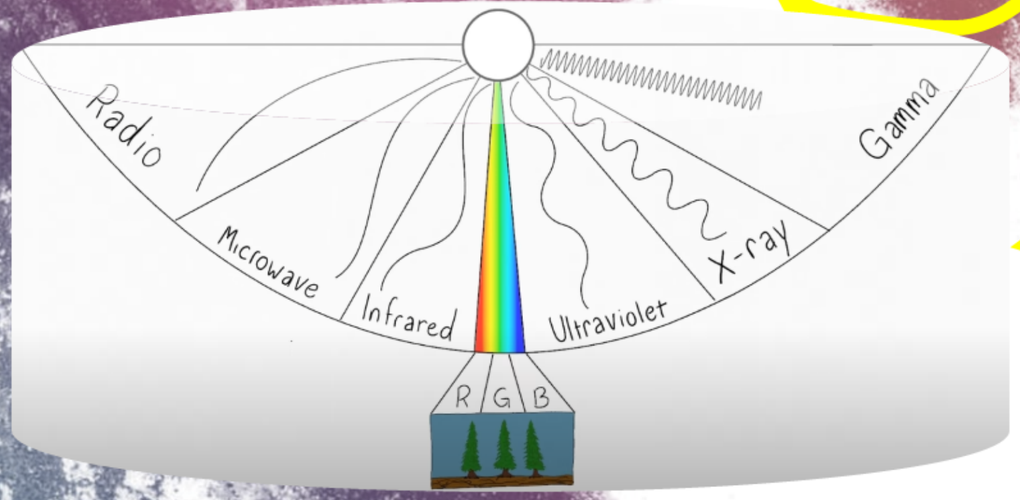
COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR LAND COVER CLASSIFICATION USING HYPERSPECTRAL IMAGES

University of the Philippines - Diliman

AI 201 Mini-Project | 1st semester, A.Y. 2023-2024 | January 16, 2024

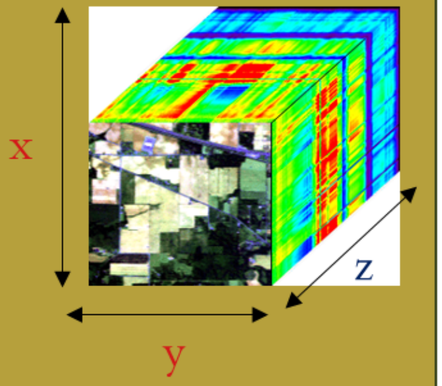
Mary Nathalie Dela Cruz | Elsa Joy Horiondo

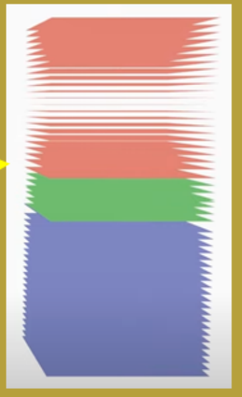
INTRODUCTION



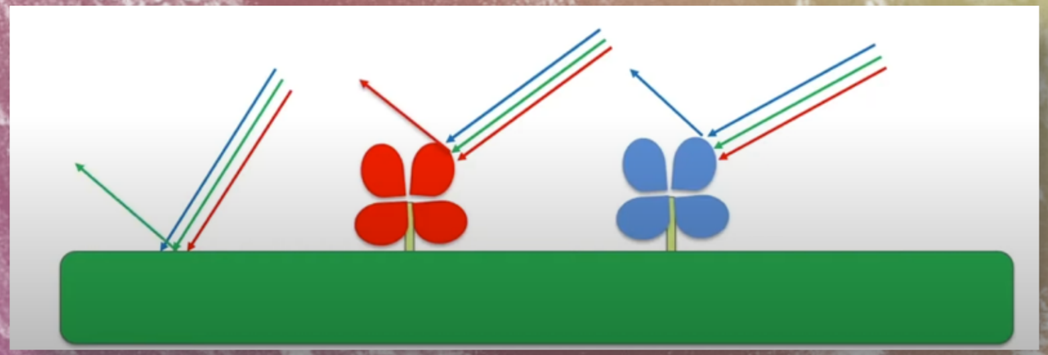
Hyperspectral Imaging (HSI)

- Imaging + Spectroscopy
- Data Cube
  - Spatial (x- and y-axis)
  - Wavelength (z-axis)
- Principles
  - EM spectrum records thousands of unique wavelengths
  - Grouped into bands
  - Uses hundreds of bands



