

How to validate models?

CLOVER MS Data Analysis Software

1 First steps

This guide is the same for MALDI and FTIR spectra

1. Choose the Prediction Model desired to validate on study view section.

2. Click on Validate.

3. Upload or select the files needed to validate and press Submit.

4. These files will be presented to the algorithm as Blind Data to be aligned with the prediction base.

5. Group the replicated spectra (if any) by name or per group selecting the number of replicates.

6. Add as many steps as necessary and press Next again.

7. Select the samples from each category to be validated on the model and press Next.

8. Then, you obtain the Prediction Result.

Create Categories → Create Peak Matrix → Perform the Classification Algorithm and save the Prediction Model

2 Validation

1. Choose the Prediction Model desired to validate on study view section.

2. Click on Validate.

3. Upload or select the files needed to validate and press Submit.

4. These files will be presented to the algorithm as Blind Data to be aligned with the prediction base.

5. Group the replicated spectra (if any) by name or per group selecting the number of replicates.

6. Add as many steps as necessary and press Next again.

2.1

4. Apply (or not) Prediction Model parameters to reduce noise on samples and press Next.

5. Group the replicated spectra (if any) by name or per group selecting the number of replicates.

6. Add as many steps as necessary and press Next again.

2.2

7. Select the samples from each category to be validated on the model and press Next.

8. Then, you obtain the Prediction Result.

Validation Analysis Metrics & Tools

Hyperplanes and Predicted Data (SVM Analysis)

ROC Curves

PRC Curves

Confusion Matrix

Actual / Predicted	Subspecies 1	Strain 1	Strain 2	% Correct
Subspecies 1	7	1	1	77.78 %
Strain 1	0	5	0	100 %
Strain 2	3	0	7	70 %

Score: 79.17 %

Balanced Accuracy: 82.59 %

Validation Result Table

Name	Control Category	Predicted Category (F1)	Probability (F1)
specie_a_subspecie_1_specie_a_subspecie_1_sample_0_F1	Subspecies 1	Subspecies 1	99.95 %
specie_a_subspecie_2_specie_a_subspecie_1_sample_0_F1	Subspecies 2	Subspecies 1	99.95 %
specie_a_subspecie_1_specie_a_subspecie_1_sample_(10)_0_E12	Subspecies 1	Subspecies 1	99.39 %
specie_a_subspecie_2_specie_a_subspecie_1_sample_(10)_0_E12	Subspecies 2	Subspecies 1	99.39 %
specie_a_subspecie_1_specie_a_subspecie_1_sample_(11)_0_B1	Subspecies 1	Subspecies 1	98.85 %
specie_a_subspecie_2_specie_a_subspecie_1_sample_(11)_0_B1	Subspecies 2	Subspecies 1	98.85 %
specie_a_subspecie_1_specie_a_subspecie_1_sample_(12)_0_A10	Subspecies 1	Subspecies 1	96.92 %
specie_a_subspecie_1_specie_a_subspecie_1_sample_(13)_0_H8	Subspecies 1	Subspecies 1	95.07 %

Distance Plot (Random Forest)

The Predicted Category is calculated by the algorithm meanwhile Control Category is the one uploaded before to perform validation. If both are the same category it is presented with a green tick.

Red dots correspond to the validation data uploaded.



Clover Bioanalytical Software Copyright© 2015

Edif. Centro de Empresas PTS. Av. del Conocimiento, 41 18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com

support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide

How to validate models?

