terminal of more than one arc.

- Close the decision tree, by clicking outside of the decision tree window.
- Now, click on the **Best Performing Model** button to return to this model.

Note that the Best Performing Model's predictive algorithm is Classification Random Forests, which is not easily represented in a way that a human can follow the model's logic. This is why JADBio also produces the Decision Tree.

- Click on the **Download Predictions**.

The downloaded **analysis\_predictions.txt** table is best viewed in a spread sheet. In it, JADBio describes each of the analyzed samples and two new columns, based on the cross validation of the best configuration. The new columns are **Prediction** and **Difficult to Predict**. The Difficult to Predict column is rated **TRUE** or **FALSE** based on the likelihood that the predicted outcome is correct. Also, for each sample you will see the effect of the **probability** the sample would be predicted as Bladder cancer type and 13 more cancer types.

	A	B	C	D	E	F	G
1	Sample name	Prob ( class = Bladder )	Prob ( class = Breast )	Prob ( class = CNS )	Prob ( class = Colorectal )	Prob ( class = Leukemia )	Prob ( class = Lung )
2	sample0	0,2458	0,1634	0,0068	0,0030	0,0081	0,0886
3	sample1	0,1618	0,2413	0,0096	0,0273	0,0093	0,1435
4	sample2	0,1305	0,2845	0,0013	0,0495	0,0104	0,0981
5	sample3	0,1287	0,2785	0,0027	0,0240	0,0010	0,1591
6	sample4	0,0528	0,1195	0,0013	0,0493	0,0175	0,0557
7	sample5	0,1415	0,2185	0,0000	0,0502	0,0000	0,1155
8	sample6	0,0943	0,1249	0,0059	0,0700	0,0058	0,1052
9	sample7	0,1371	0,4173	0,0017	0,0033	0,0033	0,0590
10	sample8	0,0257	0,0286	0,0043	0,0770	0,0237	0,2038
11	sample9	0,2317	0,1082	0,0106	0,0849	0,0113	0,0842
12	sample10	0,0201	0,0053	0,1843	0,0189	0,0235	0,0967
13	sample11	0,0412	0,0017	0,3368	0,0233	0,0495	0,0435
14	sample12	0,0067	0,0108	0,0671	0,0063	0,0050	0,0226
15	sample13	0,0250	0,0299	0,0364	0,0443	0,0062	0,0400
16	sample14	0,0064	0,0010	0,0967	0,0333	0,0267	0,0528
17	sample15	0,0091	0,0068	0,0320	0,0536	0,0132	0,0273
18	sample16	0,0071	0,0008	0,0158	0,0661	0,0312	0,2197
19	sample17	0,1702	0,1676	0,0096	0,0199	0,0027	0,1183
20	sample18	0,1076	0,1066	0,0467	0,0078	0,0474	0,1336

Figure 38 Downloaded predictions viewed in spreadsheet

## Performance Overview

The performance of a classifier (i.e. for Colorectal cancer vs all other cancers) is described by the Area Under the Curve (AUC) of the ROC curve, where the true positive rate (sensitivity) is plotted against the false positive rate (1-specificity). The best ROC curves are the ones where X (false positive rate) = 0 and Y (true positive rate) = 1.

In this case, AUC is 0.934 [0.903, 0.963] indicating that the gene expression profiles carry useful information in order to predict different types of cancer.