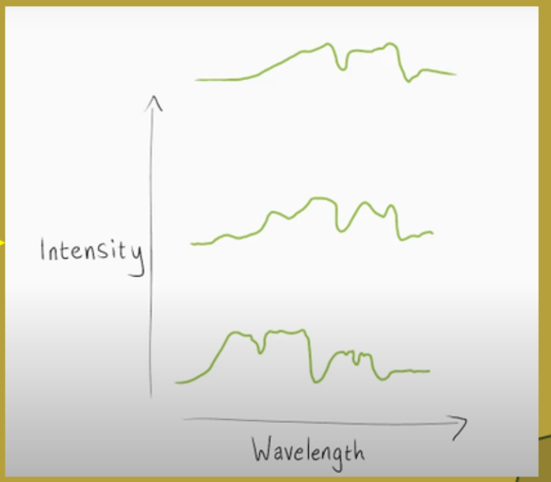


INTRODUCTION



Hyperspectral Imaging (HSI)

- Spectral Fingerprint
  - Objects reflect and absorb (vegetation, minerals, building materials)
- Hyperspectral sensors can precisely detect these spectra



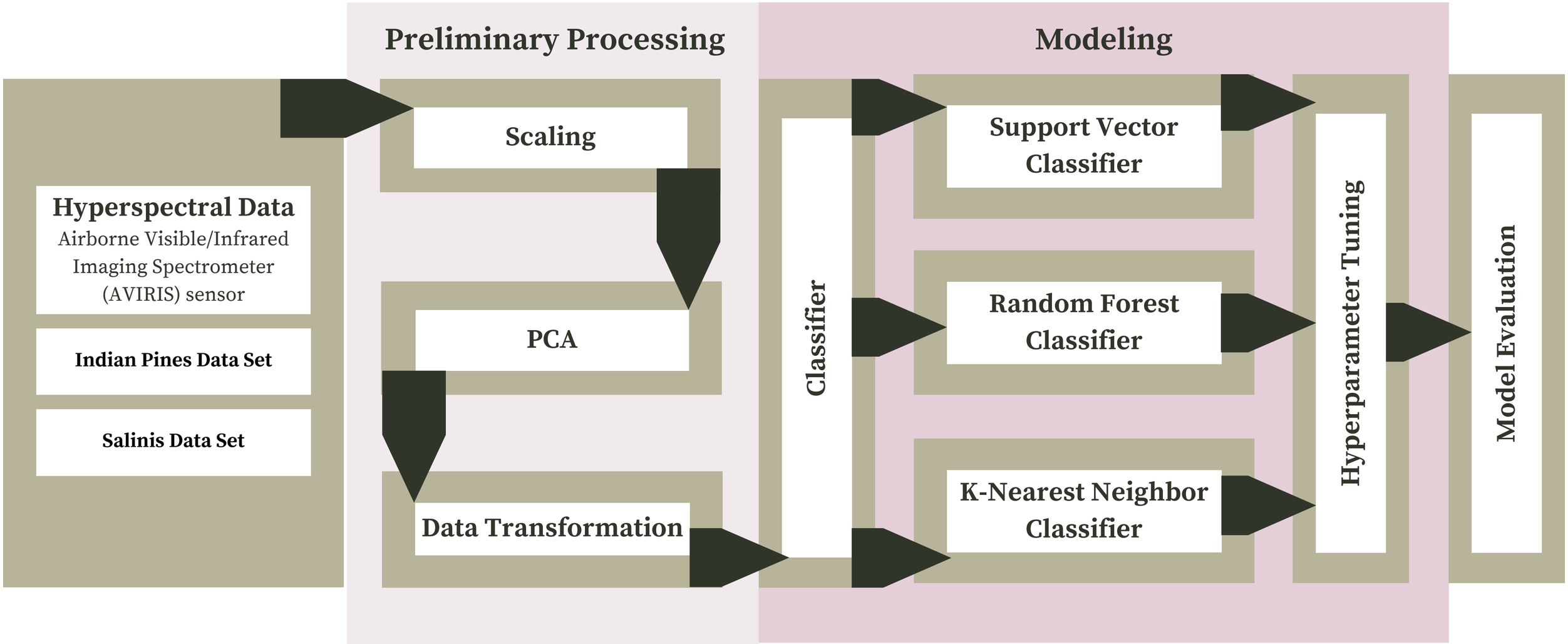
INTRODUCTION

Hyperspectral Imaging (HSI)

- Relevance
  - Agriculture
  - Art conservation
  - Medicine
  - Pollution Mapping
  - Surveillance
  - Geological mapping

OBJECTIVES

- To analyze HSI of Indian Pines and Salinas dataset using three classification methods
- To determine the optimal hyperparameters for the models
- To gauge the algorithms thru evaluation metrics (Precision, Recall, F1)


Preliminary Processing
Modeling
Hyperspectral Data
Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor
Indian Pines Data Set
Salinis Data Set
Scaling
PCA
Data Transformation
Classifier
Support Vector Classifier
Random Forest Classifier
K-Nearest Neighbor Classifier
Hyperparameter Tuning
Model Evaluation

