Machine learning applied to hyperspectral images

Authors:

Louis Guo (louis.guo)

Ilan Guenet (ilan.guenet)

Raphaël Ramelet (raphael.ramelet)

Table of content

1. Introduction
2. Load data
3. Preprocessing
4. Transformers
5. Multiclass classification (Indiana and Salinas)
6. Impact of the gaussian blur
7. Performance of the best classifier

Introduction

Problem to solve

- Develop a hyperspectral image classification pipeline

- Be able to segment our pictures into given categories

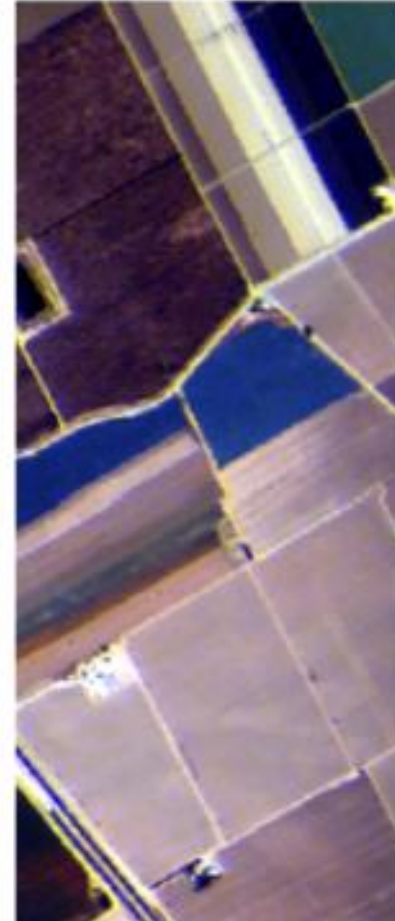- Have satisfactory and consistent results to support the coherence of our proposal

Load data
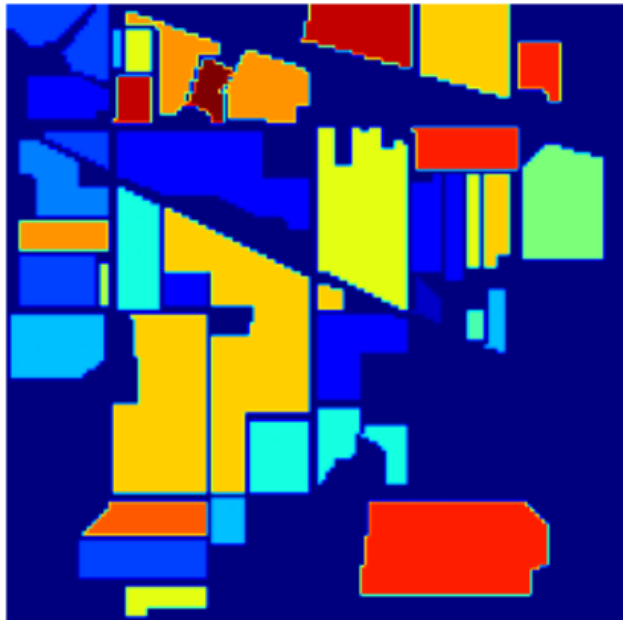
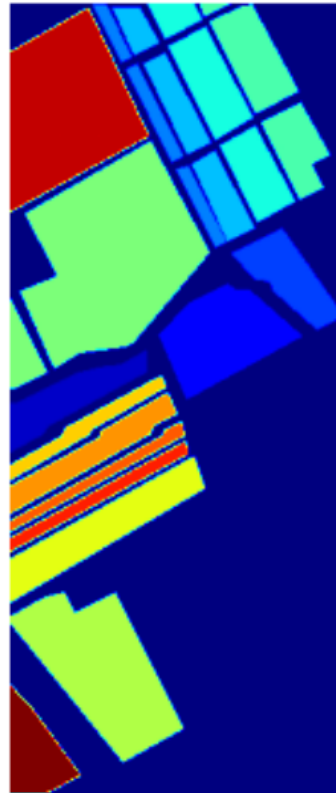# Hyperspectral Images

Indiana (Original)

Salinas (Original)

Labels

Labels of the Indiana image

Labels of the Salinas image

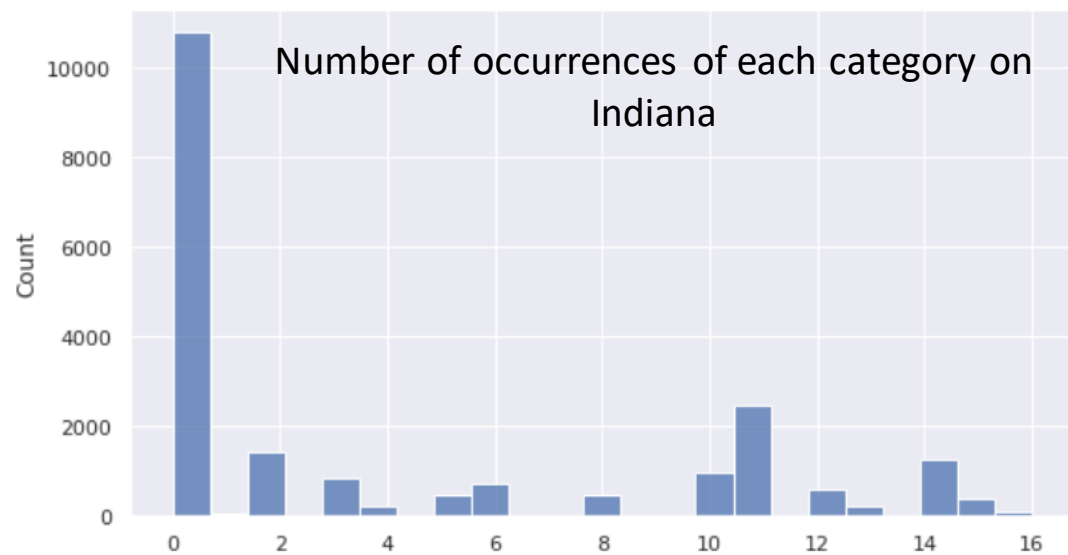# Unbalanced categories

Categories from Indiana		Categories from Salinas	
#	Category name	#	Category name
0	Other	0	Other
1	Alfalfa	1	Brocoli_green_weeds_1
2	Corn-notill	2	Brocoli_green_weeds_2
3	Corn-mintill	3	Fallow
4	Corn	4	Fallow_rough_plow
5	Grass-pasture	5	Fallow_smooth
6	Grass-trees	6	Stubble
7	Grass-pasture-mowed	7	Celery
8	Hay-windrowed	8	Grapes_untrained
9	Oats	9	Soil_vinyard_develop
10	Soybean-notill	10	Corn_senesced_green_weeds
11	Soybean-mintill	11	Lettuce_romaine_4wk
12	Soybean-clean	12	Lettuce_romaine_5wk
13	Wheat	13	Lettuce_romaine_6wk
14	Woods	14	Lettuce_romaine_7wk
15	Buildings-Grass-Trees-Drives	15	Vinyard_untrained
16	Stone-Steel-Towers	16	Vinyard_vertical_trellis

Number of occurrences of each category on Indiana

Number of occurrences of each category on Salinas
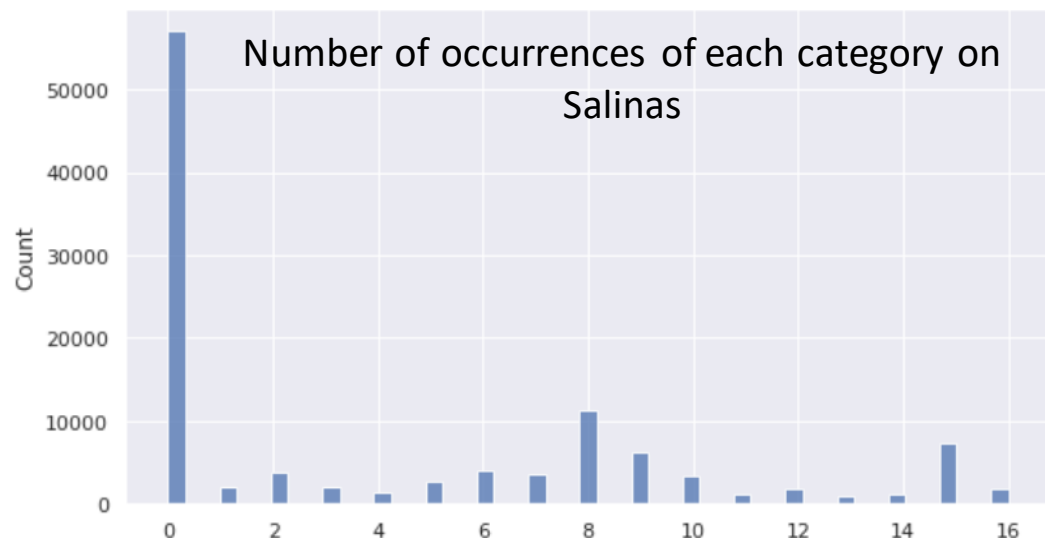
Preprocessing