CLOVER MS Data Analysis Software

1 First steps



This guide is the same for MALDI and FTIR spectra

Create Categories

Create Peak Matrix

Perform the Classification Algorithm and save the Prediction Model



cloverbiosoft.com
clovermsdataanalysis.com

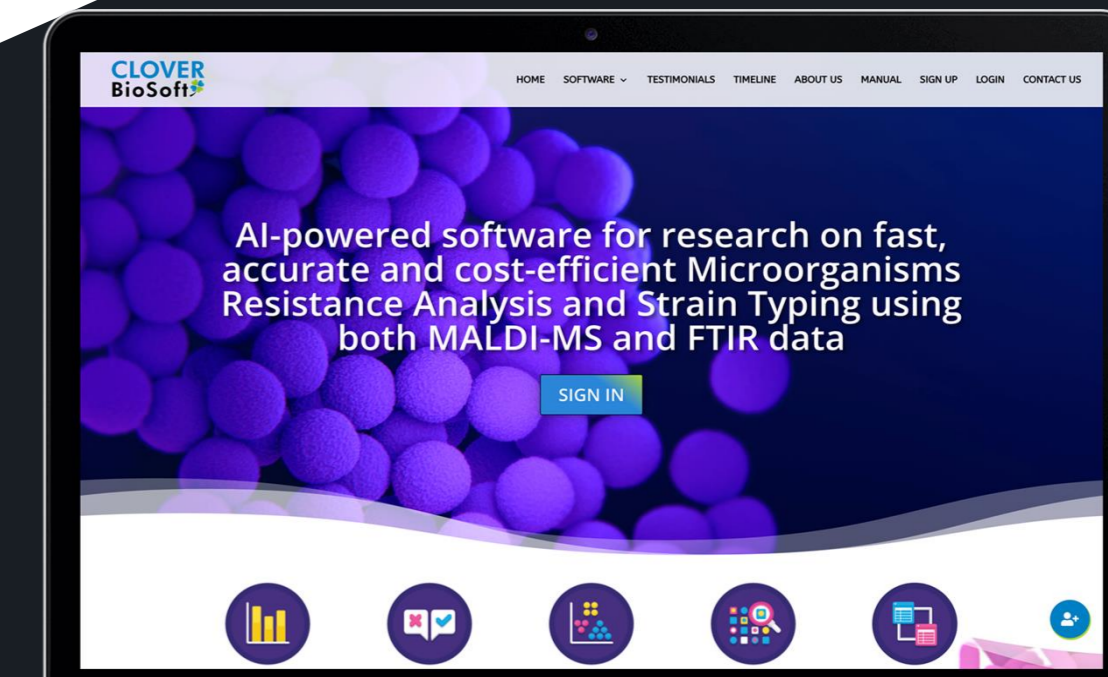
Clover Bioanalytical Software
Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com

support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide



2

Validation

The screenshots illustrate the validation workflow:

- Study View:** The 'Prediction Models' tab is selected in the breadcrumb navigation. A green box labeled '1' highlights the 'Prediction Models' tab.
- Prediction Model:** The 'Validate' button is highlighted with a green box labeled '2' and an arrow.
- Prediction Model Validation:** The 'Select files' button is highlighted with a green box labeled '3'. Below it, a list of folders is shown, including 'Validation (2 folders)' and 'Species A (2 folders)'. A green arrow points to the 'Submit' button.

1. Choose the **Prediction Model** desired to validate on study view section.

2. Click on **Validate**.

3. Upload or select the **files** needed to validate and press **Submit**.

i These files will be presented to the algorithm as **Blind Data** to be aligned with the prediction base.



cloverbiosoft.com
clovermsdataanalysis.com

Clover Bioanalytical Software
Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com

support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide



2.1

4. Apply (or not) Prediction Model parameters to **reduce noise on samples** and press **Next**.

Prediction Model Validation

Study Study 1 Prediction Model 4 - SVM

Prediction Model data

4

Study Select... Experiment Select... Prediction Select... Upload or select fil... Noise reduction Replicated spec... Category select... Prediction Res...

Apply Prediction Model parameters

Variance stabilization

Smoothing

Window length * 11 Polynomial order * 3

Baseline Subtraction

Tophat filter

Factor * 0,02

Preview

Next >

Select Spectrum for preview

specie_a_subspecie_1_specie_a_subspecie_1

specie_a_subspecie_1_specie_a_subspecie_1_sample_0_F1_1

Intensity

5. Group the **replicated spectra** (if any) by name or per group selecting the number of replicates.

6. Add as many steps as necessary and press **Next** again.

Prediction Model Validation

Study Study 1 Prediction Model 4 - SVM

Prediction Model data

Study Sele... Experiment Sel... Prediction Sel... Upload or sele... Noise redu... 6 Replicated spectra 7 Category sel... 8 Prediction ...

