Machine Learning and Multispectral Unmanned Aerial Vehicle Imagery Data for Agriculture.

Georgios Batsis
https://github.com/gbatsis/VYSegML

Machine Learning Course @ MSc Program of AI - NCSR Demokritos & University of Piraeus.

Image Segmentation

Image segmentation → Image is broken down into subgroups (Segments) → Assigning labels to pixels

Image segmentation – Deep Leaning →

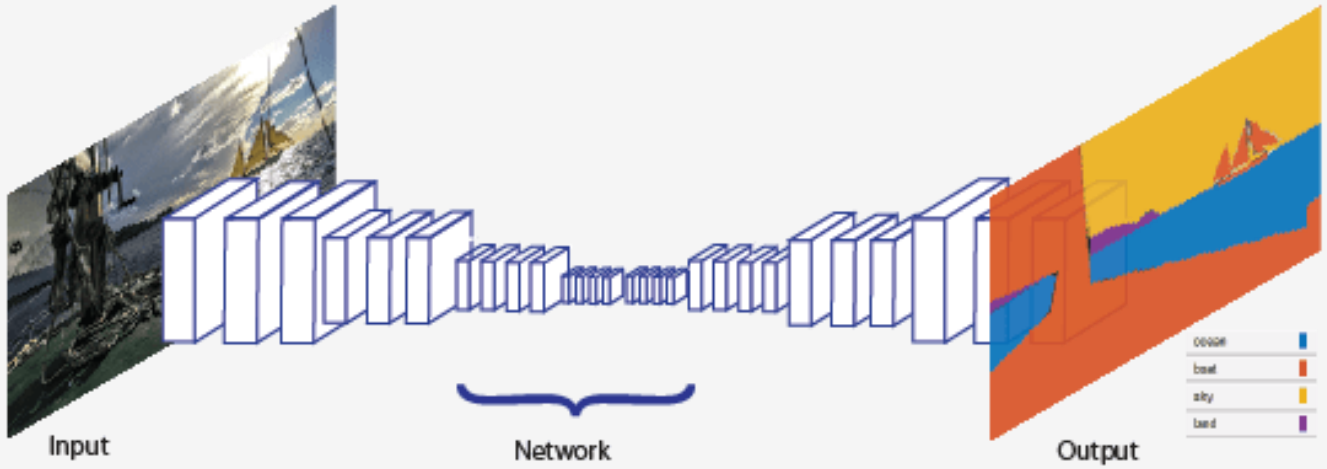
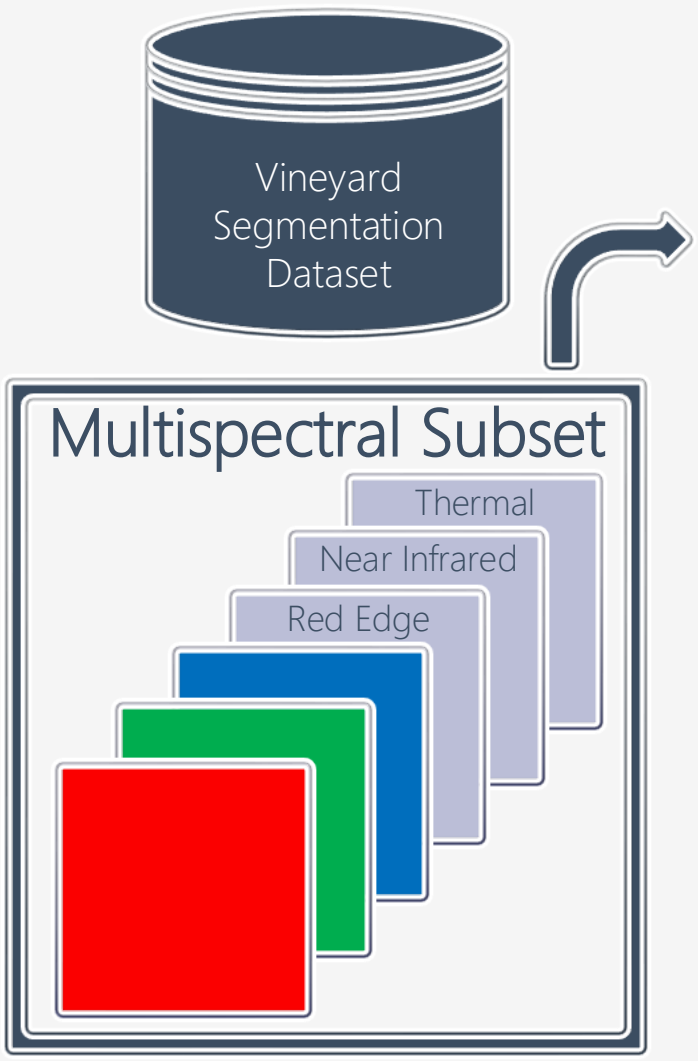
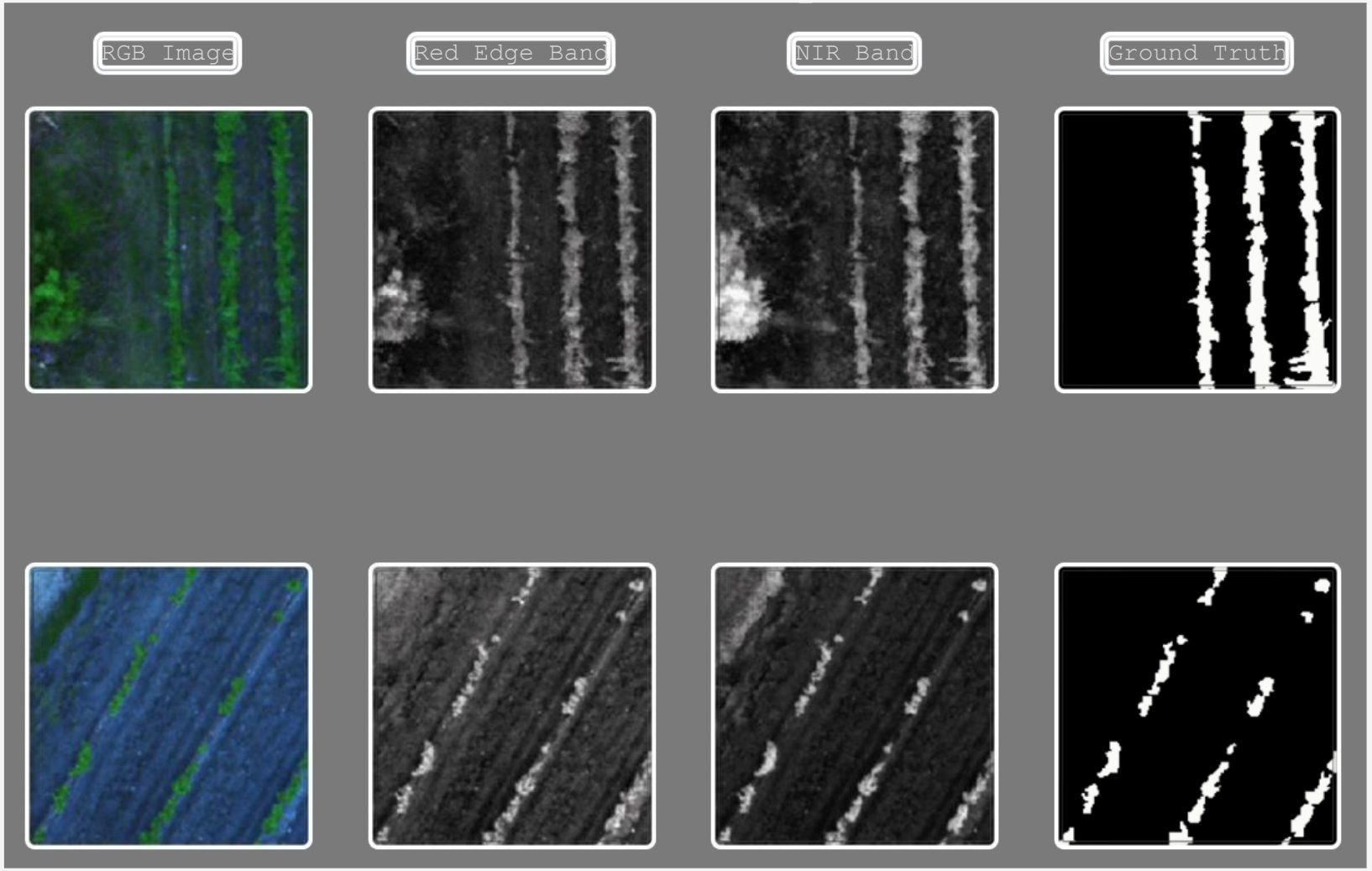