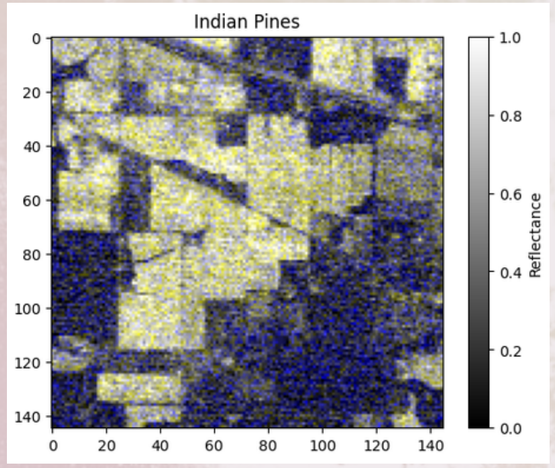
HYPERSPECTRAL DATA
from AVIRIS Sensor

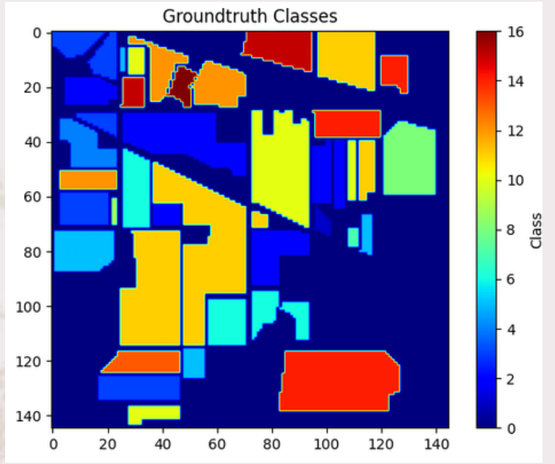
Indian Pines Data Set

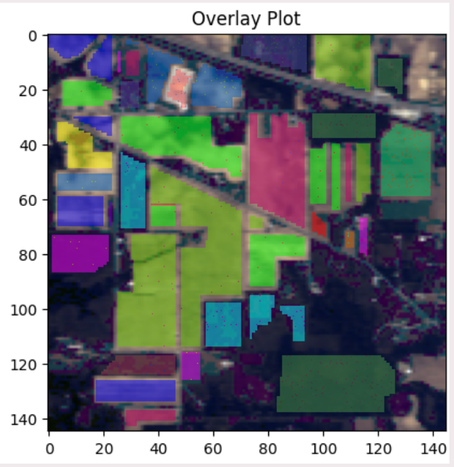
Location: Indian Pines test site in
North-Western Indiana, U.S.A.
spatial extent of 145 × 145 pixels
per band with 200 bands

Class	1	2	3	4	5	6	7	8
Name	Alfalfa	Corn notill	Corn mintill	Corn	Grass pasture	Grass trees	Grass pasture mowed	Hay windrowed
Samples	46	1428	830	237	483	730	28	478

Class	9	10	11	12	13	14	15	16
Name	Oats	Soybean notill	Soybean mintill	Soybean clean	Wheat	Woods	Buildings grass trees drives	Stone steel towers
Samples	20	972	2455	593	205	1265	386	93





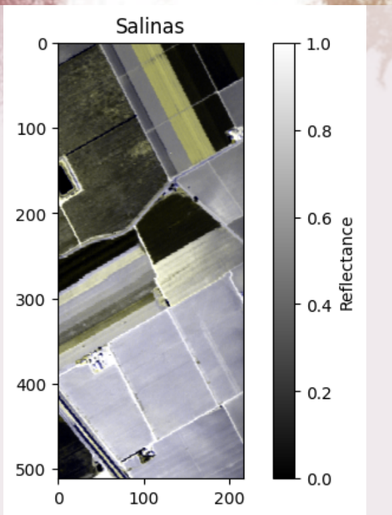


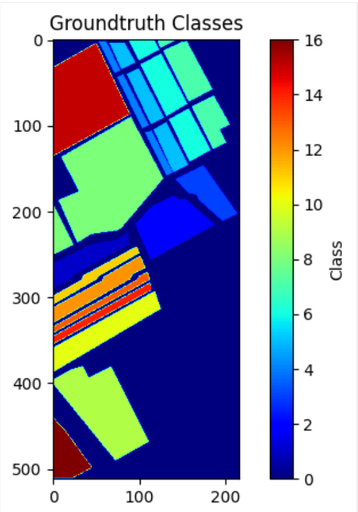
Salinas Data Set

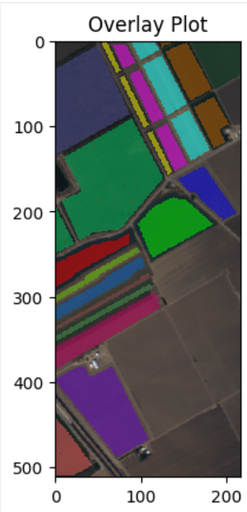
Location: Salinas Valley,
California, U.S.A.
spatial extent of 512 × 217 pixels
per band with 204 bands