Please, group the replicated samples, if any. Remove the samples you don't want to include on the preprocessing:

Next >

X Reset all

Step 1 Add step 6

Auto replicates By sample name By replicates per group Replicates per group: 3 Apply 5

X Remove Set as replicated Select All Reset

Columns	Total samples: 72	
specie_a_subspecie_1_specie_a_subspecie_1_sample_0_F1	specie_a_subspecie_1_strain_(2)_spe... (10)_0_A11	specie_a_subspecie_2_strain_1_speci... e_(35)_0_G2
specie_a_subspecie_1_specie_a_subspecie_1_sample_0_F1	specie_a_subspecie_1_strain_(2)_spe... (10)_0_A11	specie_a_subspecie_2_strain_1_speci... e_(35)_0_G2
specie_a_subspecie_1_specie_a_subspecie_1_sample_0_F1	specie_a_subspecie_1_strain_(2)_spe... (10)_0_A11	specie_a_subspecie_2_strain_1_speci... e_(35)_0_G2
specie_a_subspecie_1_specie_a_subspecie_1_sample_0_F1	specie_a_subspecie_2_specie_a_subspecie_2_sample_0_F1	specie_a_subspecie_2_strain_(2)_spe... sample_0_C7



cloverbiosoft.com
clovermsdataanalysis.com

Clover Bioanalytical Software
Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com

support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide

How to validate models?

CLOVER MS Data Analysis Software

2.2

7. Select the **samples from each category** to be validated on the model and press **Next**.

7

Subspecies 1 Subspecies 2

Samples (+ Add)

specie_a_subspecie_2_specie_a_subspecie_1_sample_0_F1 Validation/Subspecies 2/Strain 2/specie_a_subspecie_2_specie_a_subspecie_1_sample_0_F1_1.csv

specie_a_subspecie_2_specie_a_subspecie_1_sample_(10)_0_E12 Validation/Subspecies 2/Strain 2/specie_a_subspecie_2_specie_a_subspecie_1_sample_(10)_0_E12_1.csv

specie_a_subspecie_2_specie...

specie_a_subspecie_2_strain_...

specie_a_subspecie_2_strain_...

8

Validation Result: 19/24 Export full table to CSV

Name	Control Category	Predicted Category (#1)	Probability (#1)
✓ specie_a_subspecie_1_specie_a_subspecie_1_sample_0_F1	Subspecies 1	Subspecies 1	99.95 %
✗ specie_a_subspecie_2_specie_a_subspecie_1_sample_0_F1	Subspecies 2	Subspecies 1	99.95 %
✓ specie_a_subspecie_1_specie_a_subspecie_1_sample_(10)_0_E12	Subspecies 1	Subspecies 1	99.39 %
✗ specie_a_subspecie_2_specie_a_subspecie_1_sample_(10)_0_E12	Subspecies 2	Subspecies 1	99.39 %

8. Then, you obtain the **Prediction Result**.



cloverbiosoft.com
clovermsdataanalysis.com

Clover Bioanalytical Software
Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