Gaussian blur

Result of a gaussian blur applied images:
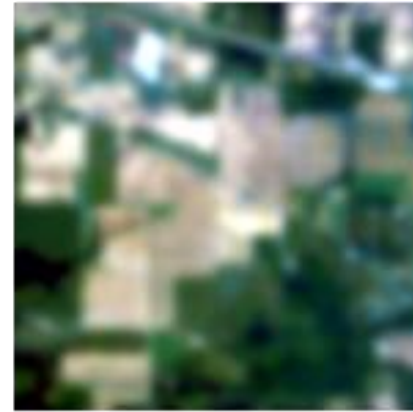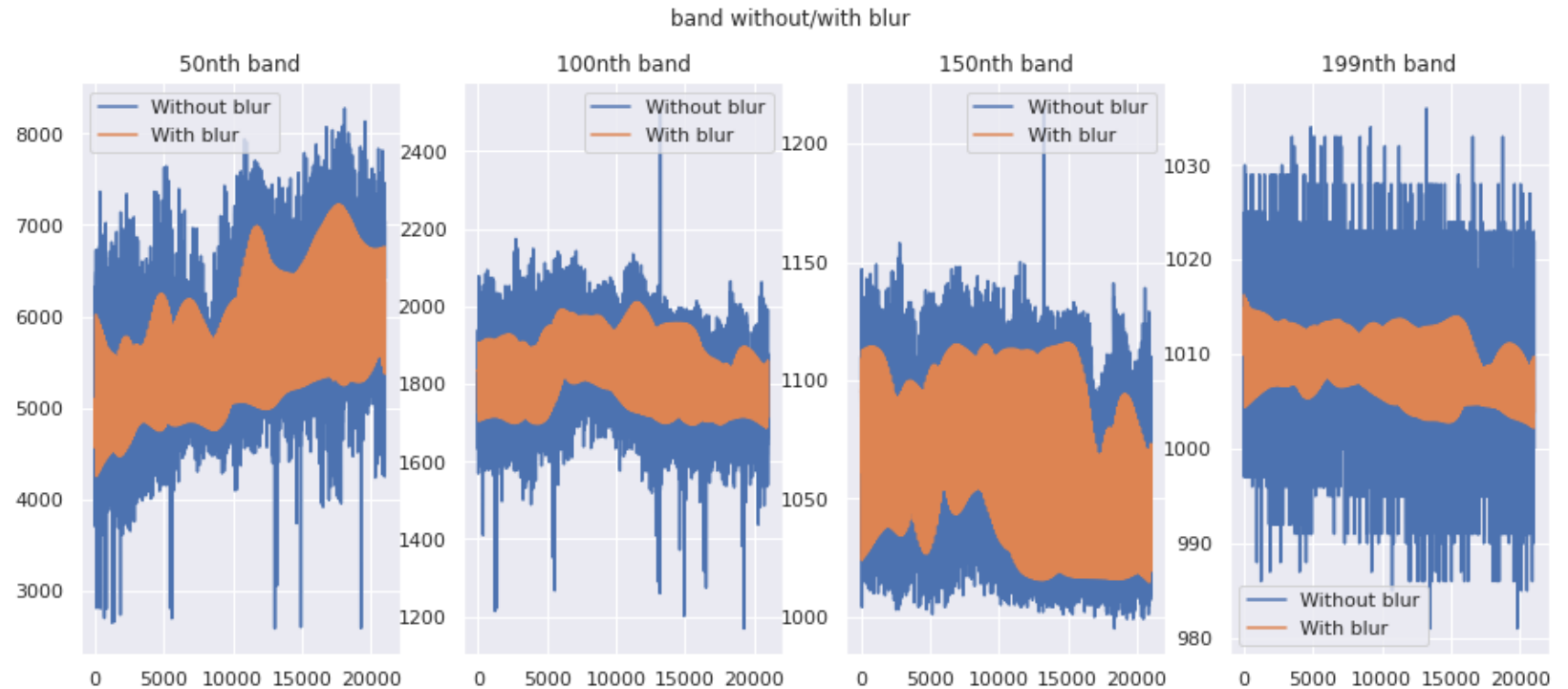
Why using a gaussian blur as a preprocessing step?

- Hyperspectral images have noise (instrumental noise)
- can be partly removed by using smoothing techniques

- smooth spikes
- Lower variance

band without/with blur

Transformers

Standard Scaler

$$x_{new} = \frac{x - u}{s}$$

Original distribution

Standardized data distribution
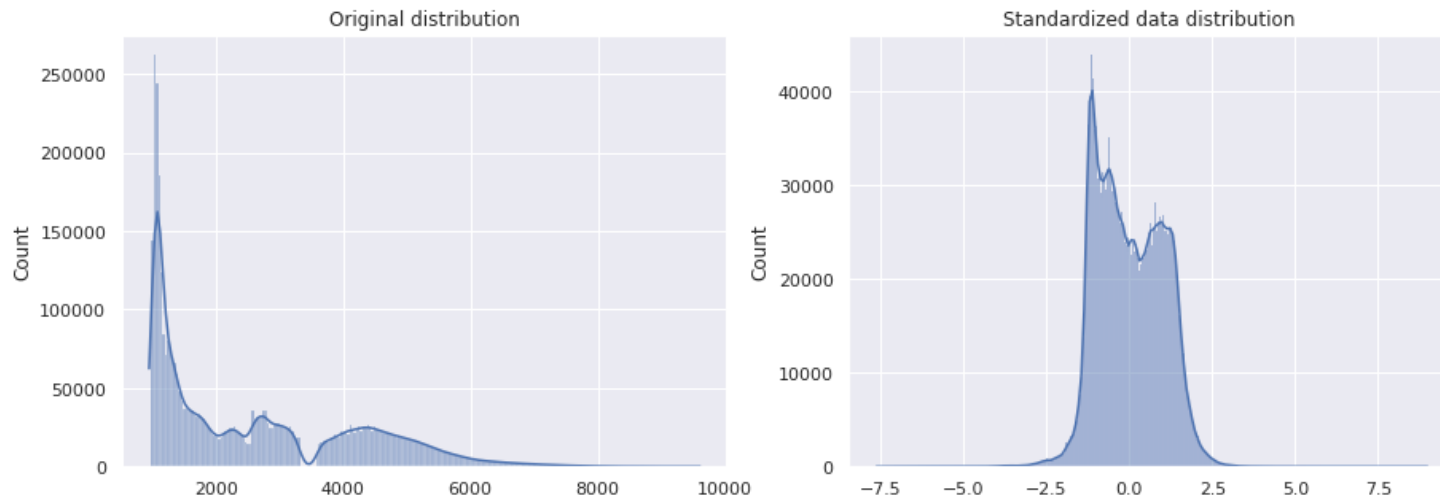
- scale the data to look like a standard normal distributed data
- Some algorithms assume that all features are centered around 0 and have variance in the same order (e.g. the RBF kernel of SVM).
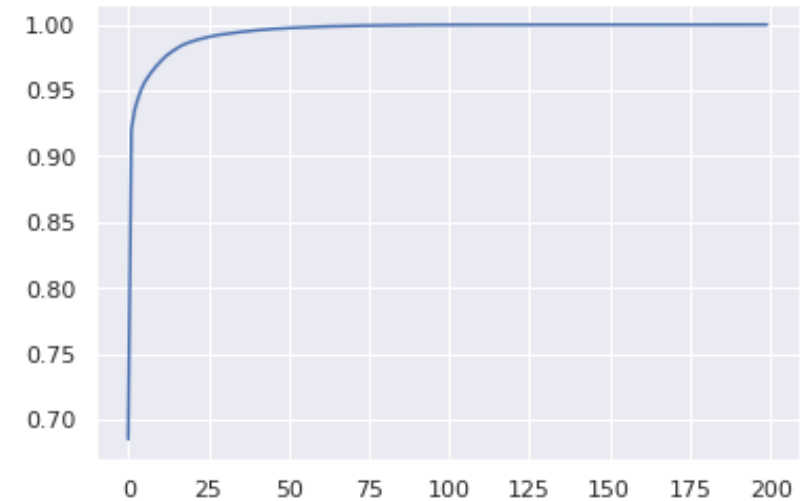
Dimension reduction (PCA)

Inverse projection 5 Principal components

Original picture

Cumulative sum of the explained ratio

- Project data in a new space
- Most of the information is contained in the first principal components
- The last components hold the information of noise
- Select the number of principal components based on a threshold

With vs Without the 'other' category

LinearSVC:
- Accuracy with: 0.86, without: 0.96
- f1 macro average score with: 0.79, without: 0.97
- f1 weigthed average score with: 0.86, without: 0.96

RandomForest:
- Accuracy with: 0.98, without: 1.00
- f1 macro average score with: 0.90, without: 1.00
- f1 weigthed average score with: 0.98, without: 1.00