Class	1	2	3	4	5	6	7	8
Name	Broccoli green weeds 1	Broccoli green weeds 2	Fallow	Fallow rough plough	Fallow smooth	Stubble	Celery	Grapes untrained
Samples	2009	3726	1976	1394	2678	3959	3579	11271

Class	9	10	11	12	13	14	15	16
Name	Soil vineyard develop	Corn senesced green weeds	Lettuce romaine 4 wk	Lettuce romaine 5wk	Lettuce romaine 6wk	Lettuce romaine 7wk	Vineyard untrained	Vineyard vertical trellis
Samples	6203	3278	1068	1927	916	1070	7268	1807







PRE-PROCESSING

SCALING

$$z = (x - u) / s$$

x is the original feature vector
u is the mean of feature vector
s is the standard deviation of vector

PCA

Covariance Computation

$$C_{ij} = 1/(n-1) * \Sigma (x_i - u_i) * (x_j - u_j)$$

x_i and x_j are data vectors for bands i and j
u_i and u_j are their respective means
n is the number of pixels

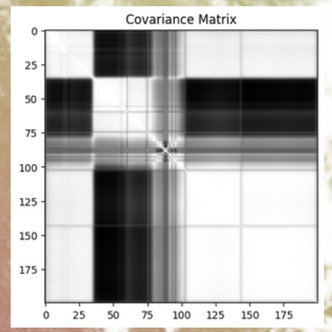
Eigendecomposition

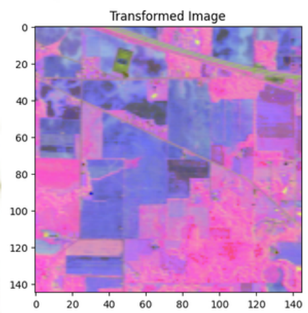
Data Transformation

$$y = dot(E^T , x)$$

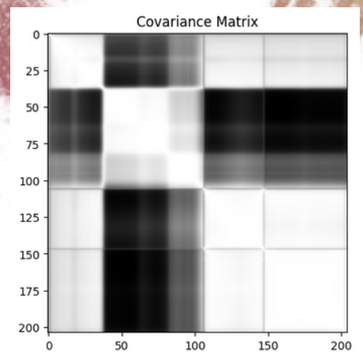
E is the matrix of eigenvectors
x is the original data vector
y is the transformed data vector

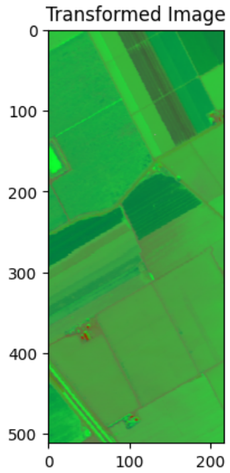
Indian Pine Dataset
200 bands > 108 bands





Salinas Dataset
204 bands > 19 bands





Data Transformation

- 3D Data Cube
- 2D Data Frame
- Train-Test Split

pixel_row	pixel_col	band-1	band-108	class
0	0	-10.573985	-0.083758	3
0	1	-13.157707	-0.060955	3
0	2	-14.070350	-0.139894	3
0	3	-13.356123	-0.064724	3
0	4	-11.591195	-0.114516	3

MODEL DEVELOPMENT

INDIAN PINES

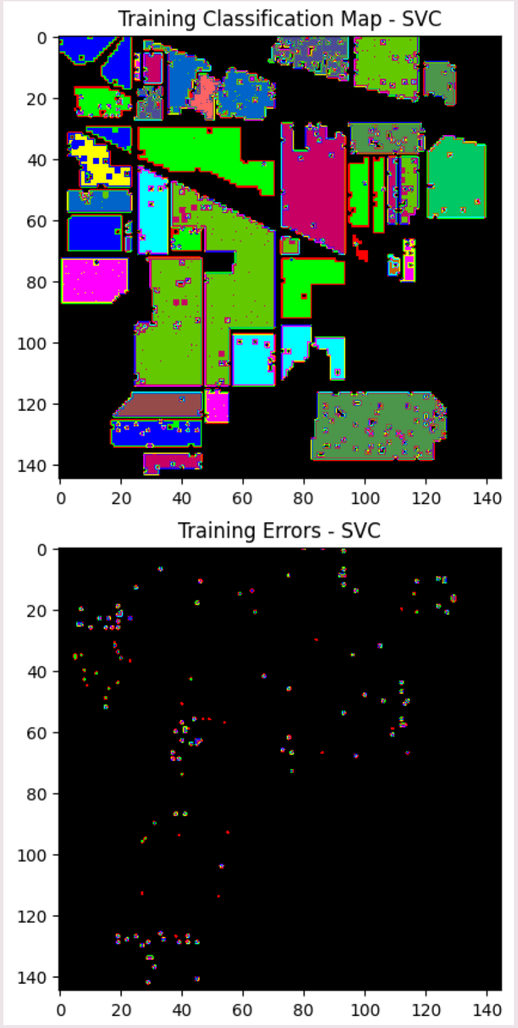
	Optimal hyperparameters	Accuracy output
Support Vector Classifier	C = 100 gamma = scale kernel = rbf time = 83.1 s	83.85%
Random Forest Classifier	n_estimators = 200 time = 185.3 s	71.85%
K-Nearest Neighbor Classifier	metric = manhattan n_neighbors = 9 weights = distance time = 32.7 s	73.41%