Image segmentation – Machine (Non Deep) Leaning → Direct Pixel classification using informative Features.

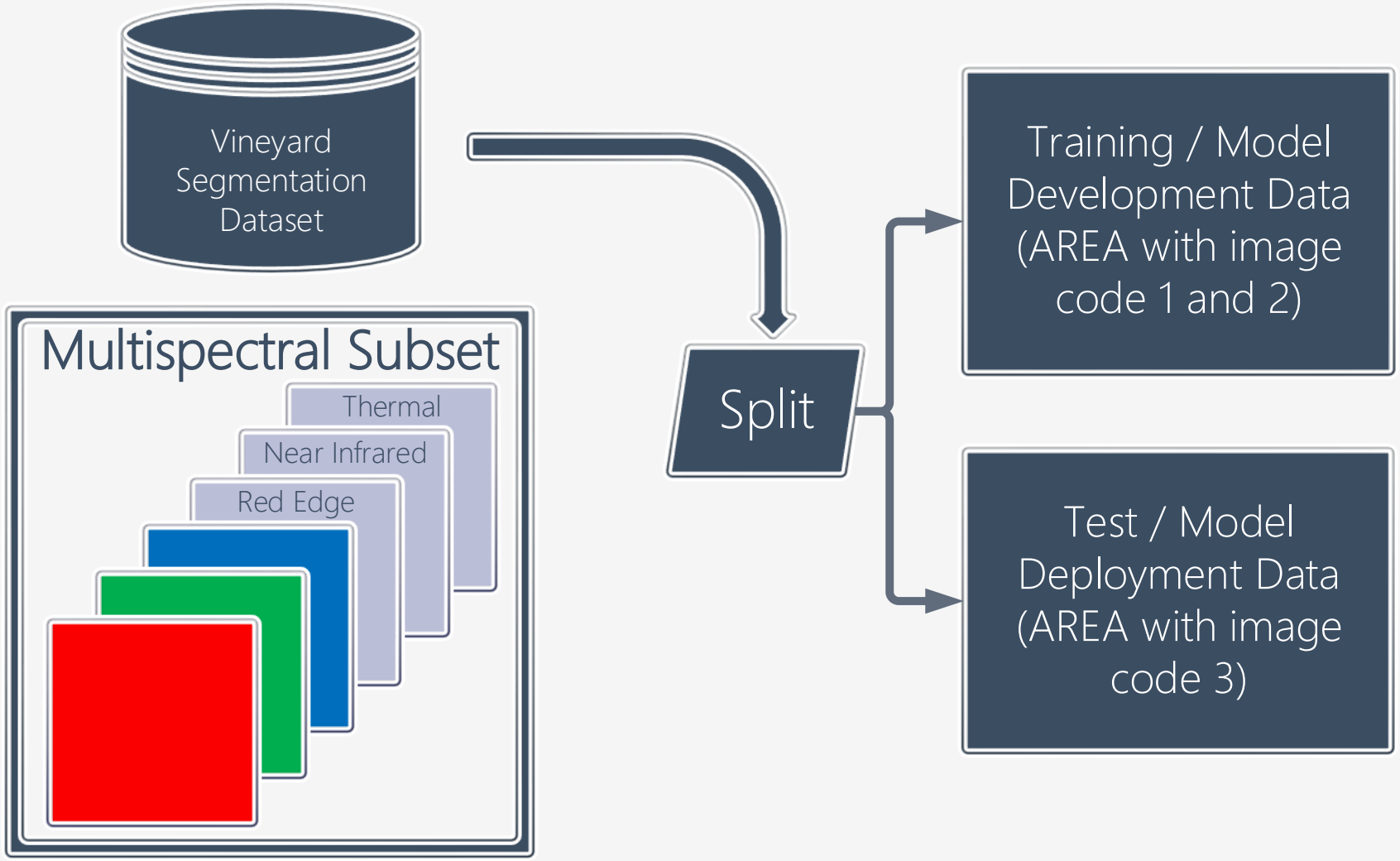
Dataset





https://github.com/Cybonic/DL_vineyard_segmentation_study

Dataset



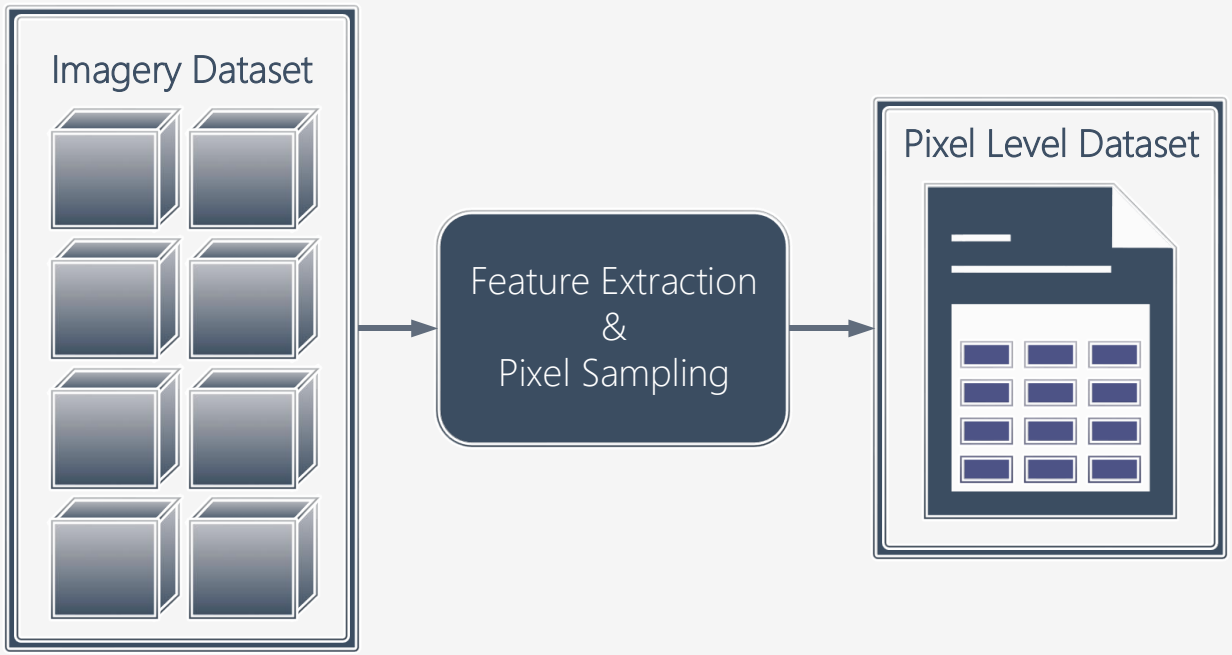
https://github.com/Cybonic/DL_vineyard_segmentation_study.git