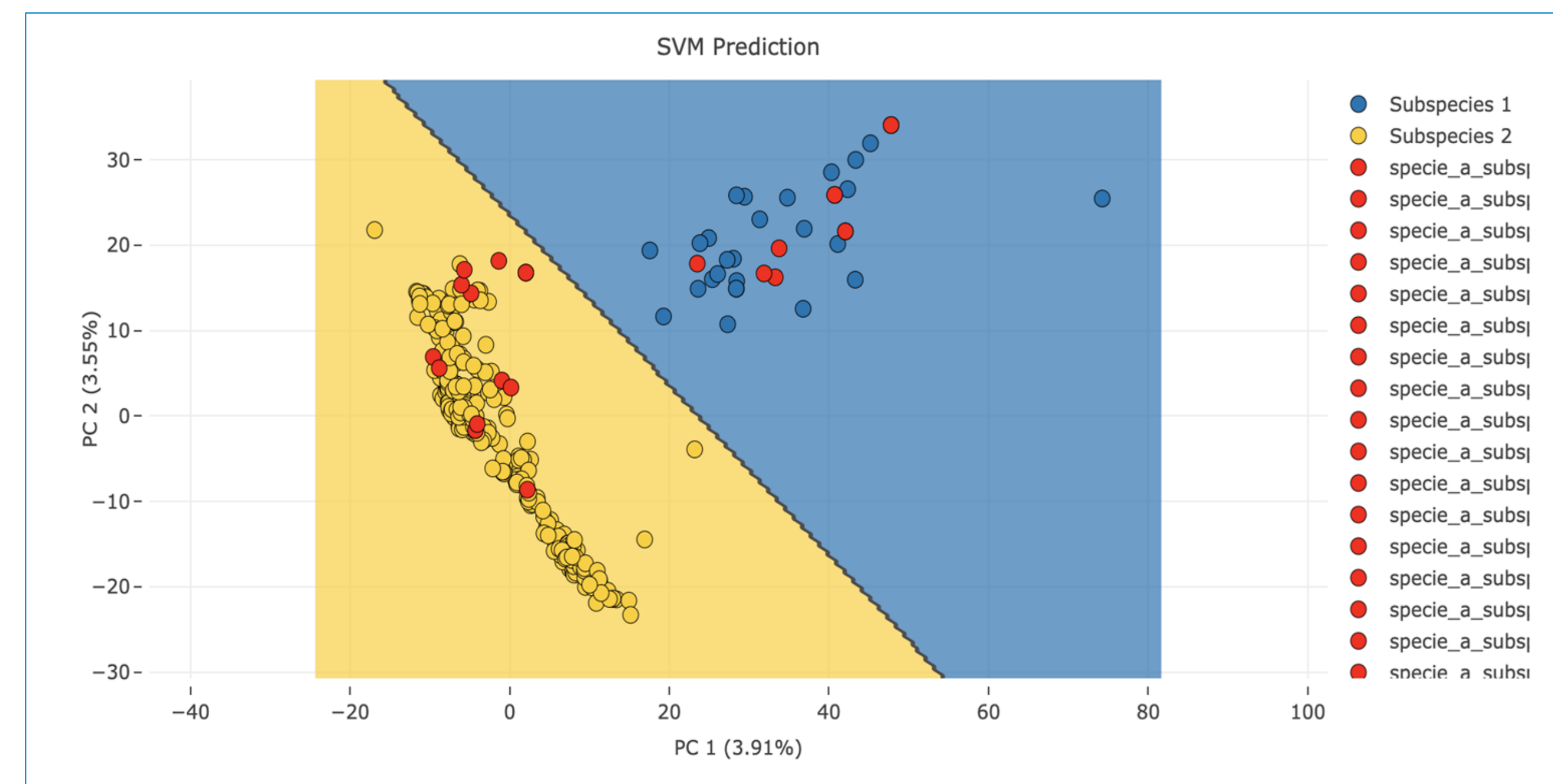
Phone: +34 958 991 543
email: info@cloverbiosoft.com