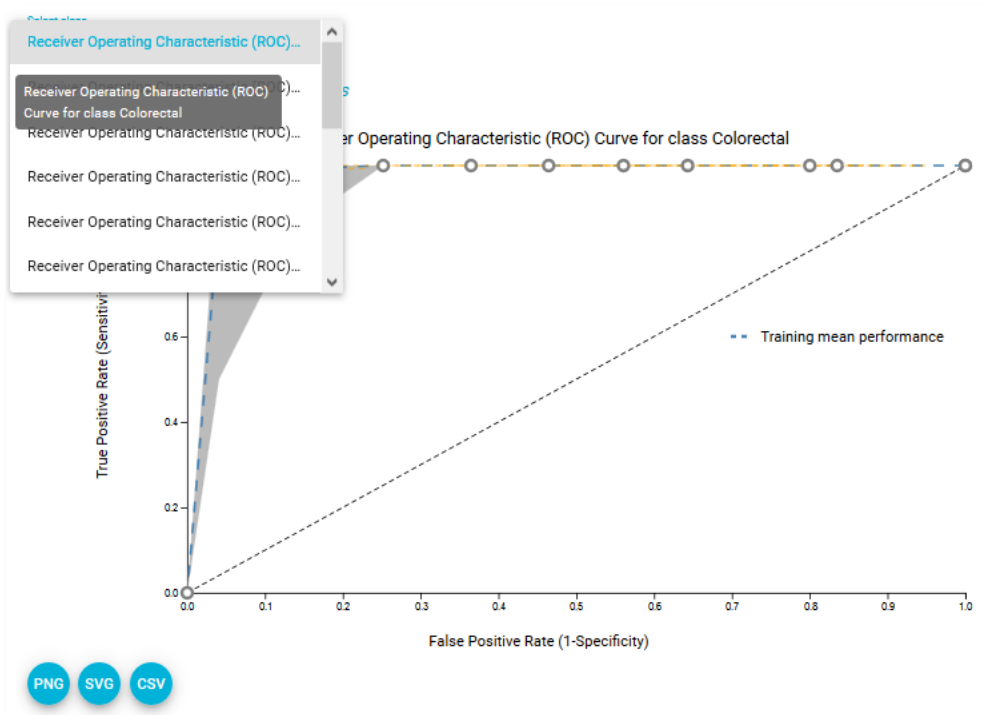


Figure 39 Receiver operating characteristic (ROC) curve plot for class Colorectal cancer

- Click on the position (0.98, 0.14) on the ROC curve. This will open a small window that displays the predictive performance at the selected threshold (0.06).

False Positive Rate: 0.14[0.09, 0.20]  
 True Positive Rate: 0.98[0.81,1.00]  
 Threshold: 0.06

		Predicted Class	
		No	Yes
Observed Class	No	0.81 [0.75, 0.87]	0.13 [0.08, 0.19]
	Yes	0.00 [0.00, 0.01]	0.06 [0.03, 0.09]

**\* Average [2.5% low confidence, 97.5% upper confidence]**

Figure 40 Predictive performance for the selected threshold

Below the ROC Curve you will find many different metrics based on your **Best Performing Model**.

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

- Hover over any “i” adjacent to a metric for an explanation of the score.
- Click on the Area under the ROC Curve metric to display confidence intervals and distribution graph.

How you set the thresholds will be determine the overall sensitivity and specificity for the classifier.