Changing Form of Dataset

Image segmentation – Machine (Non Deep) Leaning → Direct Pixel classification using informative Features.

It is essential that we should change the form of Dataset to fit Machine Learning Classifiers and to perform Pixel Level Classification.

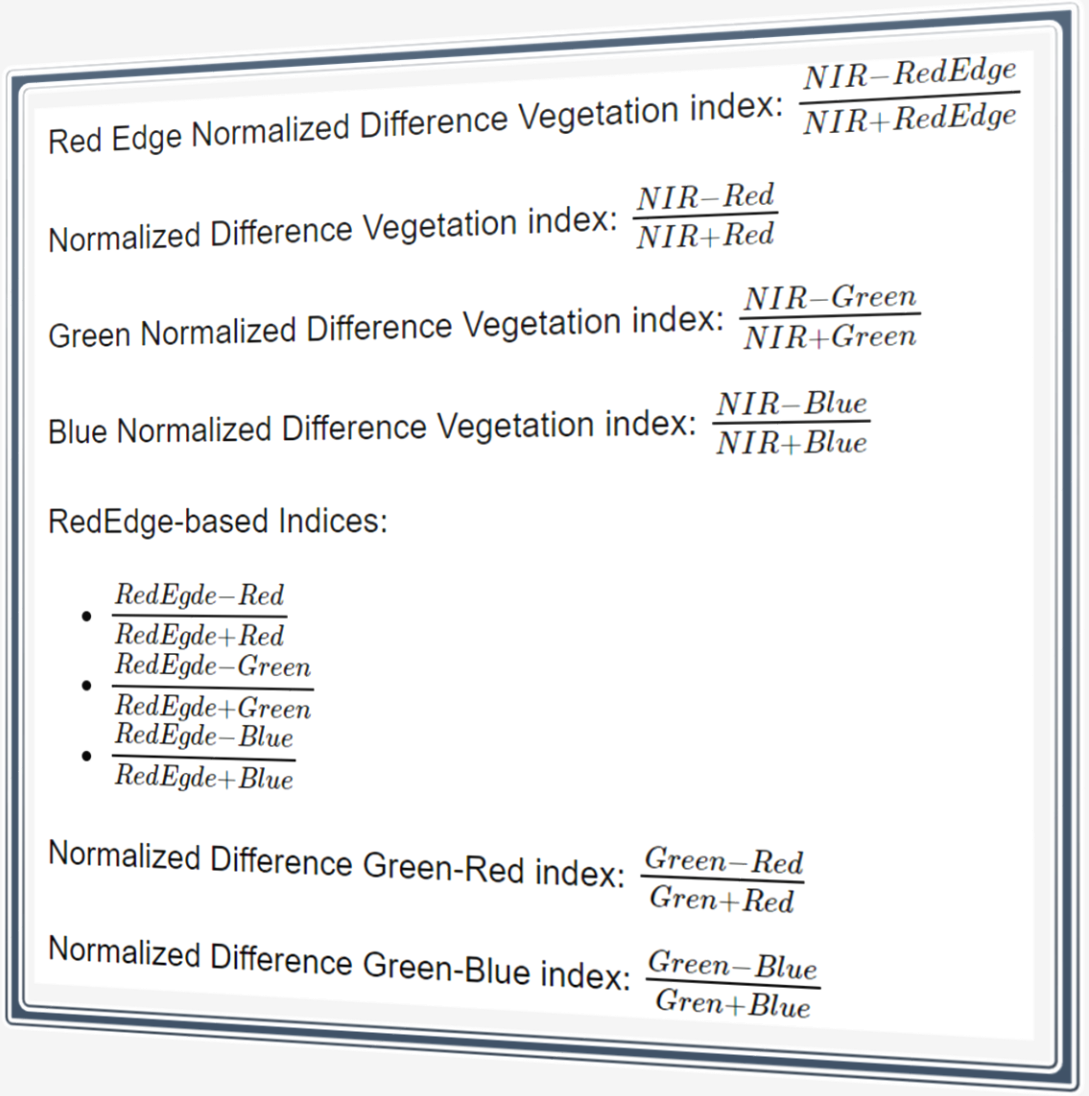
Extract a Pixel – Level Dataset from original Imagery Dataset.



Feature Extraction

Vegetation Indices

- *A Vegetation Index (VI) is a mathematical combination of several spectral Bands which provides information about relationship between vegetation or crop patterns.*
- *A Vegetation Index (VI) is a spectral transformation metric for measuring the presence and state of vegetation.*
- After its calculation, VI simulates a single channel image.