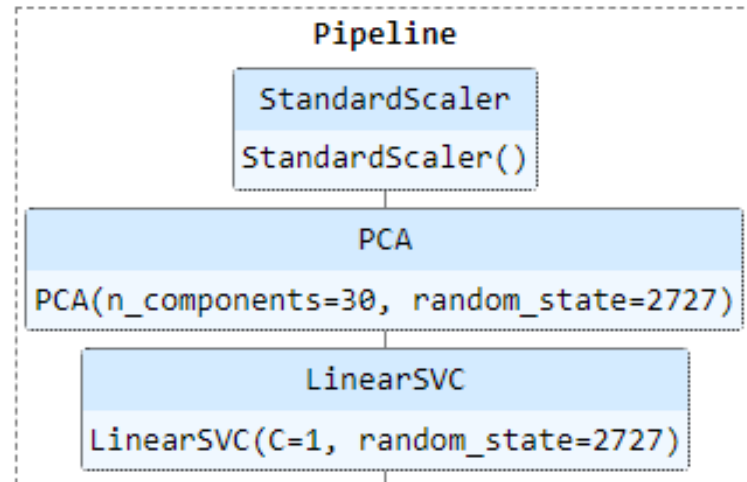# Multiclass classification (Indiana)

Indiana (Original)

- One Versus All
- N predicters
- Based on SVM
  (support vector machine algorithm)
- Find hyperplane that best divide
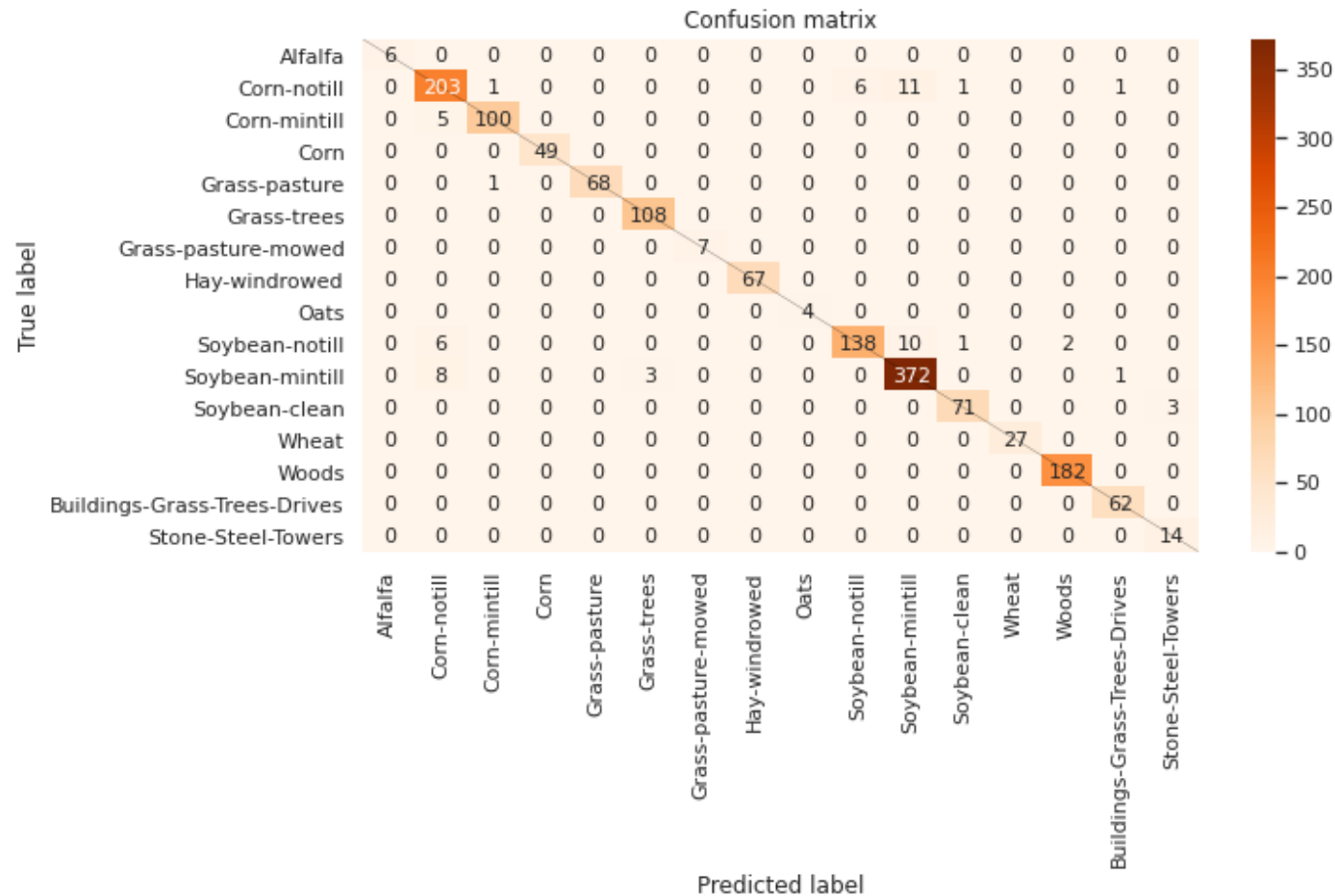 samples

Linear SVC (Indiana)

Final score


```
Score: 0.961.
f1 macro score: 0.974.
f1 weighted score: 0.961.
```

Score according to each category.

	precision	recall	f1-score	support
Alfalfa	1.00	1.00	1.00	6
Corn-notill	0.91	0.91	0.91	223
Corn-mintill	0.98	0.95	0.97	105
Corn	1.00	1.00	1.00	49
Grass-pasture	1.00	0.99	0.99	69
Grass-trees	0.97	1.00	0.99	108
Grass-pasture-mowed	1.00	1.00	1.00	7
Hay-windrowed	1.00	1.00	1.00	67
Oats	1.00	1.00	1.00	4
Soybean-notill	0.96	0.88	0.92	157
Soybean-mintill	0.95	0.97	0.96	384
Soybean-clean	0.97	0.96	0.97	74
Wheat	1.00	1.00	1.00	27
Woods	0.99	1.00	0.99	182
Buildings-Grass-Trees-Drives	0.97	1.00	0.98	62
Stone-Steel-Towers	0.82	1.00	0.90	14
accuracy			0.96	1538
macro avg	0.97	0.98	0.97	1538
weighted avg	0.96	0.96	0.96	1538

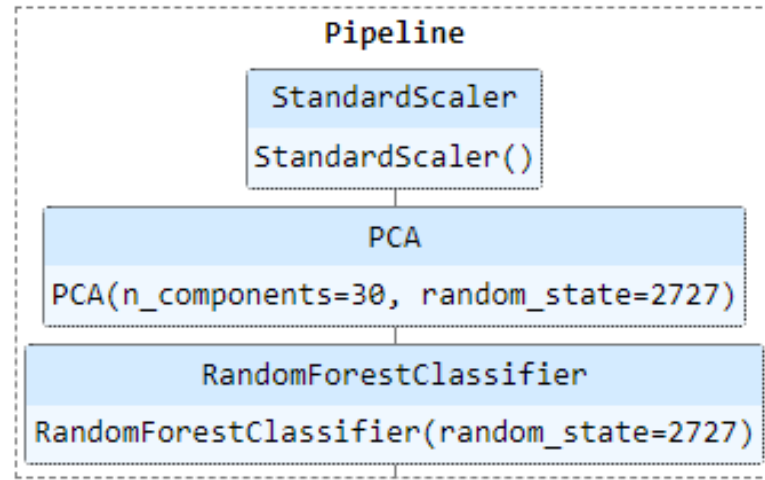# Linear SVC (Indiana)

The confusion matrix

Confusion matrix

Random Forest (Indiana)

- Inherently multiclass
- Collection of decision trees
- The class with the most votes along the decision trees wins

Final score

Score according to each category.


```
Pipeline
StandardScaler
StandardScaler()
PCA
PCA(n_components=30, random_state=2727)
RandomForestClassifier
RandomForestClassifier(random_state=2727)
```

```
Score: 0.999.
f1 macro score: 0.997.
f1 weighted score: 0.999.
```

	precision	recall	f1-score	support
Alfalfa	1.00	1.00	1.00	6
Corn-notill	1.00	1.00	1.00	223
Corn-mintill	1.00	1.00	1.00	105
Corn	1.00	1.00	1.00	49
Grass-pasture	1.00	0.99	0.99	69
Grass-trees	1.00	1.00	1.00	108
Grass-pasture-mowed	1.00	1.00	1.00	7
Hay-windrowed	1.00	1.00	1.00	67
Oats	1.00	1.00	1.00	4
Soybean-notill	1.00	1.00	1.00	157
Soybean-mintill	1.00	1.00	1.00	384
Soybean-clean	0.99	1.00	0.99	74
Wheat	1.00	1.00	1.00	27
Woods	1.00	1.00	1.00	182
Buildings-Grass-Trees-Drives	1.00	1.00	1.00	62
Stone-Steel-Towers	1.00	0.93	0.96	14
accuracy			1.00	1538
macro avg	1.00	0.99	1.00	1538
weighted avg	1.00	1.00	1.00	1538