### 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.

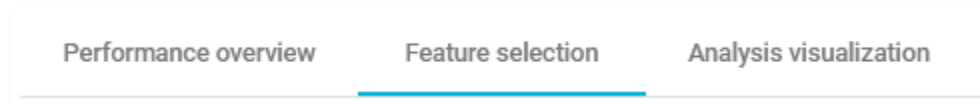


Figure 41 Feature selection

### 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, one signature with 25 features was selected.

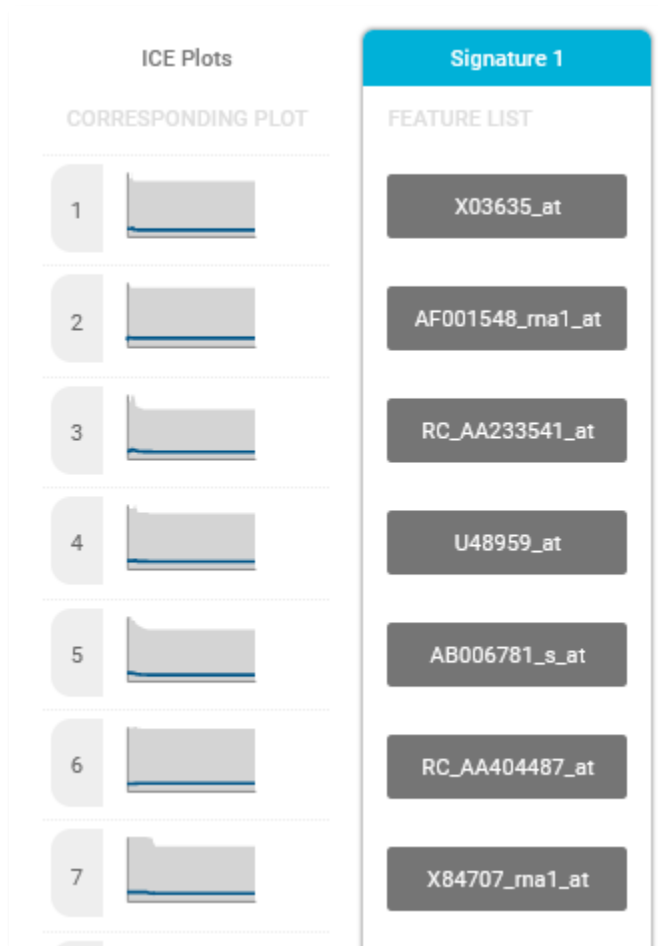


Figure 42 Part of the selected signature

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 the X03635\_at feature to enlarge the ICE plot.
- Use the pulldown to select the **Class Colorectal**.

This opens the ICE plot for the prediction of a Colorectal classification. In this plot, you can see that as the X03635\_at expression level increases, the probability of colorectal classification stays stable.

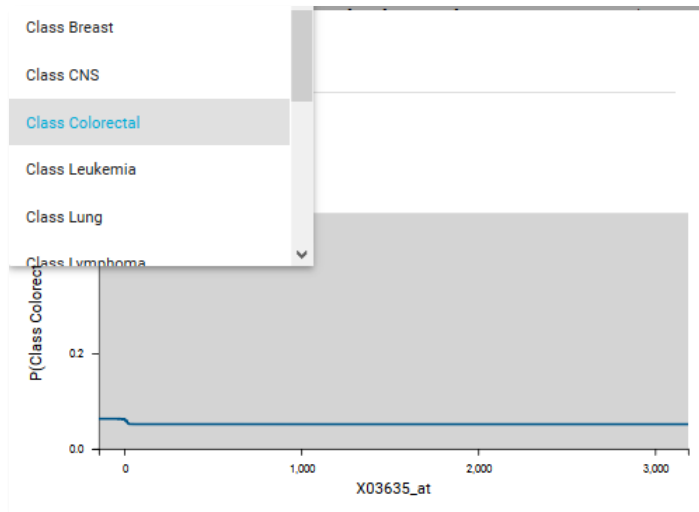


Figure 43 Individual Contribution Expectation plot for X03635\_at

### 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 to diagnose multiple cancers. 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, the user, 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 25 features of the selected signature and their relative performance.



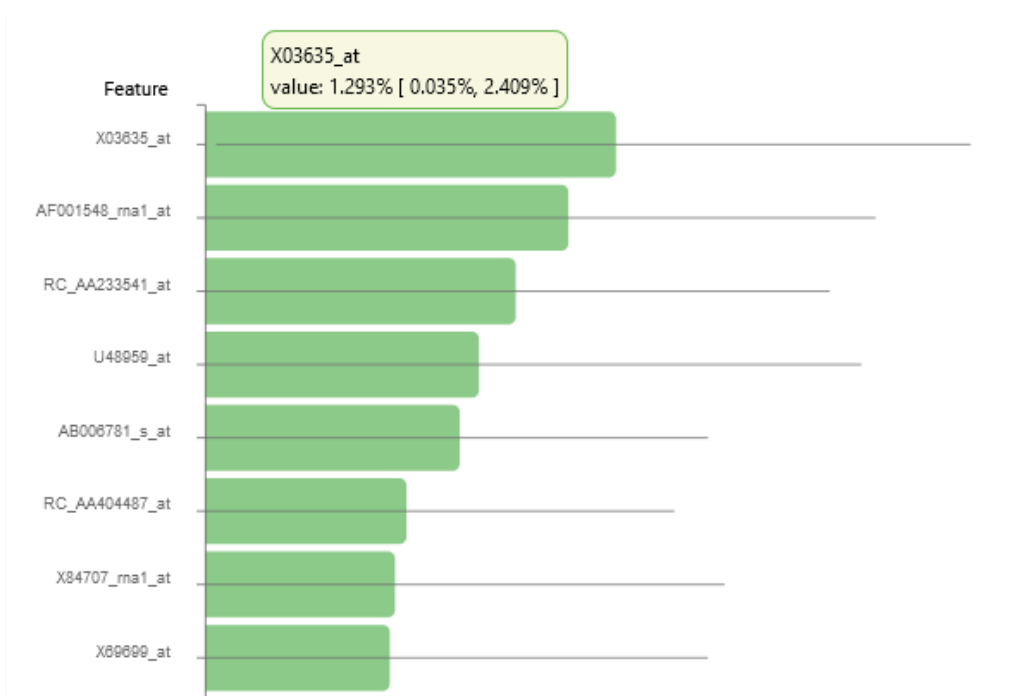


Figure 44 Part of the Feature Importance plot

## Analysis Visualization

- Select the **Analysis Visualization** tab.

**Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP)** is a dimension reduction technique that attempts to learn the high-dimensional manifold on which the original data lays, and then maps it down to two dimensions. UMAP is ideal for non-linear models.

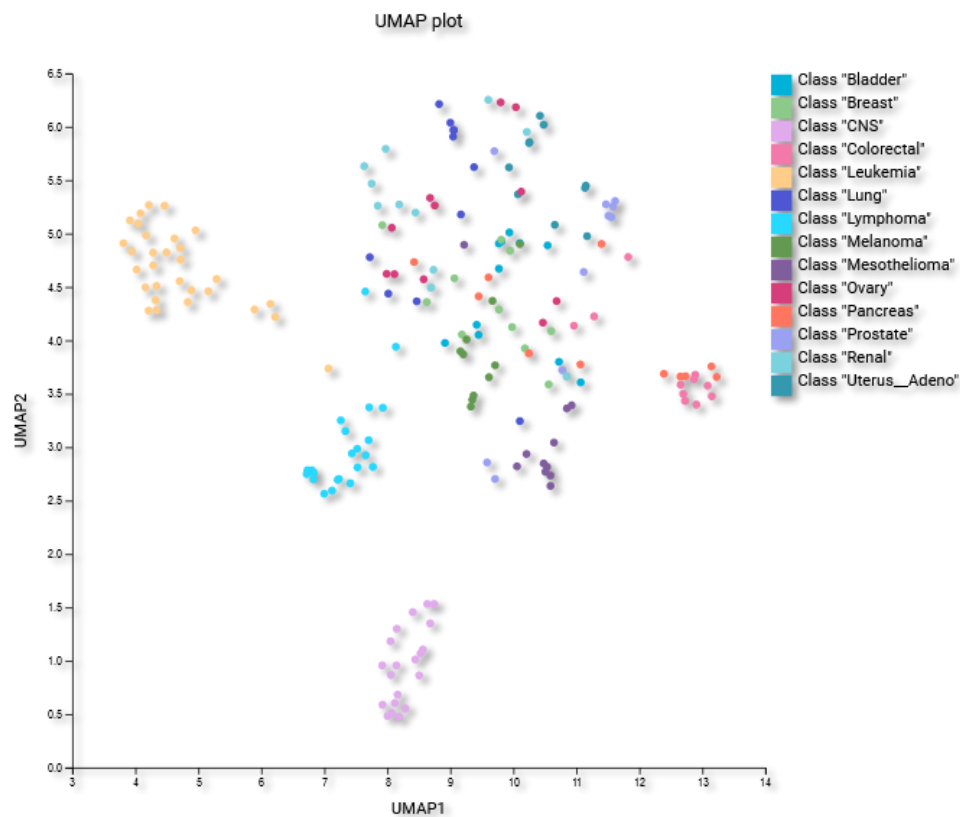


Figure 45 Uniform Manifold Approximation Projection. UMAP (The image used here is from the downloaded PNG, rather than a screen shot).

- Change the **Display as** selection from UMAP to PCA.

**Principal component analysis (PCA)** is another dimensionality reduction technique that seeks the linear combinations (Principal Components) of the original features, such that the derived features capture maximal variance. JADBio performs PCA with only the features included in the signature.