Feature Extraction

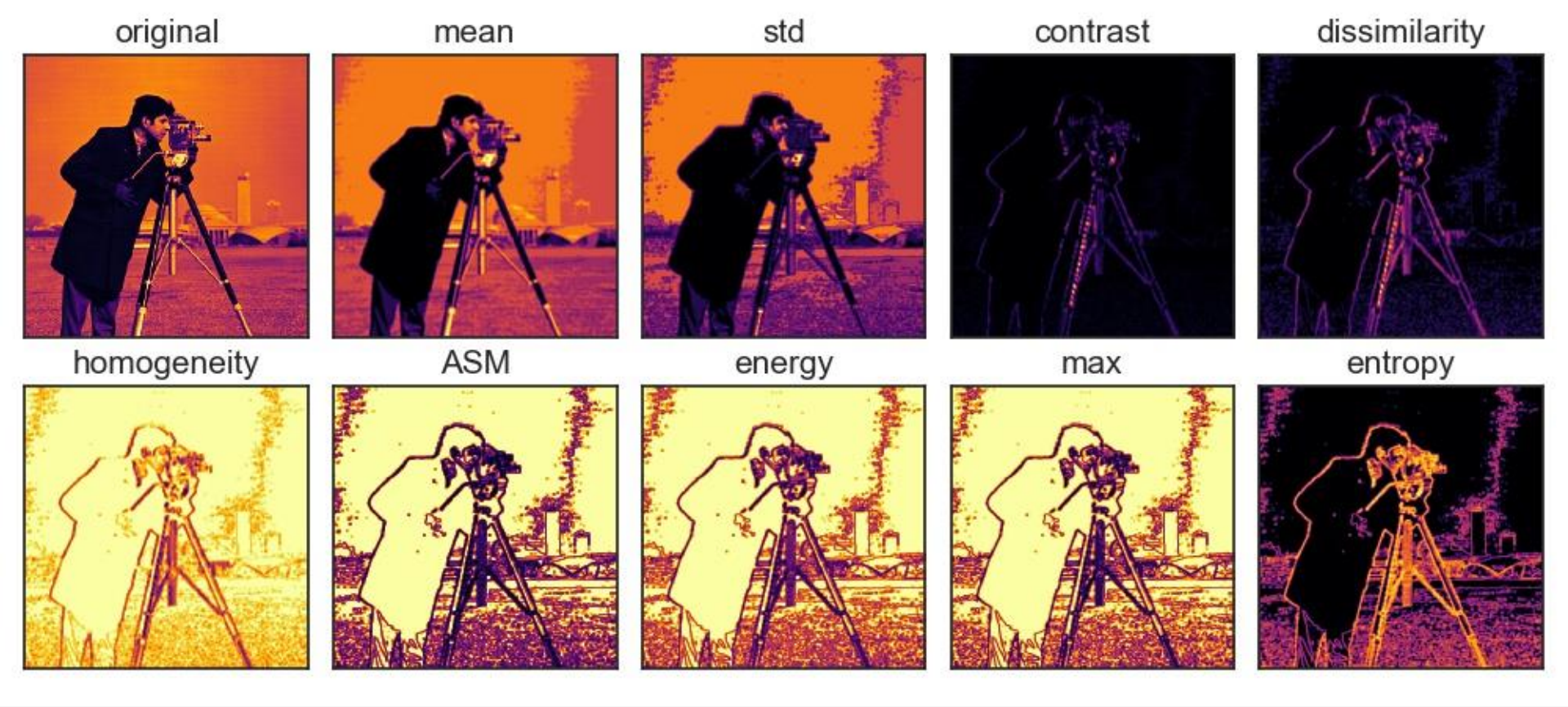
Texture Information

- ***Gray Level Co-occurrence Matrix:*** *Tabulation of how often pairs of pixel with specific values and in a specified spatial relationship occur in an image.*
- GLCM contains information about the relationship of intensity of a pixel and its neighborhood defined using a window.
- Texture Information are statistical measures extracted from the aforementioned matrix.
- Each measure simulates a single pixel image.

- Mean
- Standard Deviation
- Max
- Entropy
- Contrast = $\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2$
- Homogeneity = $\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i,j)}{1+(i-j)^2}$
- Angular second Moment (ASM) $\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)^2$
- Energy $\sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)^2}$