Random Forrest (Indiana)
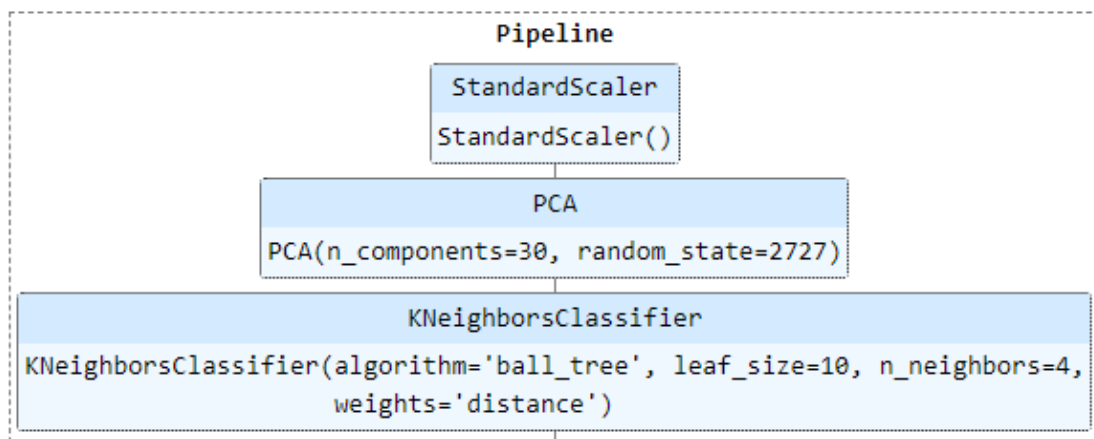
The confusion matrix

Confusion matrix
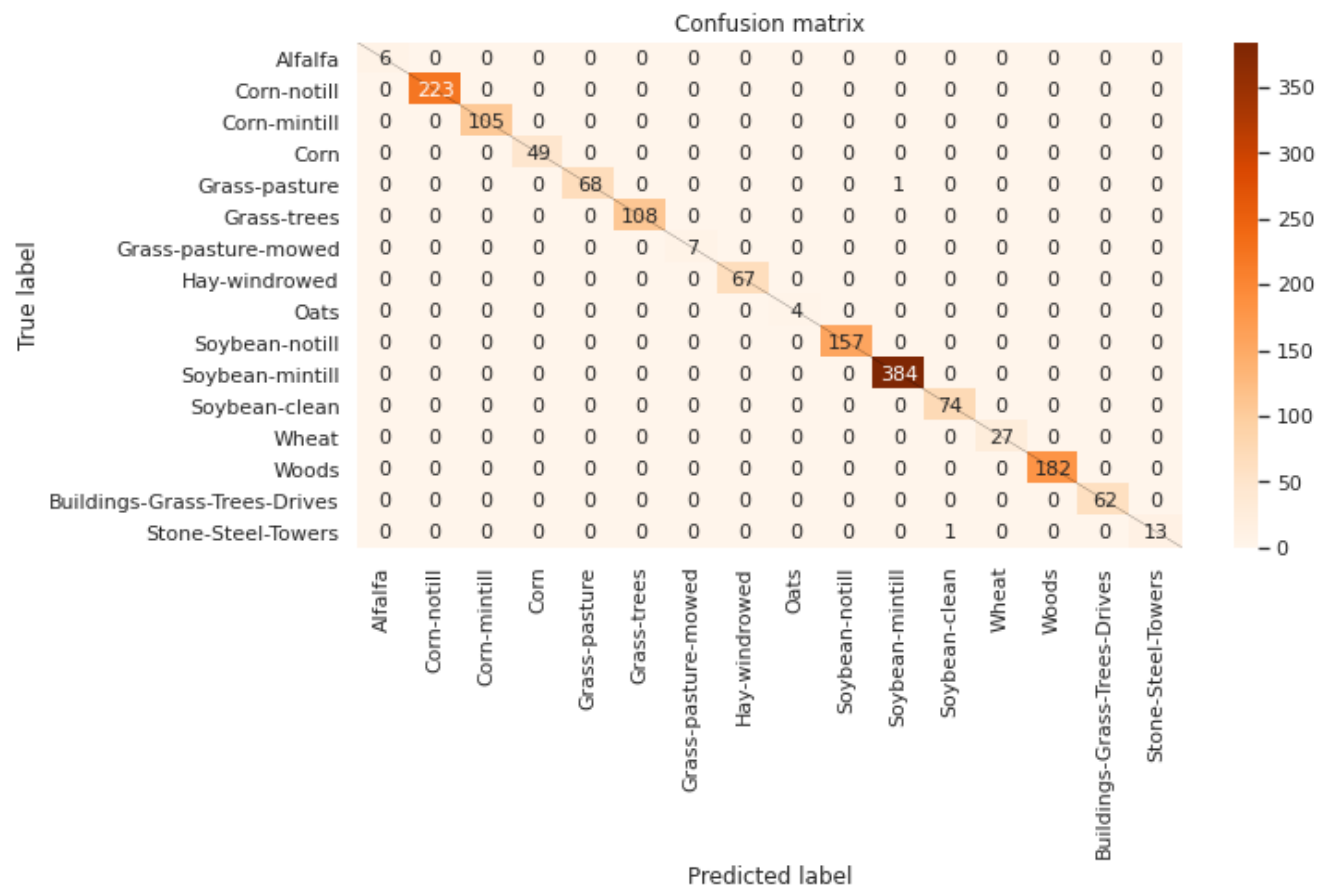
K Nearest Neighbours (Indiana)

- Inherently multiclass
- no real training step
- Naively, assigns according to the majority class in the neighbor-hood of the points
- Can be weighted by the distance

Final score

Score according to each category.


```
Pipeline
    StandardScaler
  StandardScaler()
        PCA
  PCA(n_components=30, random_state=2727)
        KNeighborsClassifier
  KNeighborsClassifier(algorithm='ball_tree', leaf_size=10, n_neighbors=4,
                       weights='distance')
```

Score: 0.996.
f1 macro score: 0.998.
f1 weighted score: 0.996.

	precision	recall	f1-score	support
Alfalfa	1.00	1.00	1.00	6
Corn-notill	1.00	1.00	1.00	223
Corn-mintill	1.00	1.00	1.00	105
Corn	1.00	1.00	1.00	49
Grass-pasture	1.00	1.00	1.00	69
Grass-trees	0.99	1.00	1.00	108
Grass-pasture-mowed	1.00	1.00	1.00	7
Hay-windrowed	1.00	1.00	1.00	67
Oats	1.00	1.00	1.00	4
Soybean-notill	0.99	0.98	0.99	157
Soybean-mintill	0.99	1.00	0.99	384
Soybean-clean	1.00	1.00	1.00	74
Wheat	1.00	1.00	1.00	27
Woods	1.00	0.99	1.00	182
Buildings-Grass-Trees-Drives	1.00	1.00	1.00	62
Stone-Steel-Towers	1.00	1.00	1.00	14
accuracy			1.00	1538
macro avg	1.00	1.00	1.00	1538
weighted avg	1.00	1.00	1.00	1538

The confusion matrix

Confusion matrix

SVC (Indiana)

- One Versus One
- N * (N - 1) / 2 predicters
- Based on SVM (support vector machine algorithm)
- Best with RBF (radial basis function) kernel