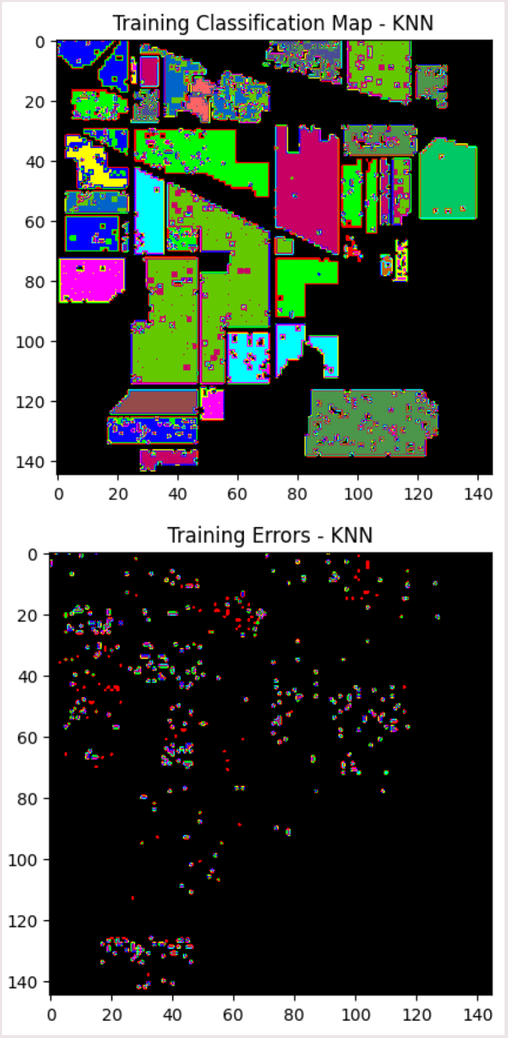
MODEL DEVELOPMENT

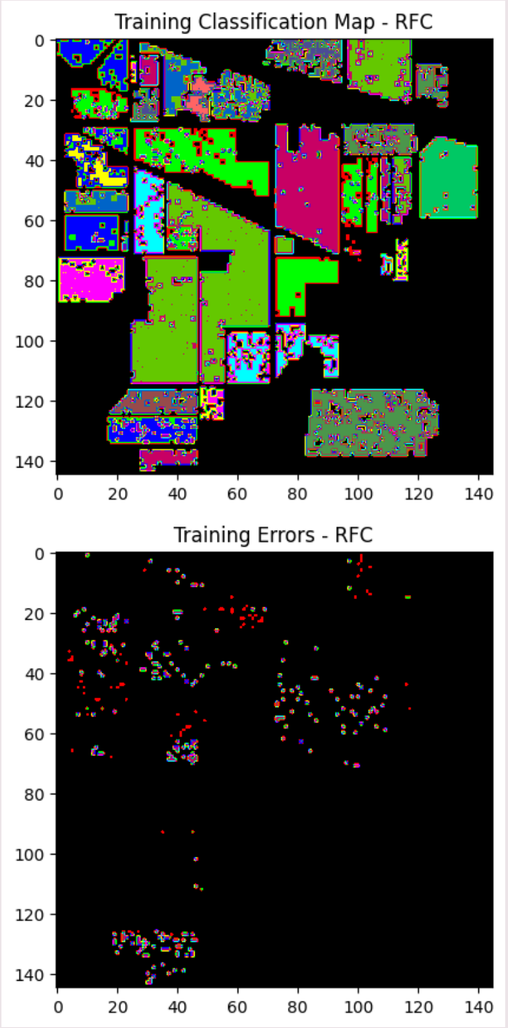
SALINAS

	Optimal hyperparameters	Accuracy output
Support Vector Classifier	C = 1000 kernel = rbf score = .920 time = 6.0 mins	91.98%
Random Forest Classifier	n_estimators = 230 time = 4.0 mins	91.95%
K-Nearest Neighbor Classifier	metric = manhattan n_neighbors = 11 weights = distance time = 6.5 mins	90.75%