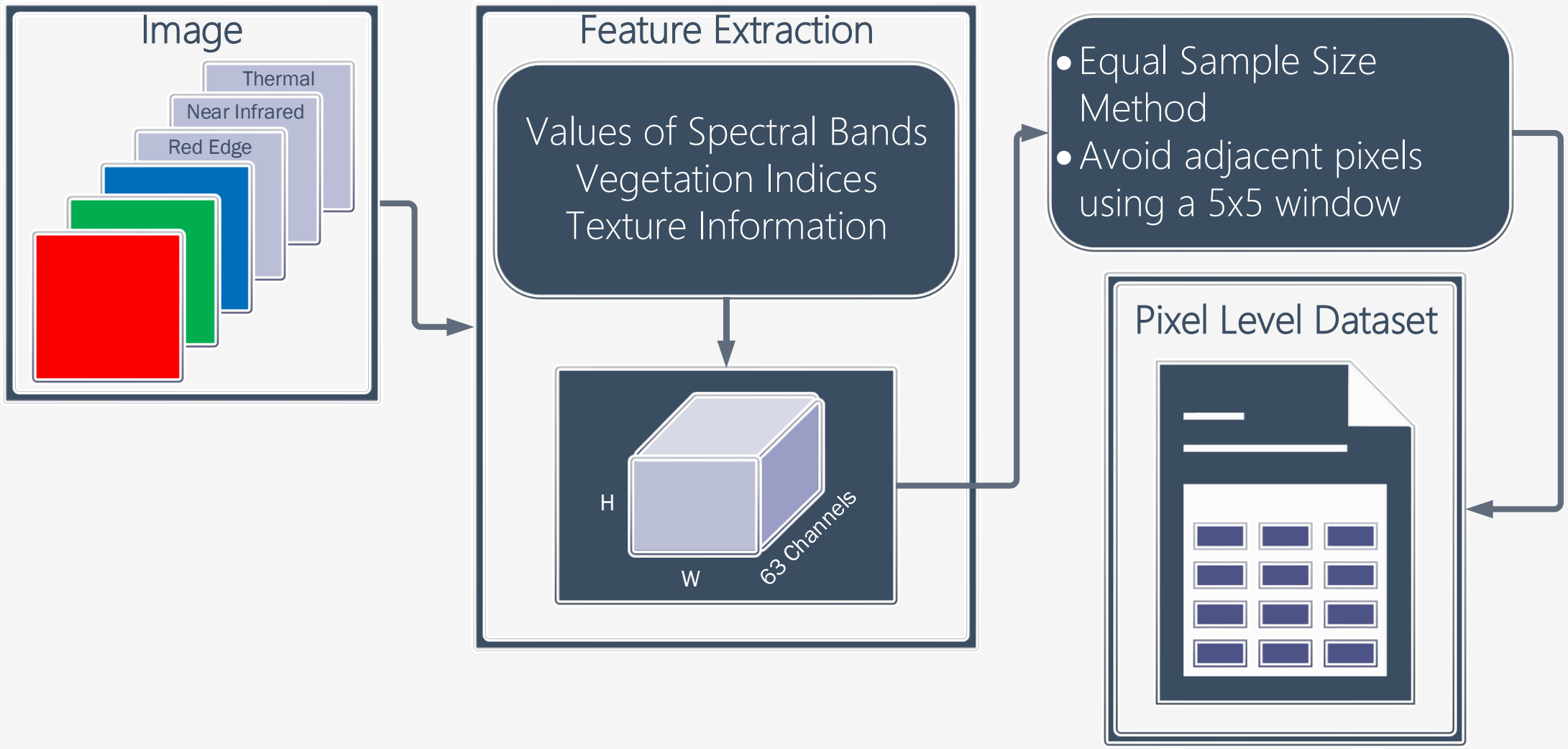
Feature Extraction

Texture Information

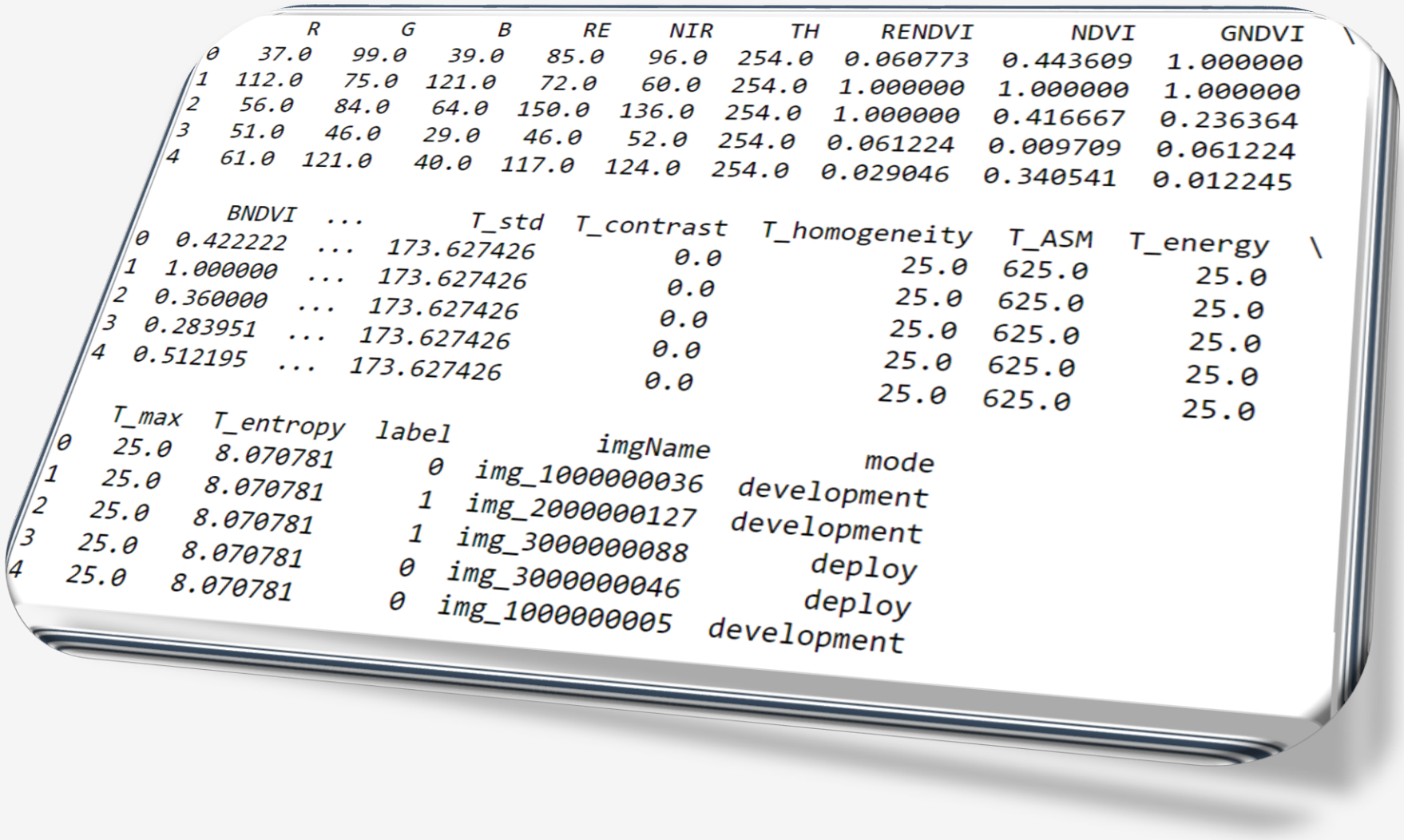


https://github.com/tzm030329/GLCM

Changing Form of Dataset using Feature Extraction & Sampling



Pixel Classification Dataset



	R	G	B	RE	NIR	TH	RENDVI	NDVI	GNDVI
0	37.0	99.0	39.0	85.0	96.0	254.0	0.060773	0.443609	1.000000
1	112.0	75.0	121.0	72.0	60.0	254.0	1.000000	1.000000	1.000000
2	56.0	84.0	64.0	150.0	136.0	254.0	1.000000	0.416667	0.236364
3	51.0	46.0	29.0	46.0	52.0	254.0	0.061224	0.009709	0.061224
4	61.0	121.0	40.0	117.0	124.0	254.0	0.029046	0.340541	0.012245

	BNDVI	...	T_std	T_contrast	T_homogeneity	T_ASM	T_energy
0	0.422222	...	173.627426	0.0	25.0	625.0	25.0
1	1.000000	...	173.627426	0.0	25.0	625.0	25.0
2	0.360000	...	173.627426	0.0	25.0	625.0	25.0
3	0.283951	...	173.627426	0.0	25.0	625.0	25.0
4	0.512195	...	173.627426	0.0	25.0	625.0	25.0

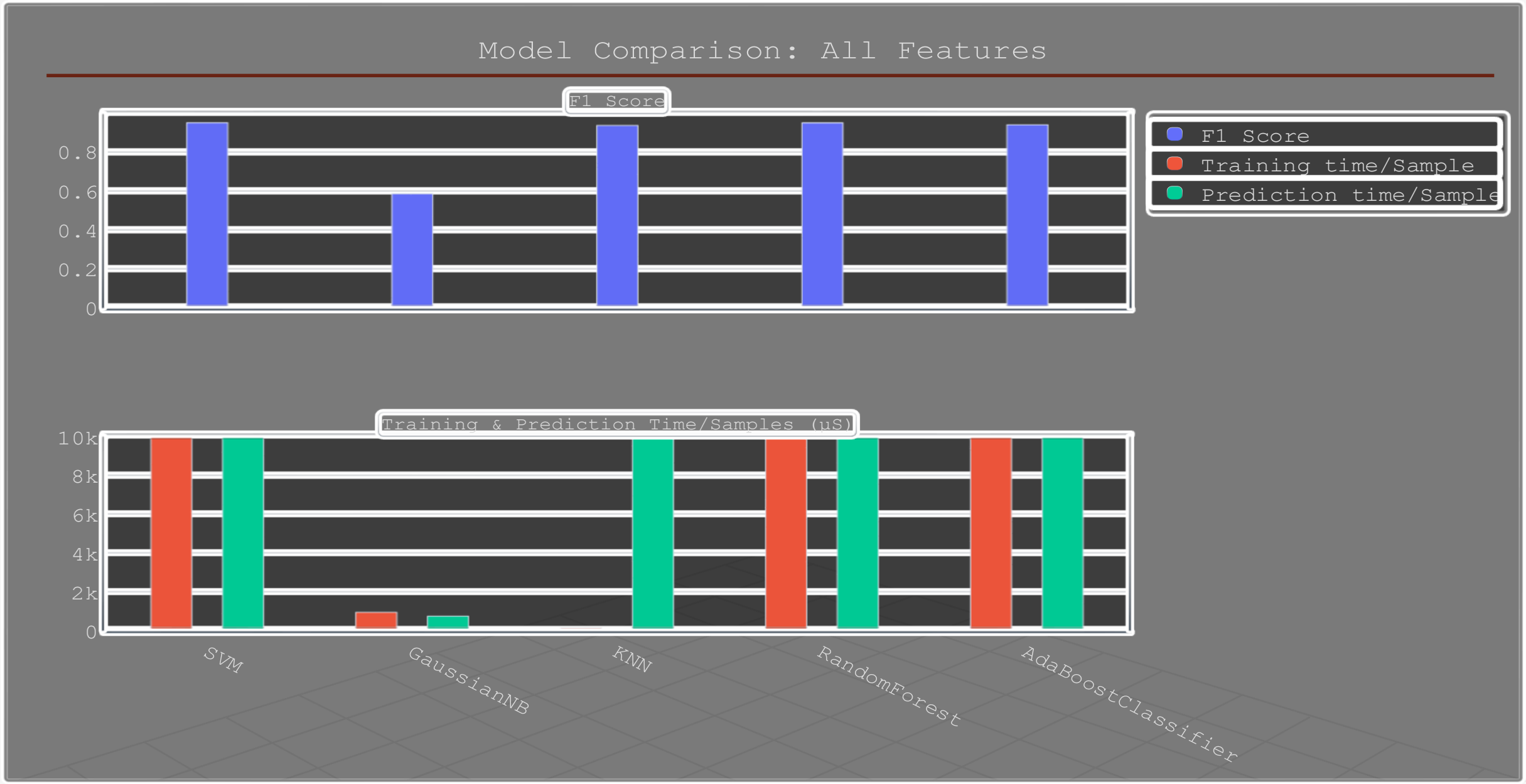
	T_max	T_entropy	label	imgName	mode
0	25.0	8.070781	0	img_1000000036	development
1	25.0	8.070781	1	img_2000000127	development
2	25.0	8.070781	1	img_3000000088	deploy
3	25.0	8.070781	0	img_3000000046	deploy
4	25.0	8.070781	0	img_1000000005	development