support@cloverbiosoft.com

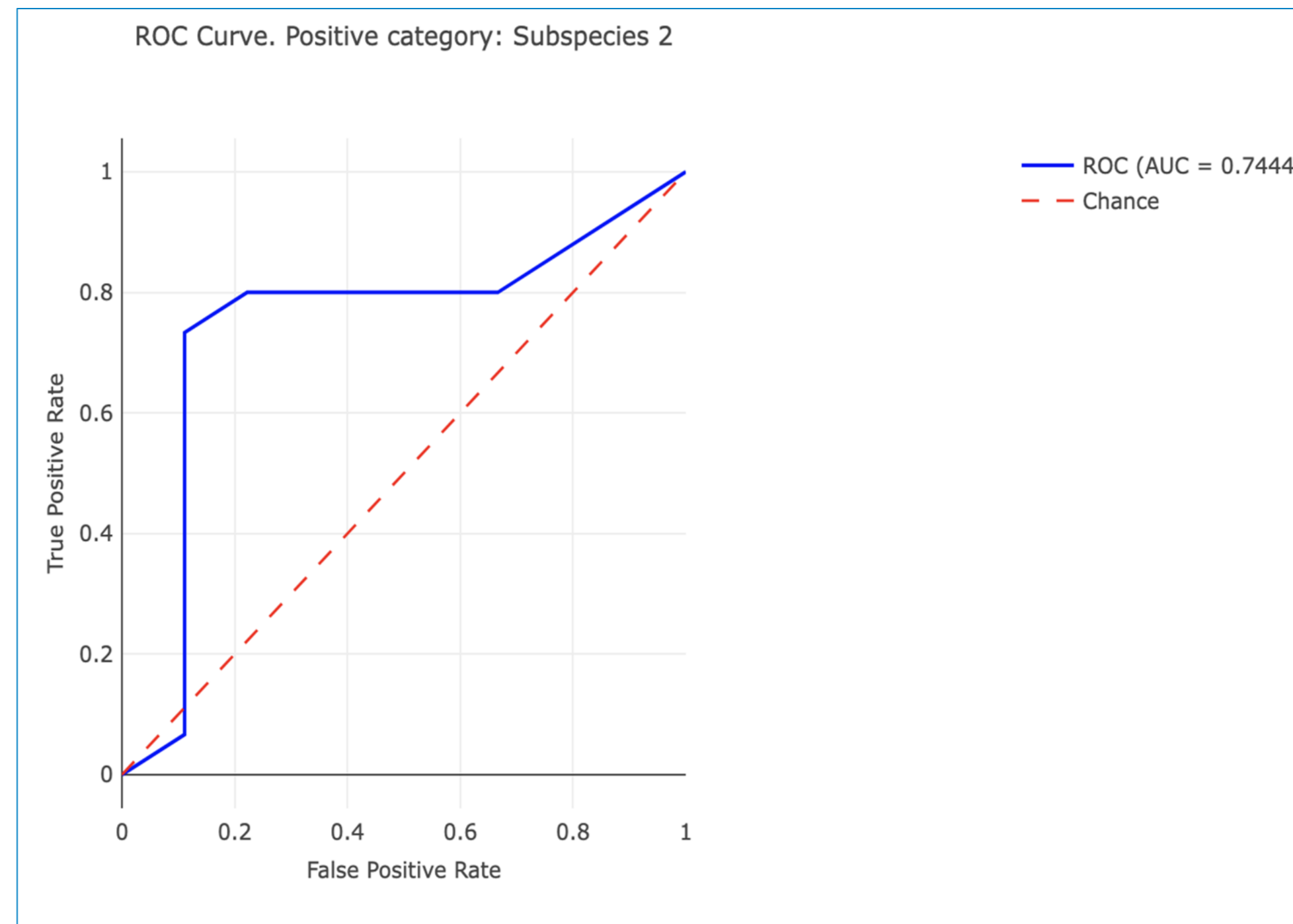
Clover MS Data Analysis Software, Quick Start Guide