INDIAN PINES RESULTS





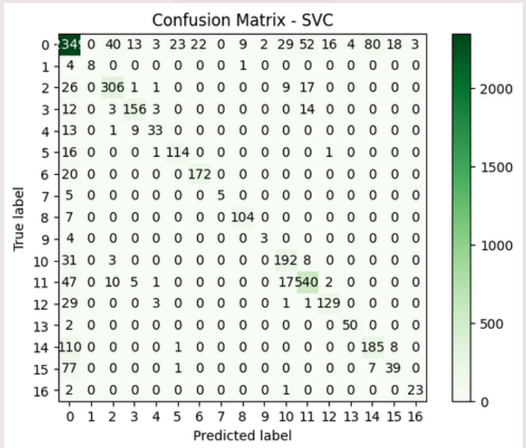


INDIAN PINES RESULTS

- **SVC** surpassed its counterparts in terms of:
  - 83.8% accuracy
  - 83% precision
  - 84% recall

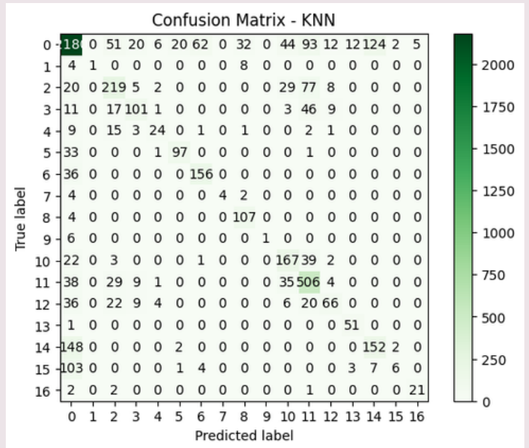
Classification Report - SVC
Accuracy: 83.85 %

	precision	recall	f1-score	support
0	0.85	0.88	0.87	2663
1	1.00	0.62	0.76	13
2	0.84	0.85	0.85	360
3	0.85	0.83	0.84	188
4	0.73	0.59	0.65	56
5	0.82	0.86	0.84	132
6	0.89	0.90	0.89	192
7	1.00	0.50	0.67	10
8	0.91	0.94	0.92	111
9	0.60	0.43	0.50	7
10	0.77	0.82	0.80	234
11	0.85	0.87	0.86	622
12	0.87	0.79	0.83	163
13	0.93	0.96	0.94	52
14	0.68	0.61	0.64	304
15	0.60	0.31	0.41	124
16	0.88	0.88	0.88	26
accuracy			0.84	5257
macro avg	0.83	0.74	0.77	5257
weighted avg	0.83	0.84	0.83	5257



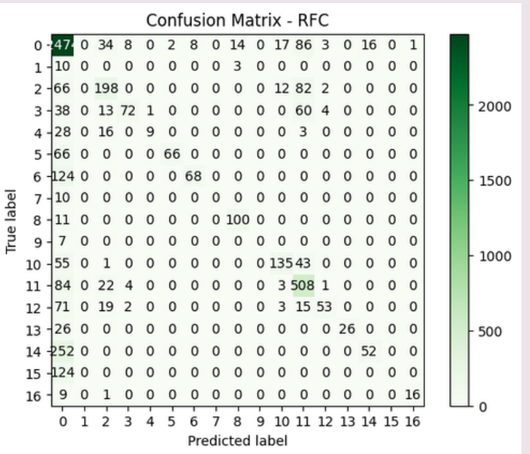
Classification Report - KNN
Accuracy: 73.41 %

	precision	recall	f1-score	support
0	0.82	0.82	0.82	2663
1	1.00	0.08	0.14	13
2	0.61	0.61	0.61	360
3	0.69	0.54	0.60	188
4	0.62	0.43	0.51	56
5	0.81	0.73	0.77	132
6	0.70	0.81	0.75	192
7	1.00	0.40	0.57	10
8	0.71	0.96	0.82	111
9	1.00	0.14	0.25	7
10	0.59	0.71	0.64	234
11	0.64	0.81	0.72	622
12	0.65	0.40	0.50	163
13	0.77	0.98	0.86	52
14	0.54	0.50	0.52	304
15	0.60	0.05	0.09	124
16	0.81	0.81	0.81	26
accuracy			0.73	5257
macro avg	0.74	0.58	0.59	5257
weighted avg	0.73	0.73	0.72	5257