Comparison of Different Classifiers

Comparison of baseline models in terms of F1 Score, training and prediction time.

Machine Learning Models: SVM-RBF, Gaussian Naïve Bayes, K-Nearest Neighbor, Random Forest, AdaBoost

- *Comparison Using All Features*
  - Implementation of Feature Selection based on importance weights during the training process of Random Forest and AdaBoost.

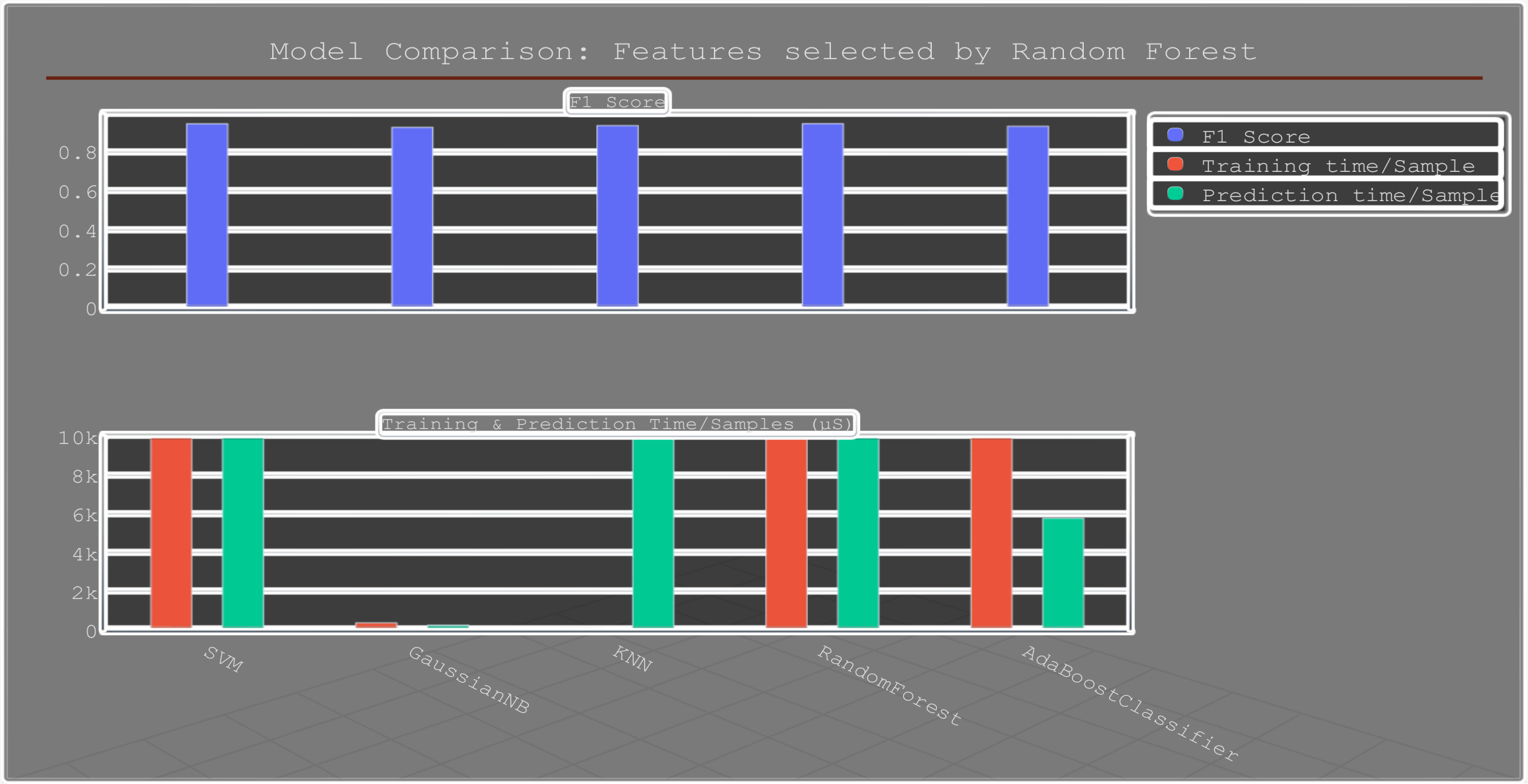

Model Comparison: All Features
F1 Score
0.8
0.6
0.4
0.2
0
F1 Score
Training time/Sample
Prediction time/Sample
Training & Prediction Time/Samples (µS)
10k
8k
6k
4k
2k
0
SVM
GaussianNB
KNN
RandomForest
AdaBoostClassifier

Comparison of Different Classifiers

Comparison of baseline models in terms of F1 Score, training and prediction time.

Machine Learning Models: SVM-RBF, Gaussian Naïve Bayes, K-Nearest Neighbor, Random Forest, AdaBoost

- ~~*Comparison Using All Features*~~
- *Comparison Using Features Selected by Random Forest*