Final score

Score according to each category.


```
                                precision    recall  f1-score   support

                      Alfalfa       1.00      1.00      1.00         6
                 Corn-notill       1.00      1.00      1.00       223
                Corn-mintill       1.00      1.00      1.00       105
                        Corn       1.00      1.00      1.00        49
               Grass-pasture       1.00      1.00      1.00        69
                 Grass-trees       1.00      1.00      1.00       108
         Grass-pasture-mowed       1.00      1.00      1.00         7
                Hay-windrowed       1.00      1.00      1.00        67
                         Oats       1.00      1.00      1.00         4
               Soybean-notill       1.00      1.00      1.00       157
              Soybean-mintill       1.00      1.00      1.00       384
                Soybean-clean       1.00      0.99      0.99        74
                        Wheat       1.00      1.00      1.00        27
                        Woods       1.00      1.00      1.00       182
   Buildings-Grass-Trees-Drives      1.00      1.00      1.00        62
            Stone-Steel-Towers       0.93      1.00      0.97        14

                     accuracy                           1.00      1538
                    macro avg       1.00      1.00      1.00      1538
                 weighted avg       1.00      1.00      1.00      1538
```
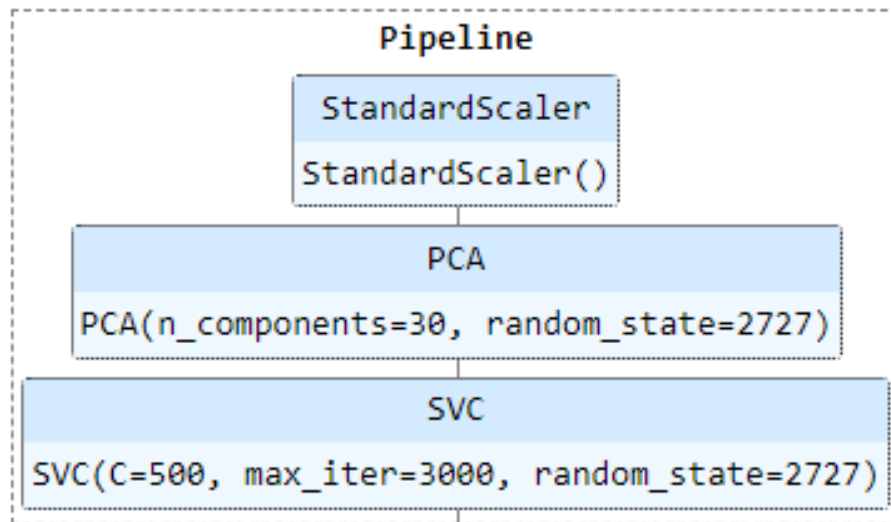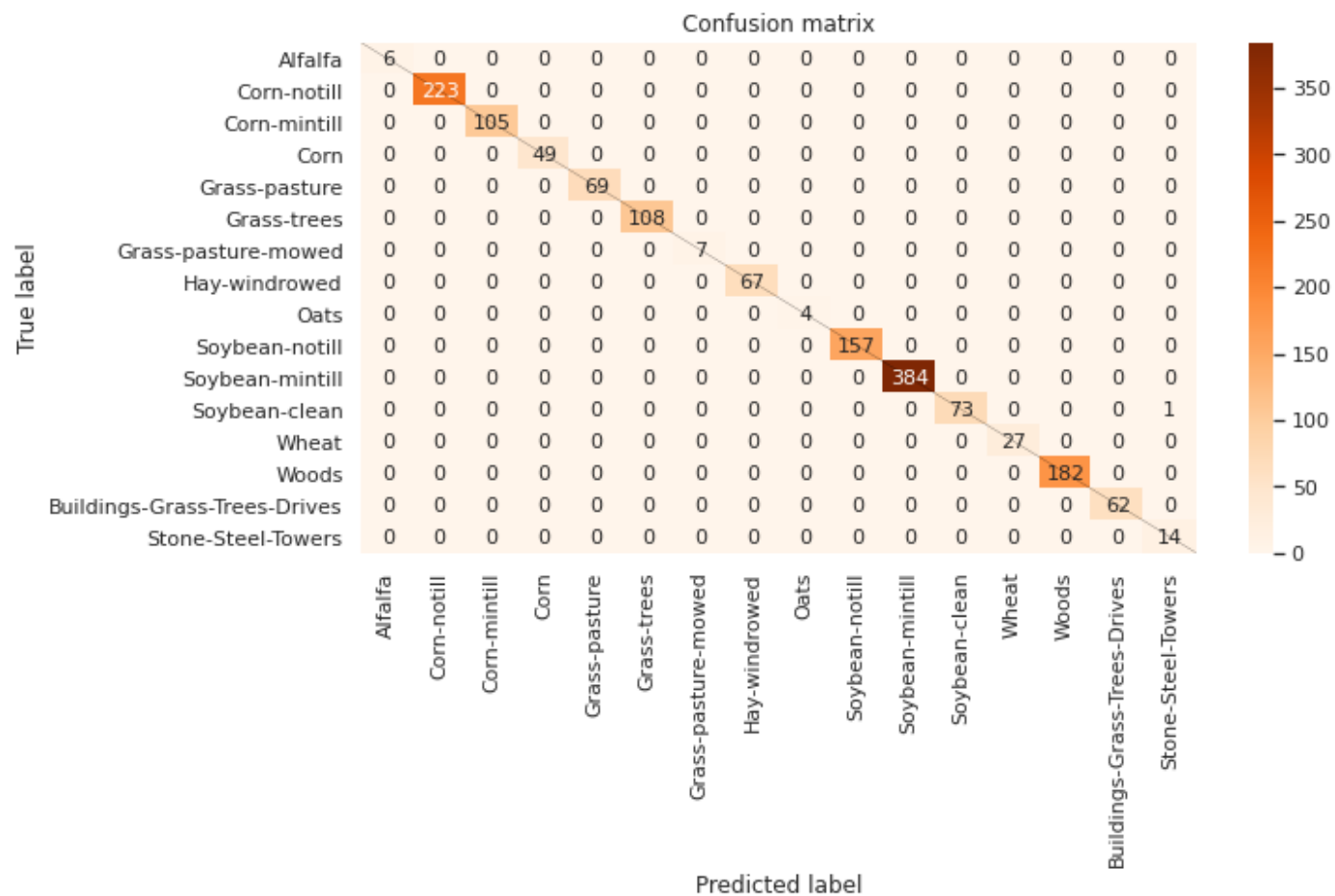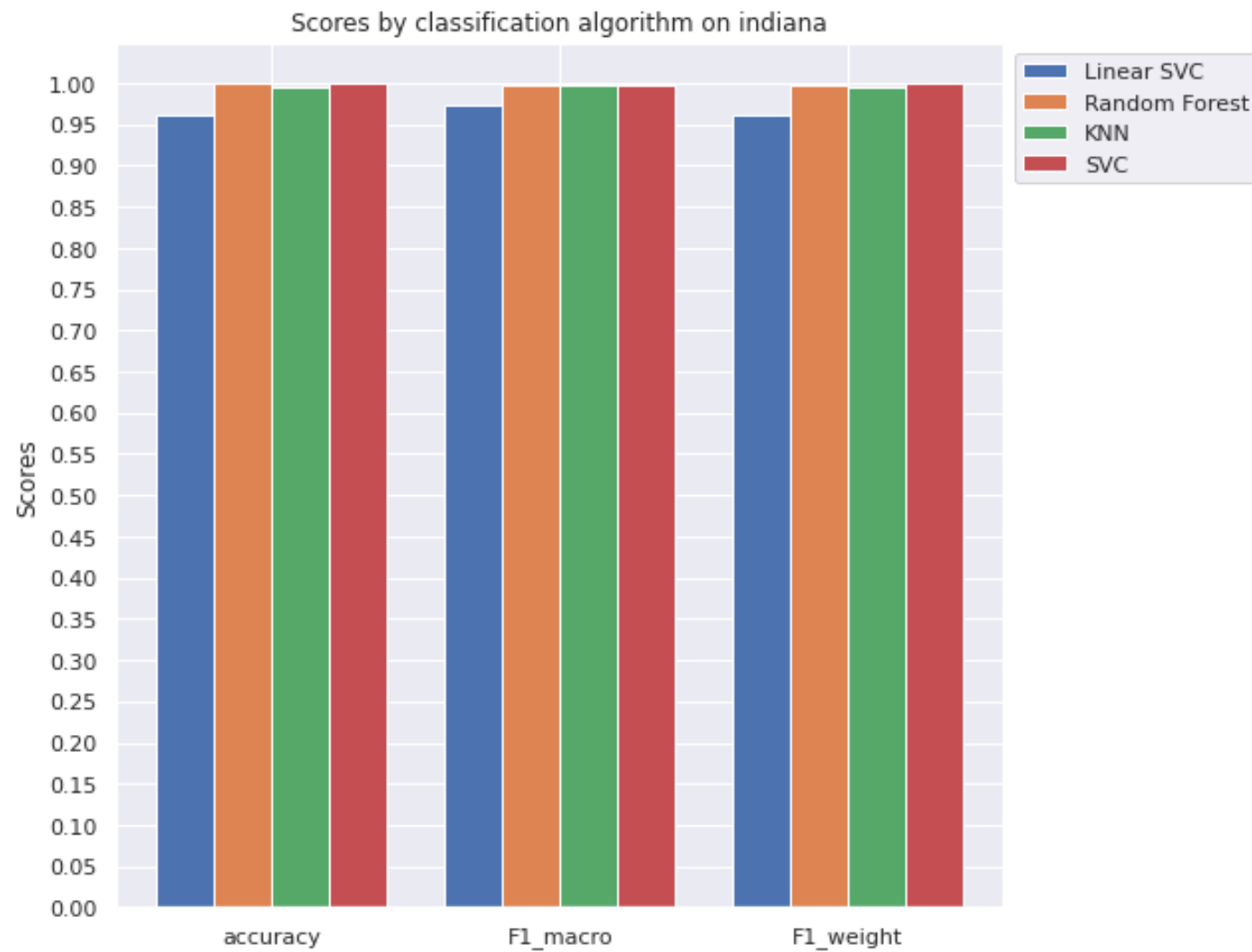
Pipeline

StandardScaler

StandardScaler()

PCA

PCA(n_components=30, random_state=2727)

SVC

SVC(C=500, max_iter=3000, random_state=2727)

Score: 0.999.
f1 macro score: 0.997.
f1 weighted score: 0.999.