Validation Analysis Metrics & Tools

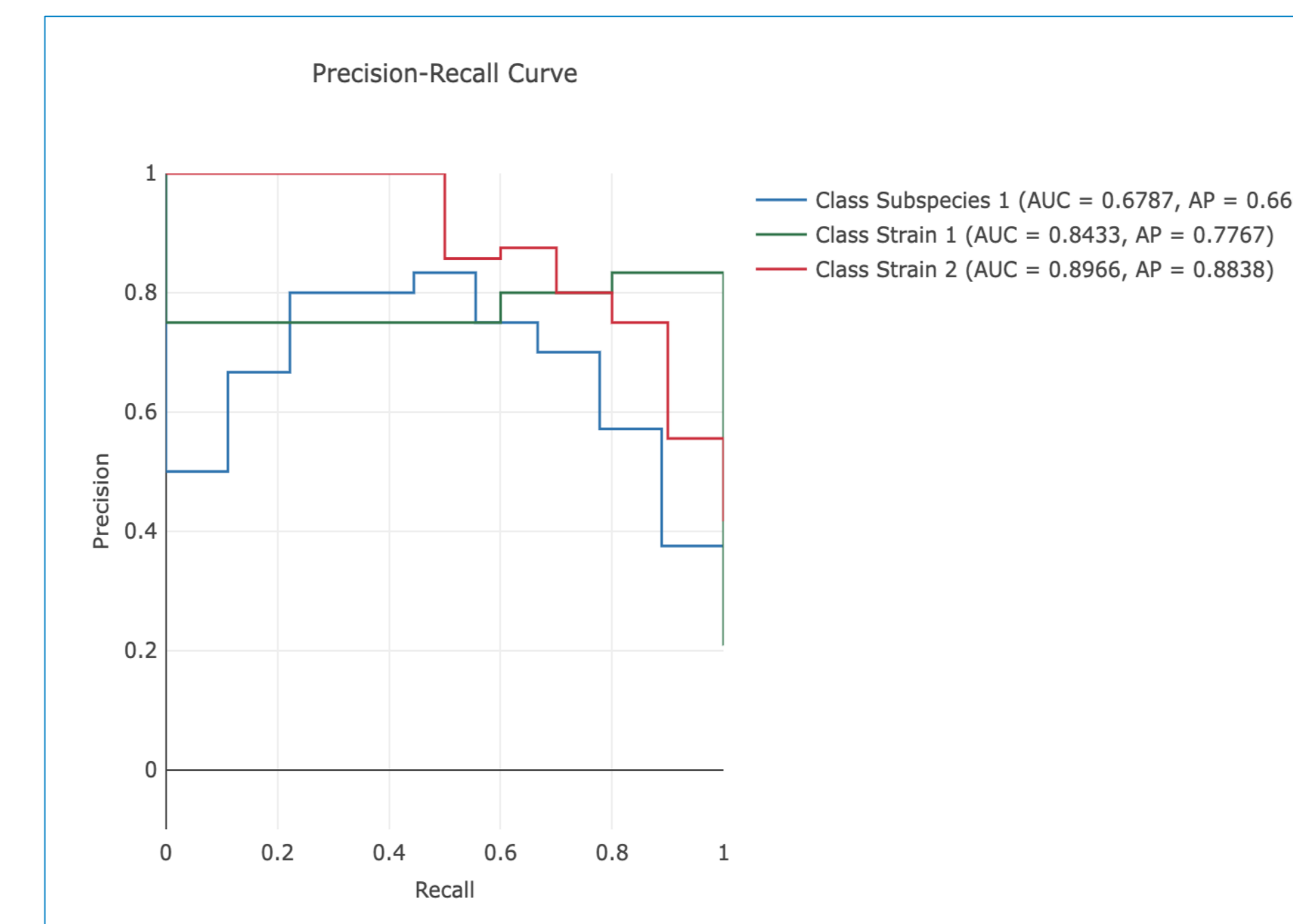
Hyperplanes and Predicted Data (SVM Analysis)



ROC Curves



PRC Curves



Confusion Matrix

Score: 79.17 %

Actual / Predicted	Subspecies 1	Strain 1	Strain 2	% Correct
Subspecies 1	7	1	1	77.78 %
Strain 1	0	5	0	100 %
Strain 2	3	0	7	70 %
				79.17 %