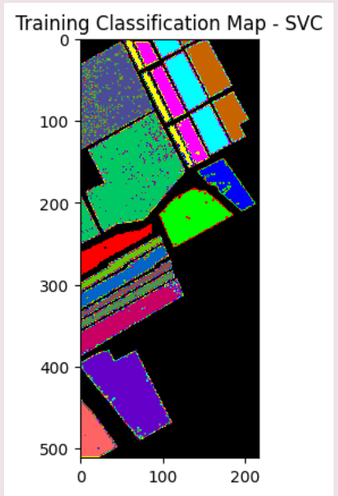


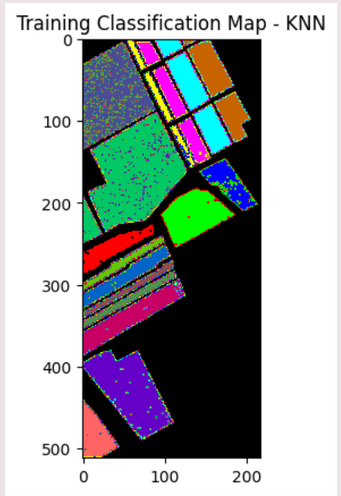
Classification Report - RFC
Accuracy: 71.85 %

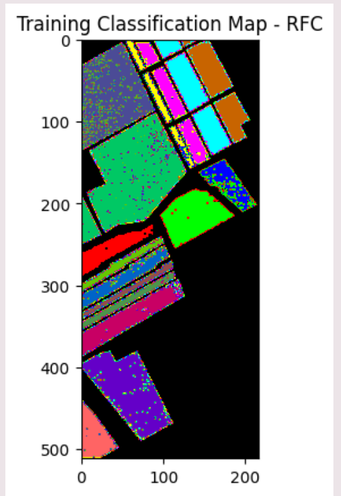
	precision	recall	f1-score	support
0	0.72	0.93	0.81	2663
1	0.00	0.00	0.00	13
2	0.65	0.55	0.60	360
3	0.84	0.38	0.53	188
4	0.90	0.16	0.27	56
5	0.97	0.50	0.66	132
6	0.89	0.35	0.51	192
7	0.00	0.00	0.00	10
8	0.85	0.90	0.88	111
9	0.00	0.00	0.00	7
10	0.79	0.58	0.67	234
11	0.64	0.82	0.72	622
12	0.84	0.33	0.47	163
13	1.00	0.50	0.67	52
14	0.76	0.17	0.28	304
15	0.00	0.00	0.00	124
16	0.94	0.62	0.74	26
accuracy			0.72	5257
macro avg	0.64	0.40	0.46	5257
weighted avg	0.72	0.72	0.68	5257

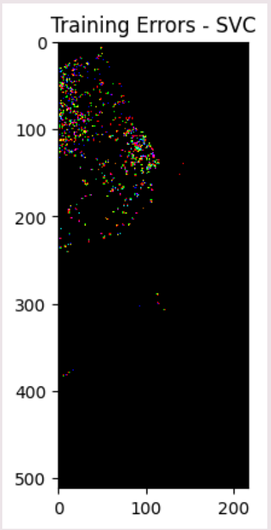


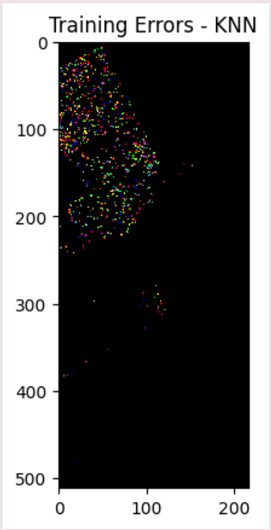
SALINAS RESULTS

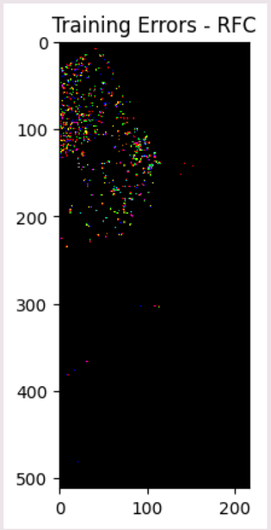












SALINAS RESULTS

There’s not much any significant difference in terms of the evaluation metrics for the Salinas dataset