SVC (Indiana)

The confusion matrix

Confusion matrix

Scores and time per classifier:

	accuracy	F1_macro	F1_weighted	Time (ms)
Linear SVC	0.961	0.9737	0.9608	25.2
Random Forest	0.9987	0.9967	0.9987	269.0
KNN	0.9961	0.9981	0.9961	301.0
SVC	0.9993	0.9974	0.9994	657.0

Benchmark (Indiana)

Scores by classification algorithm on indiana

Multiclass classification (Salinas)

Salinas (Original)

SVC (Salinas)

Final score


```
Pipeline
  StandardScaler
  StandardScaler()

  PCA
  PCA(n_components=30, random_state=2727)

  SVC
  SVC(C=500, max_iter=3000, random_state=2727)
```
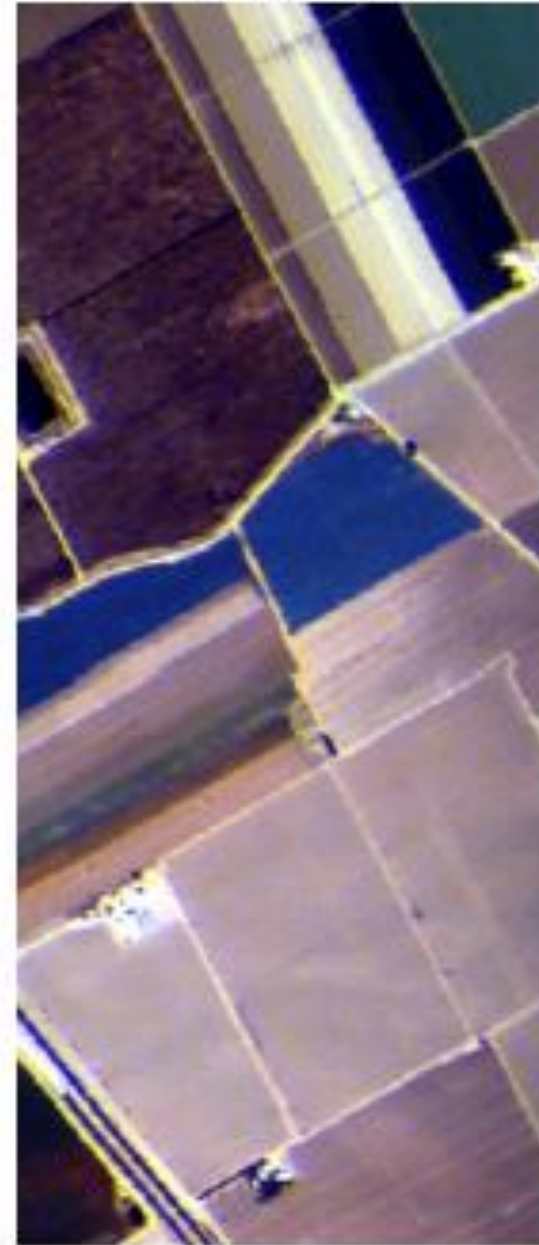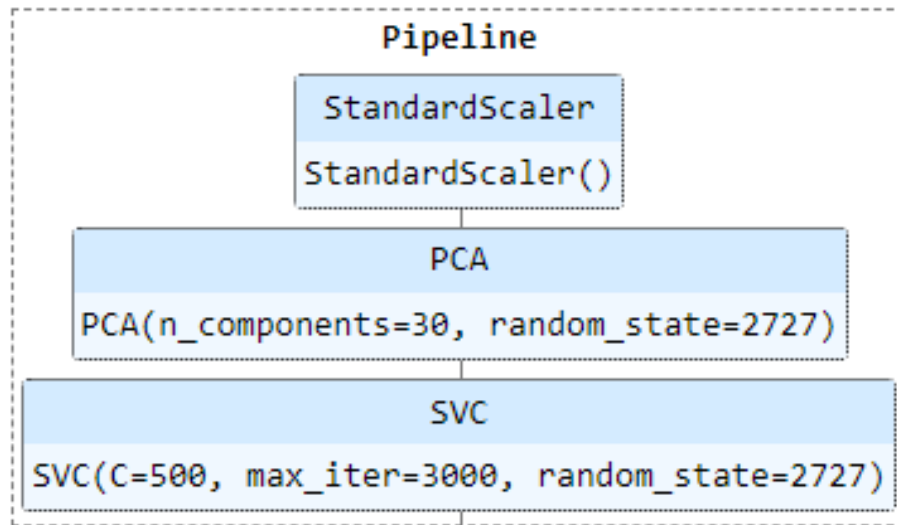
Score: 0.998.
f1 macro score: 0.999.
f1 weighted score: 0.998.

Score according to each category.

	precision	recall	f1-score	support
Brocoli_green_weeds_1	1.00	1.00	1.00	291
Brocoli_green_weeds_2	1.00	1.00	1.00	559
Fallow	1.00	1.00	1.00	278
Fallow_rough_plow	1.00	1.00	1.00	235
Fallow_smooth	1.00	1.00	1.00	422
Stubble	1.00	1.00	1.00	621
Celery	1.00	1.00	1.00	549
Grapes_untrained	1.00	0.99	1.00	1604
Soil_vinyard_develop	1.00	1.00	1.00	954
Corn_senesced_green_weeds	1.00	1.00	1.00	488
Lettuce_romaine_4wk	1.00	1.00	1.00	159
Lettuce_romaine_5wk	1.00	1.00	1.00	292
Lettuce_romaine_6wk	1.00	1.00	1.00	129
Lettuce_romaine_7wk	1.00	1.00	1.00	181
Vinyard_untrained	0.99	1.00	0.99	1097
Vinyard_vertical_trellis	1.00	1.00	1.00	261
accuracy			1.00	8120
macro avg	1.00	1.00	1.00	8120
weighted avg	1.00	1.00	1.00	8120

SVC (Salinas)

The confusion matrix

Confusion matrix with Salinas without `other` SVC classifier.
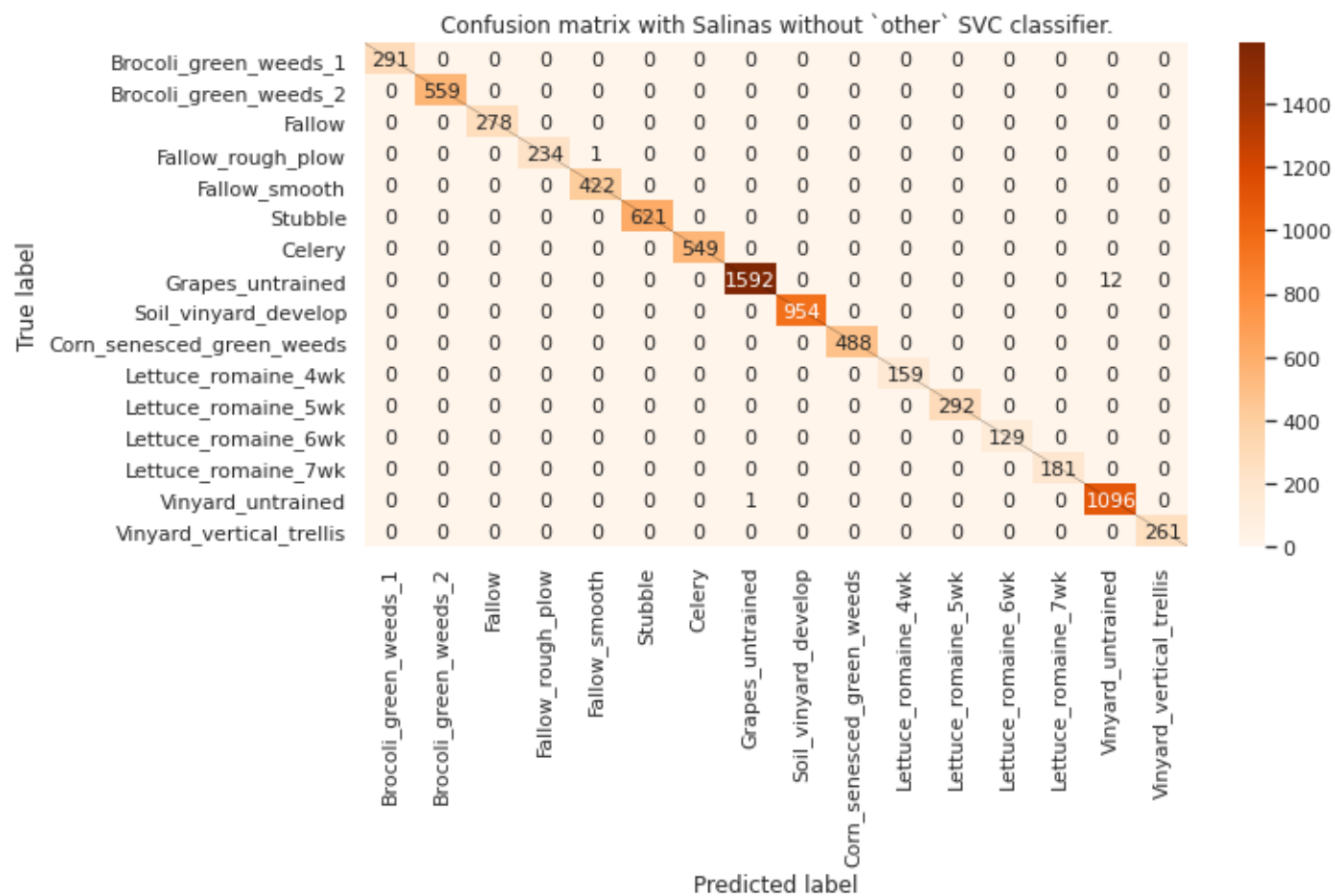
Random Forest (Salinas)