Confusion Matrix

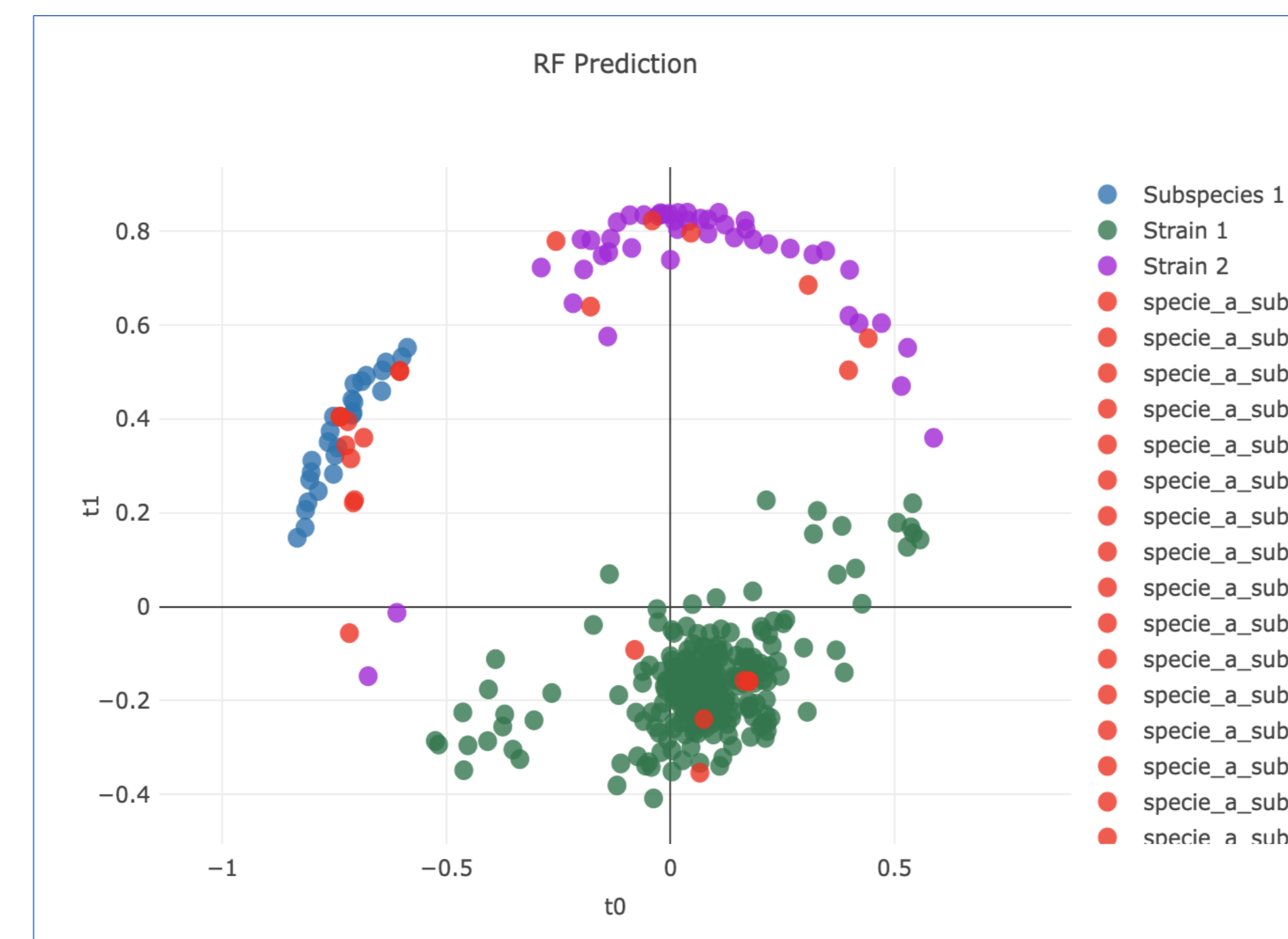
Balanced Accuracy: 82.59 %

Validation Result Table

Name	Control Category	Predicted Category (#1)	Probability (#1)
✓ specie_a_subspecie_1_specie_a_subspecie_1_sample_0_F1	Subspecies 1	Subspecies 1	99.95 %
✗ specie_a_subspecie_2_specie_a_subspecie_1_sample_0_F1	Subspecies 2	Subspecies 1	99.95 %
✓ specie_a_subspecie_1_specie_a_subspecie_1_sample_(10)_0_E12	Subspecies 1	Subspecies 1	99.39 %
✗ specie_a_subspecie_2_specie_a_subspecie_1_sample_(10)_0_E12	Subspecies 2	Subspecies 1	99.39 %
✓ specie_a_subspecie_1_specie_a_subspecie_1_sample_(11)_0_B1	Subspecies 1	Subspecies 1	98.85 %
✗ specie_a_subspecie_2_specie_a_subspecie_1_sample_(11)_0_B1	Subspecies 2	Subspecies 1	98.85 %
✓ specie_a_subspecie_1_specie_a_subspecie_1_sample_(12)_0_A10	Subspecies 1	Subspecies 1	96.92 %
✓ specie_a_subspecie_1_specie_a_subspecie_1_sample_(13)_0_H8	Subspecies 1	Subspecies 1	95.07 %

i The Predicted Category is calculated by the algorithm meanwhile Control Category is the one uploaded before to perform validation. If both are the same category it is presented with a green tick.

Distance Plot (Random Forest)



i Red dots correspond to the validation data uploaded.