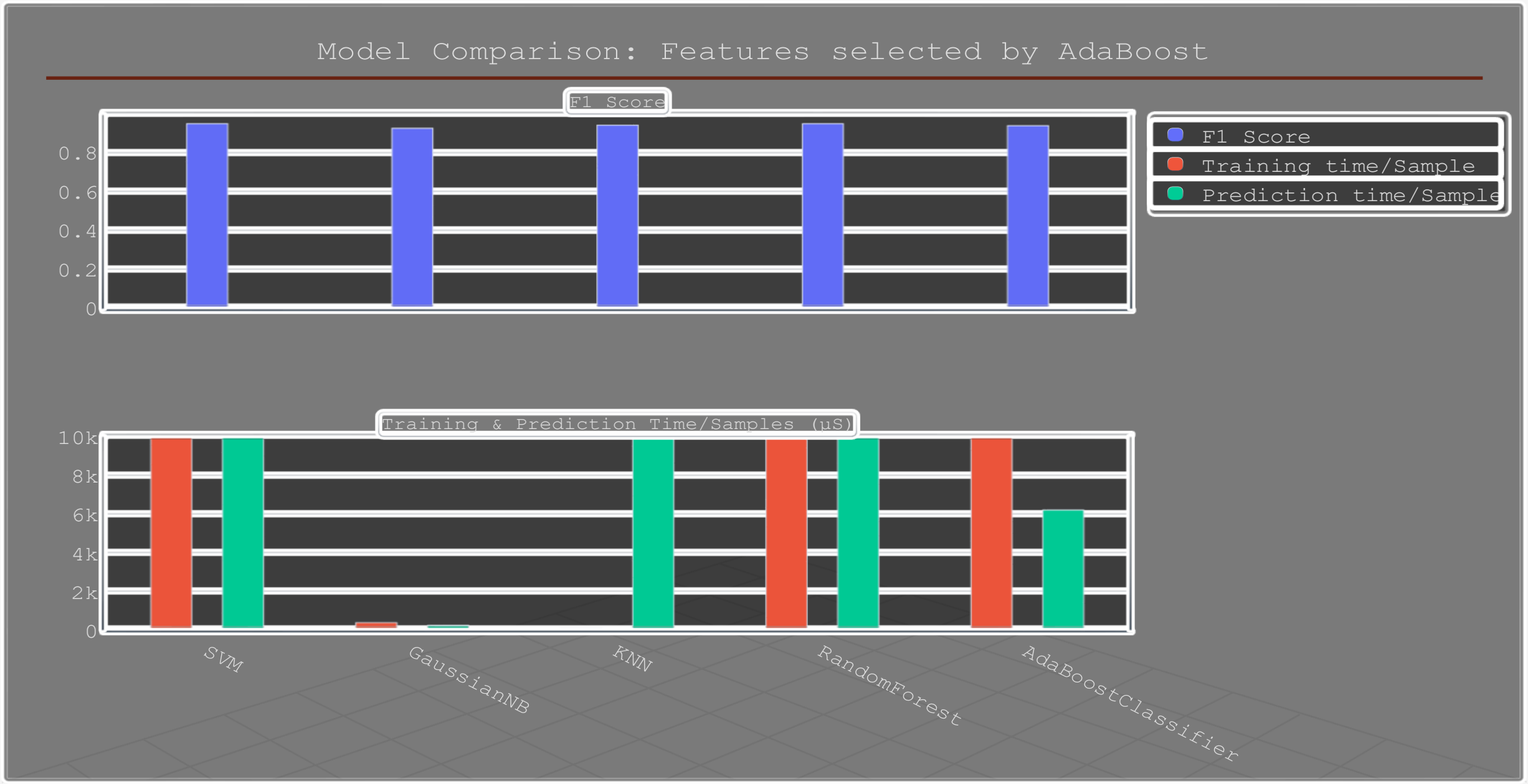

Model Comparison: Features selected by Random Forest
F1 Score
0.8
0.6
0.4
0.2
0
F1 Score
Training time/Sample
Prediction time/Sample
Training & Prediction Time/Samples (µS)
10k
8k
6k
4k
2k
0
SVM
GaussianNB
KNN
RandomForest
AdaBoostClassifier

Comparison of Different Classifiers

Comparison of baseline models in terms of F1 Score, training and prediction time.

Machine Learning Models: SVM-RBF, Gaussian Naïve Bayes, K-Nearest Neighbor, Random Forest, AdaBoost

- ~~*Comparison Using All Features*~~
- ~~*Comparison Using Features Selected by Random Forest*~~
- *Comparison Using Features Selected by AdaBoost*

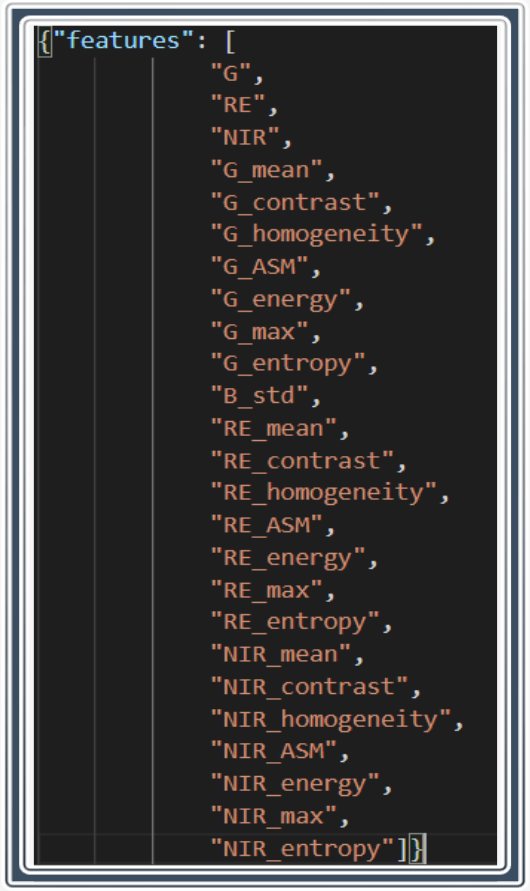

Model Comparison: Features selected by AdaBoost
F1 Score
0.8
0.6
0.4
0.2
0
F1 Score
Training time/Sample
Prediction time/Sample
Training & Prediction Time/Samples (µS)
10k
8k
6k
4k
2k
0
SVM
GaussianNB
KNN
RandomForest
AdaBoostClassifier