Final score
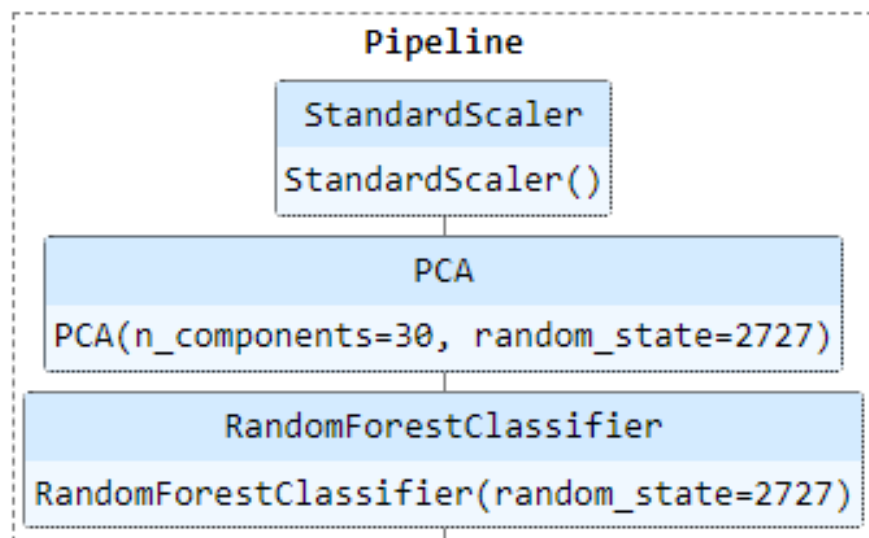

```
Score: 1.000.
f1 macro score: 1.000.
f1 weighted score: 1.000.
```

Scores of the classifier:

	precision	recall	f1-score	support
Brocoli_green_weeds_1	1.00	1.00	1.00	291
Brocoli_green_weeds_2	1.00	1.00	1.00	559
Fallow	1.00	1.00	1.00	278
Fallow_rough_plow	1.00	1.00	1.00	235
Fallow_smooth	1.00	1.00	1.00	422
Stubble	1.00	1.00	1.00	621
Celery	1.00	1.00	1.00	549
Grapes_untrained	1.00	1.00	1.00	1604
Soil_vinyard_develop	1.00	1.00	1.00	954
Corn_senesced_green_weeds	1.00	1.00	1.00	488
Lettuce_romaine_4wk	1.00	1.00	1.00	159
Lettuce_romaine_5wk	1.00	1.00	1.00	292
Lettuce_romaine_6wk	1.00	1.00	1.00	129
Lettuce_romaine_7wk	1.00	1.00	1.00	181
Vinyard_untrained	1.00	1.00	1.00	1097
Vinyard_vertical_trellis	1.00	1.00	1.00	261
accuracy			1.00	8120
macro avg	1.00	1.00	1.00	8120
weighted avg	1.00	1.00	1.00	8120

Random Forest (Salinas)

The confusion matrix

Confusion matrix with Salinas without `other` RandomForest classifier.

Score per classifier.

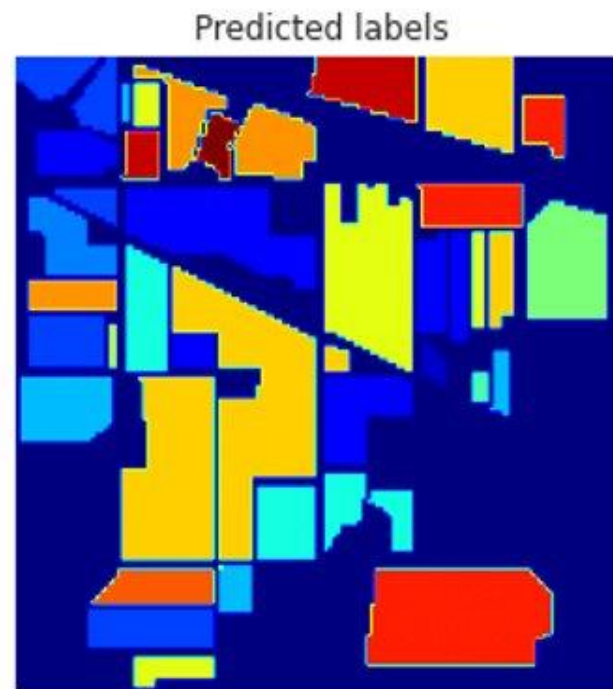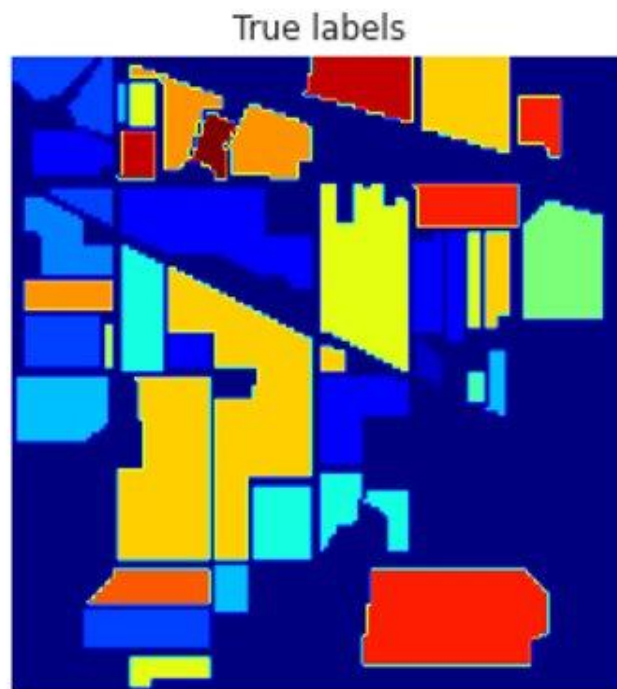|	accuracy	F1_macro	F1_weighted
SVC	0.9983	0.9992	0.9983
Random Forest	1.0	1.0	1.0
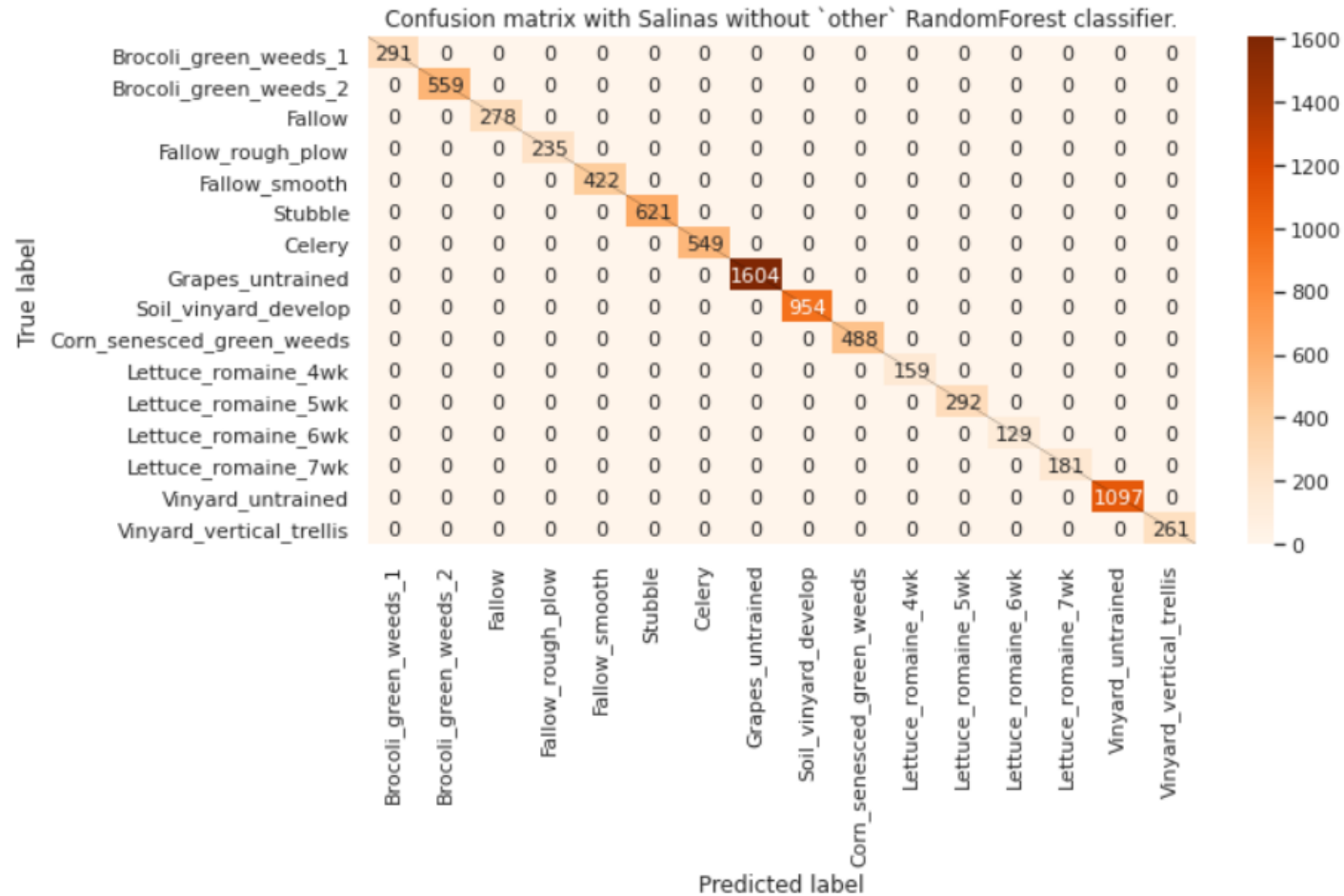
Benchmark (Salinas)