Classification Report - SVC
Accuracy: 91.98 %

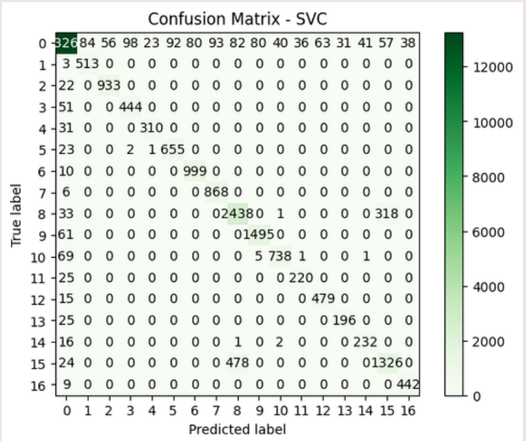
	precision	recall	f1-score	support
0	0.97	0.93	0.95	14255
1	0.86	0.99	0.92	516
2	0.94	0.98	0.96	955
3	0.82	0.90	0.85	495
4	0.93	0.91	0.92	341
5	0.88	0.96	0.92	681
6	0.93	0.99	0.96	1009
7	0.90	0.99	0.95	874
8	0.81	0.87	0.84	2790
9	0.95	0.96	0.95	1556
10	0.94	0.91	0.93	814
11	0.86	0.90	0.88	245
12	0.88	0.97	0.92	494
13	0.86	0.89	0.88	221
14	0.85	0.92	0.88	251
15	0.78	0.73	0.75	1828
16	0.92	0.98	0.95	451
accuracy			0.92	27776
macro avg	0.89	0.93	0.91	27776
weighted avg	0.92	0.92	0.92	27776

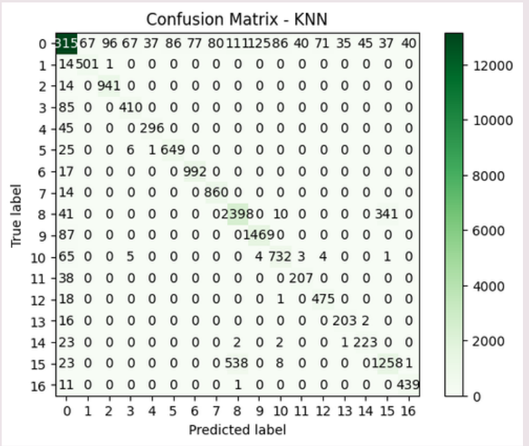
Classification Report - KNN
Accuracy: 90.75 %

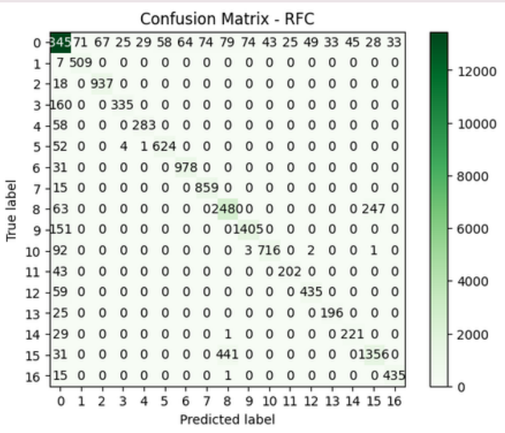
	precision	recall	f1-score	support
0	0.96	0.92	0.94	14255
1	0.88	0.97	0.92	516
2	0.91	0.99	0.94	955
3	0.84	0.83	0.83	495
4	0.89	0.87	0.88	341
5	0.88	0.95	0.92	681
6	0.93	0.98	0.95	1009
7	0.91	0.98	0.95	874
8	0.79	0.86	0.82	2790
9	0.92	0.94	0.93	1556
10	0.87	0.90	0.89	814
11	0.83	0.84	0.84	245
12	0.86	0.96	0.91	494
13	0.85	0.92	0.88	221
14	0.83	0.89	0.86	251
15	0.77	0.69	0.73	1828
16	0.91	0.97	0.94	451
accuracy			0.91	27776
macro avg	0.87	0.91	0.89	27776
weighted avg	0.91	0.91	0.91	27776

Classification Report - RFC
Accuracy: 91.55 %

	precision	recall	f1-score	support
0	0.94	0.94	0.94	14255
1	0.88	0.99	0.93	516
2	0.93	0.98	0.96	955
3	0.92	0.68	0.78	495
4	0.90	0.83	0.87	341
5	0.91	0.92	0.92	681
6	0.94	0.97	0.95	1009
7	0.92	0.98	0.95	874
8	0.83	0.89	0.86	2790
9	0.95	0.90	0.92	1556
10	0.94	0.88	0.91	814
11	0.89	0.82	0.86	245
12	0.90	0.88	0.89	494
13	0.86	0.89	0.87	221
14	0.83	0.88	0.85	251
15	0.83	0.74	0.78	1828
16	0.93	0.96	0.95	451
accuracy			0.92	27776
macro avg	0.90	0.89	0.89	27776
weighted avg	0.92	0.92	0.91	27776







Conclusion

- **Scaling and Dimensionality Reduction Via PCA**
  - Indian Pines: 200 bands to 108 bands
  - Salinas: 204 bands to 19 bands
- **Hyperparameter Tuning via Grid Search**
- **Training Errors**
  - KNN for both Indian Pines and Salinas dataset
- **Evaluation metrics**
  - SVC performed best in the classification of Indian Pines dataset
  - No classifier is significantly better than others in the classification of Salinas dataset