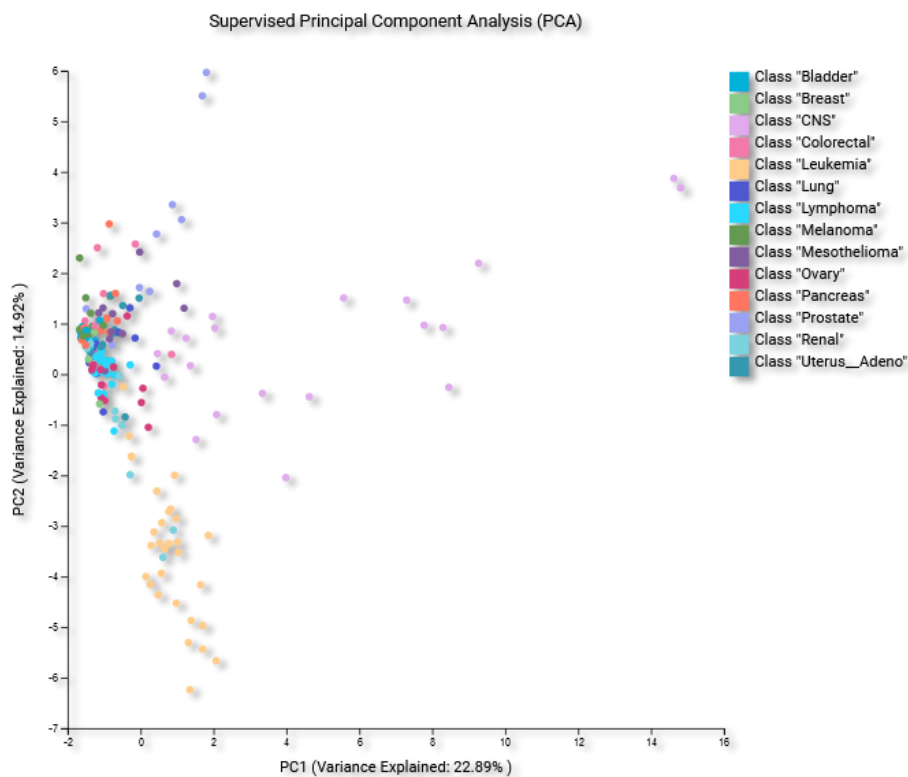


Figure 46 PCA plot

The appropriate visualizations for this model are UMAP and PCA plots, in which all cancers types predictions are visualized. Other types of analysis and other types of models would likely result in different visualizations. For instance, a survival (time to event) analysis would have resulted in a Kaplan-Meier curve.

- Scroll down to see the **Probabilities plot**.

The **Probabilities plot** shows the probability of the prediction resulting in the chosen class against all other classes. An ideal plot would have complete separation between the two classes. You can also visualize the probabilities in a box plot.

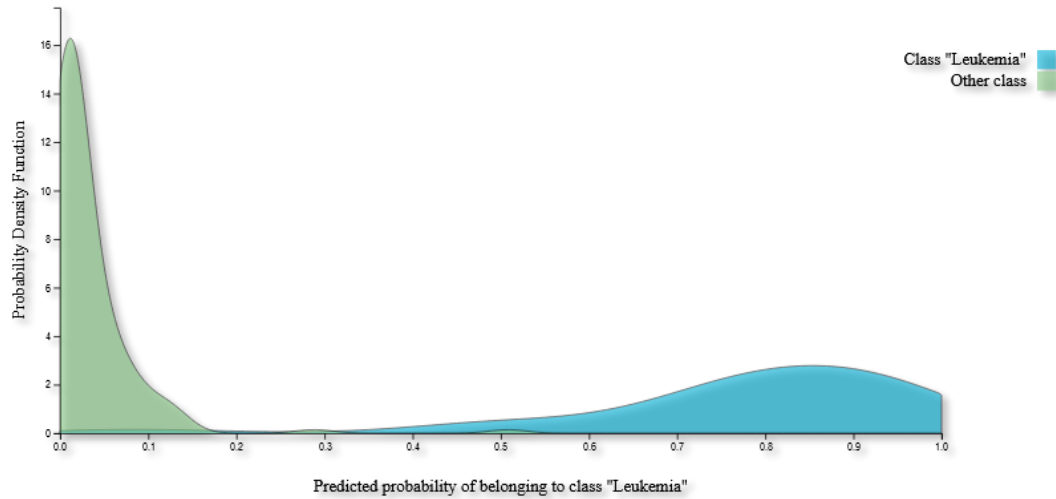


Figure 47 Probabilities plot for Class Leukemia against all other classes displaying as density plot.

*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.*