Scores by classification algorithm on salinas
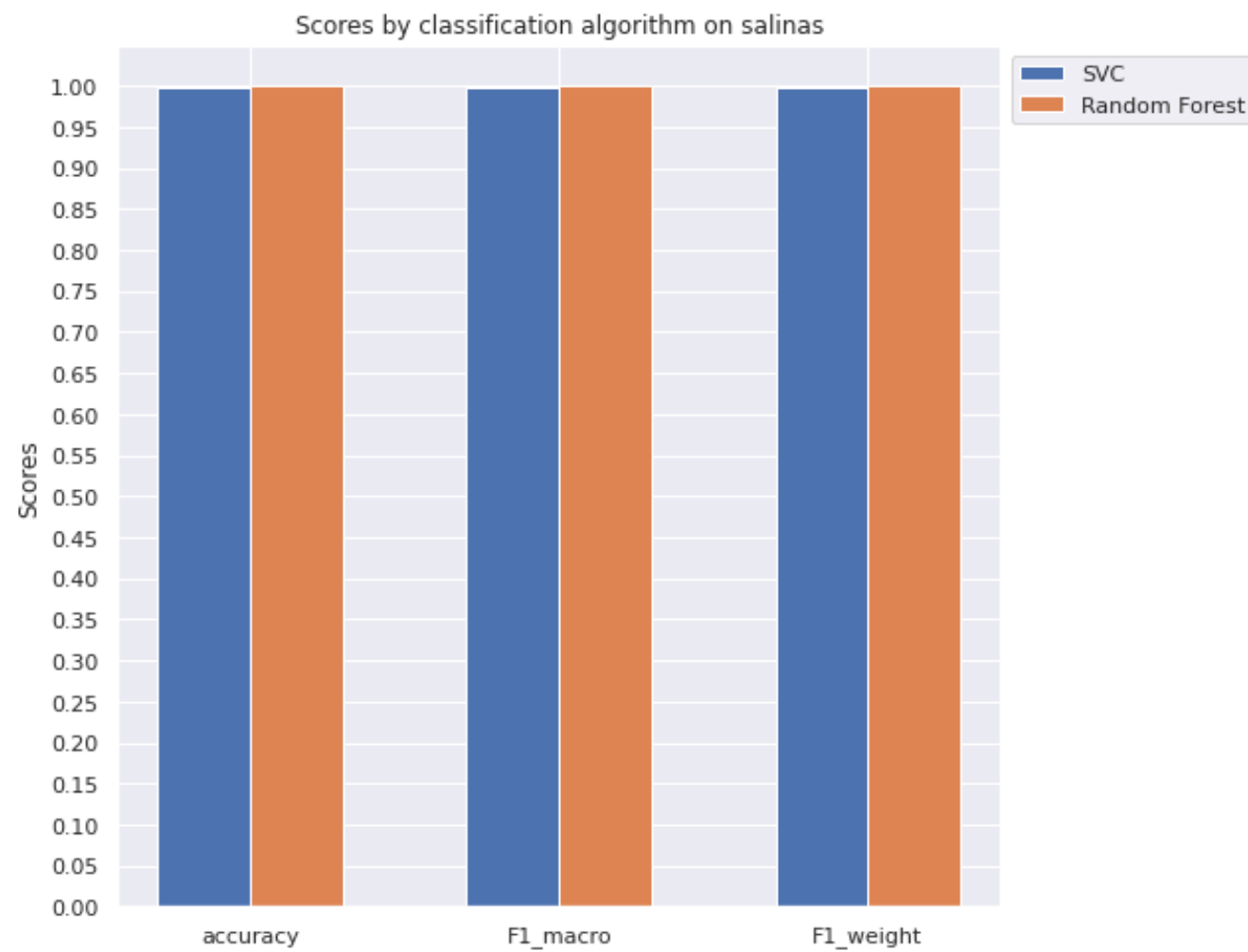
Impact of the gaussian blur
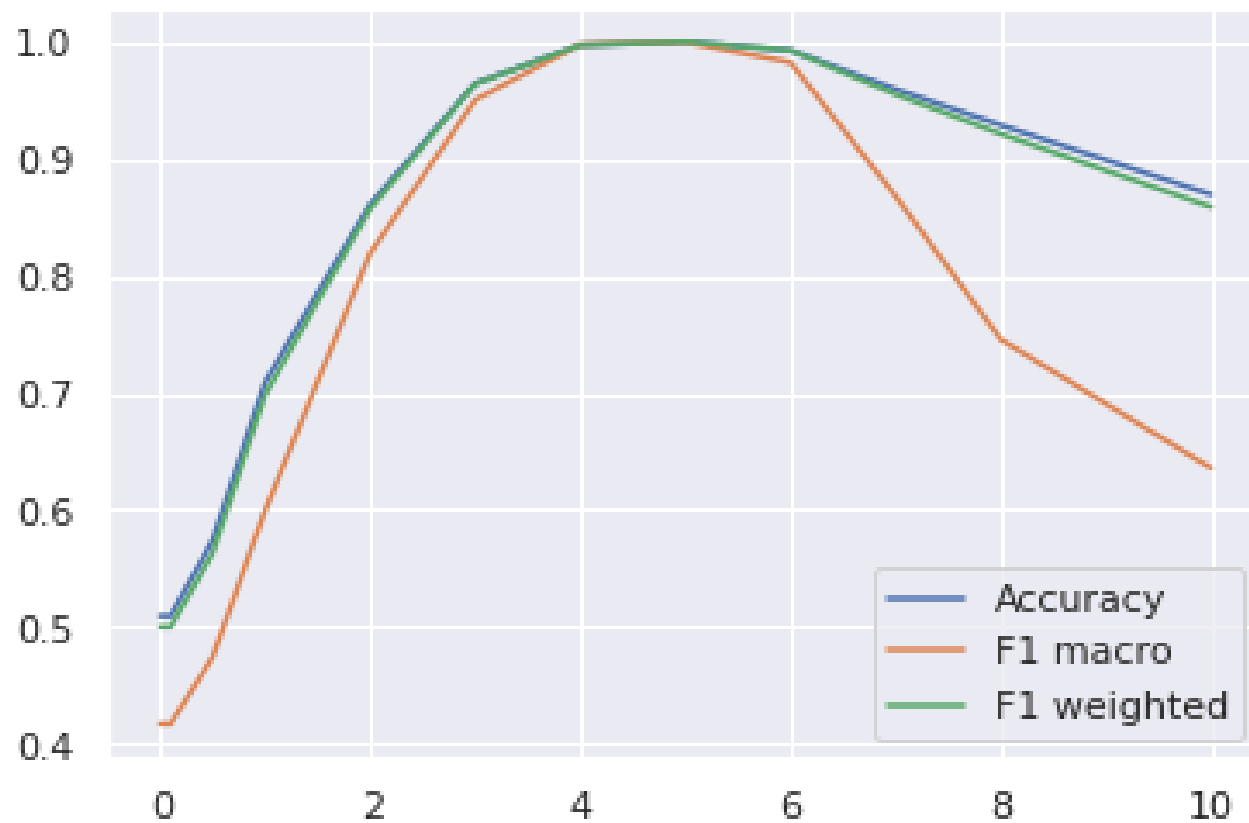
Performance of the best classifier

Random Forrest on Indiana
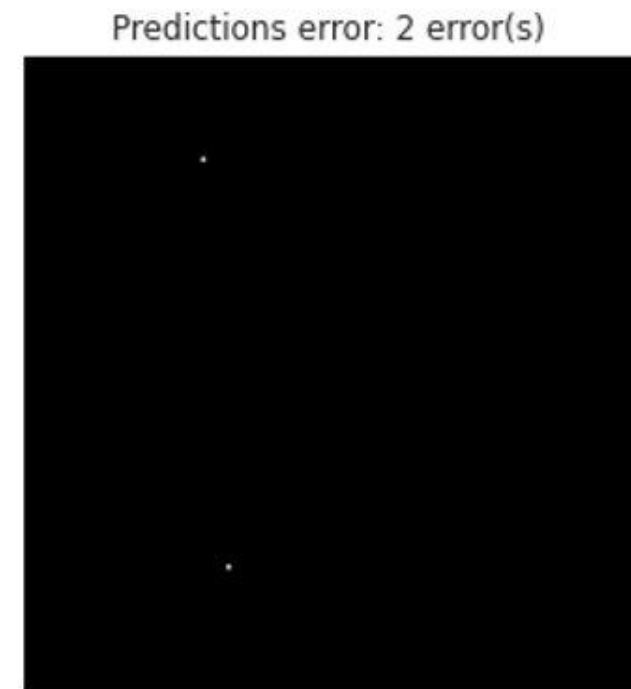
2 errors were made by the classifier.

Random Forrest on Salinas

True labels   Predicted labels   Predictions error: 0 error(s)

No errors were made by the classifier.

Conclusion