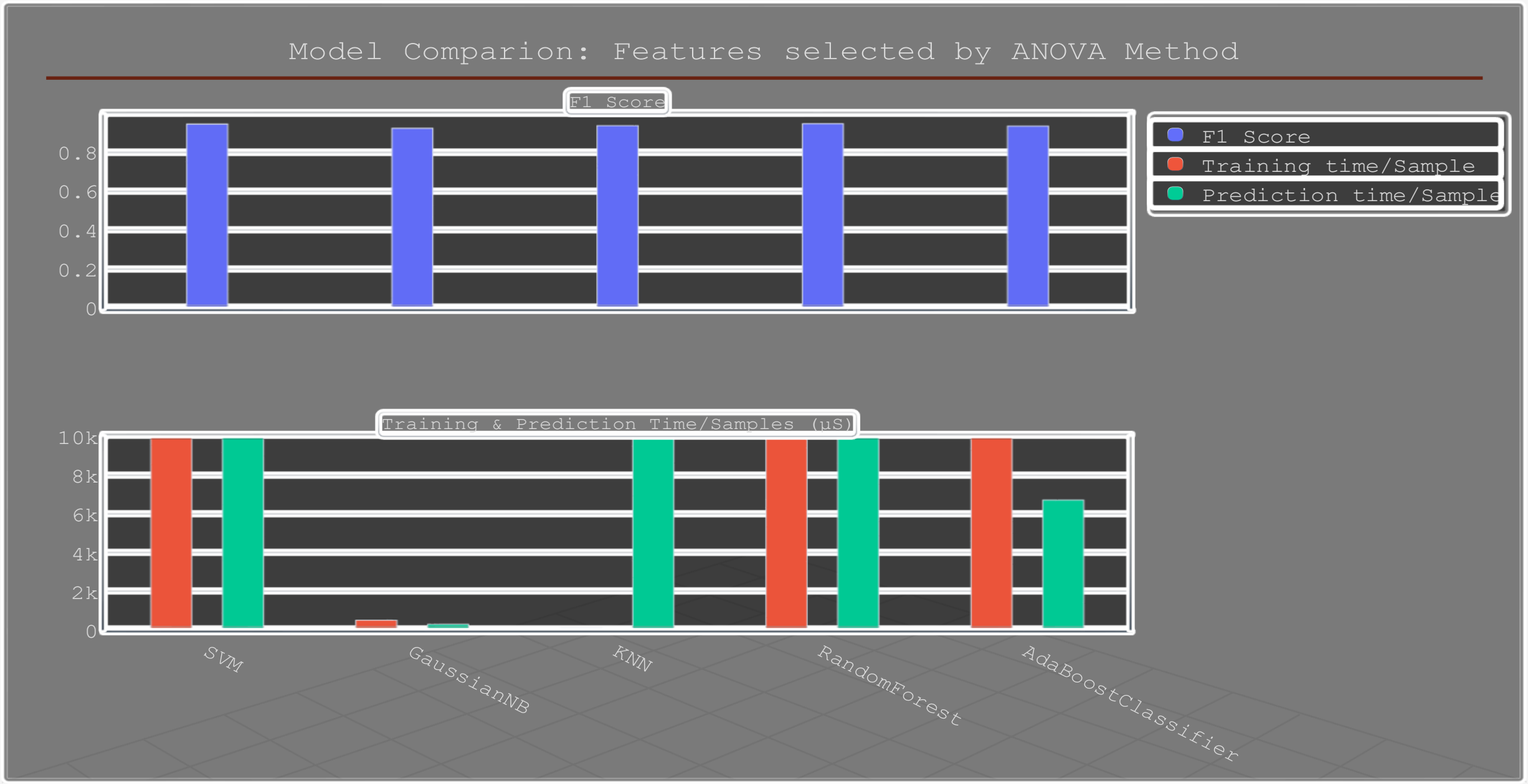
Comparison of Different Classifiers

Comparison of baseline models in terms of F1 Score, training and prediction time.

Machine Learning Models: SVM-RBF, Gaussian Naïve Bayes, K-Nearest Neighbor, Random Forest, AdaBoost

- ~~*Comparison Using All Features*~~
- ~~*Comparison Using Features Selected by Random Forest*~~
- ~~*Comparison Using Features Selected by AdaBoost*~~
- *Comparison Using Features Selected by ANOVA statistical Test*




Model Comparion: Features selected by ANOVA Method
F1 Score
0.8
0.6
0.4
0.2
0
F1 Score
Training time/Sample
Prediction time/Sample
Training & Prediction Time/Samples (µS)
10k
8k
6k
4k
2k
0
SVM
GaussianNB
KNN
RandomForest
AdaBoostClassifier

Comparison of Different Classifiers

Comparison of baseline models in terms of F1 Score, training and prediction time.

Machine Learning Models: SVM-RBF, Gaussian Naïve Bayes, K-Nearest Neighbor, Random Forest, AdaBoost

- ~~*Comparison Using All Features*~~
- ~~*Comparison Using Features Selected by Random Forest*~~
- ~~*Comparison Using Features Selected by AdaBoost*~~
- ~~*Comparison Using Features Selected by ANOVA statistical Test*~~

For this task, training dataset is used by performing a random split.

Data Preprocessing: Standardization → $x' = \frac{x - \mu}{\sigma}$

Training of Selected Classifier

Selection: Gaussian Naïve Bayes combined with Features Selected By Random Forest.

How well a classifier generalizes ?

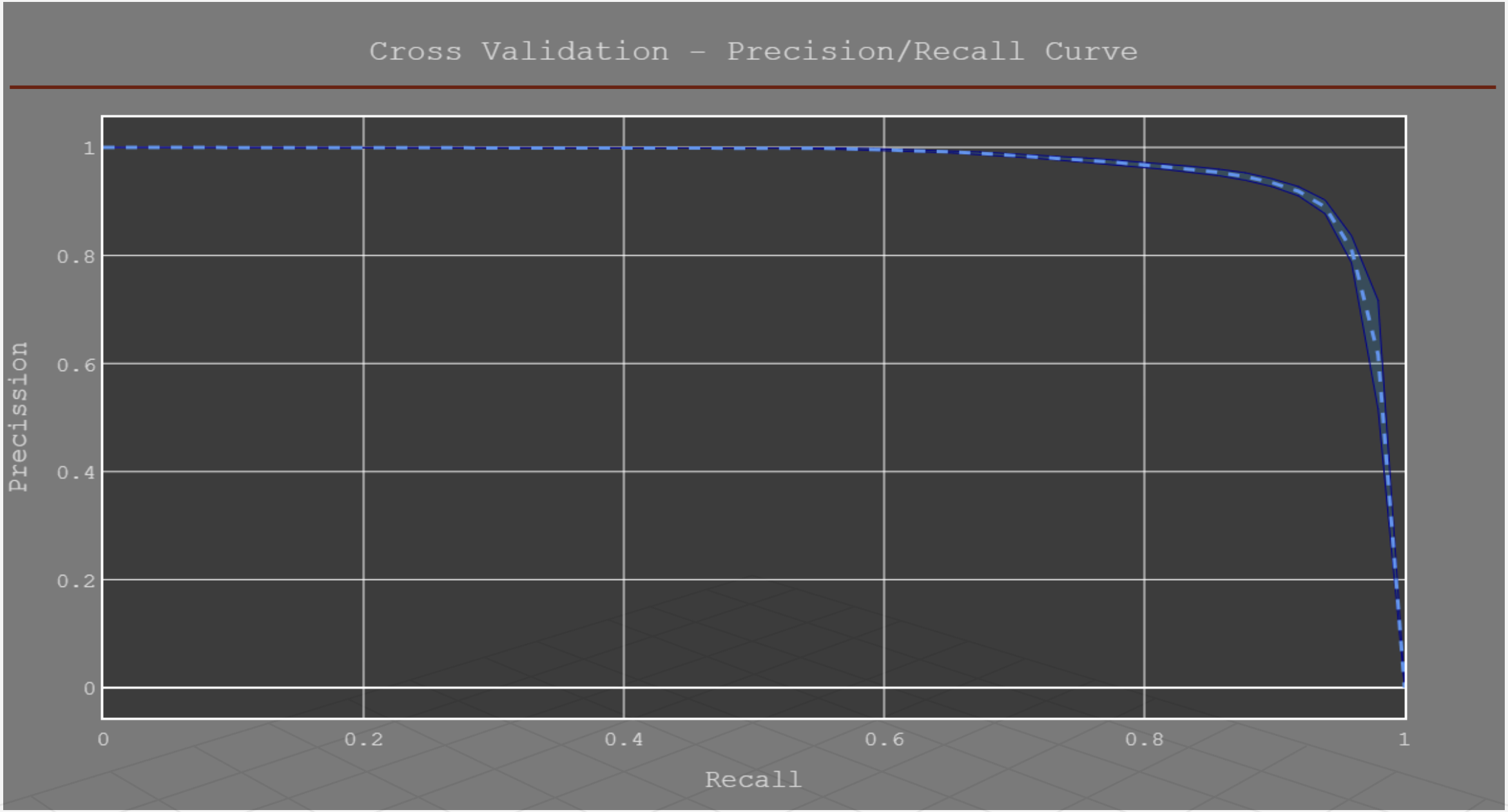
Which is the range of expected errors of the classifier ?

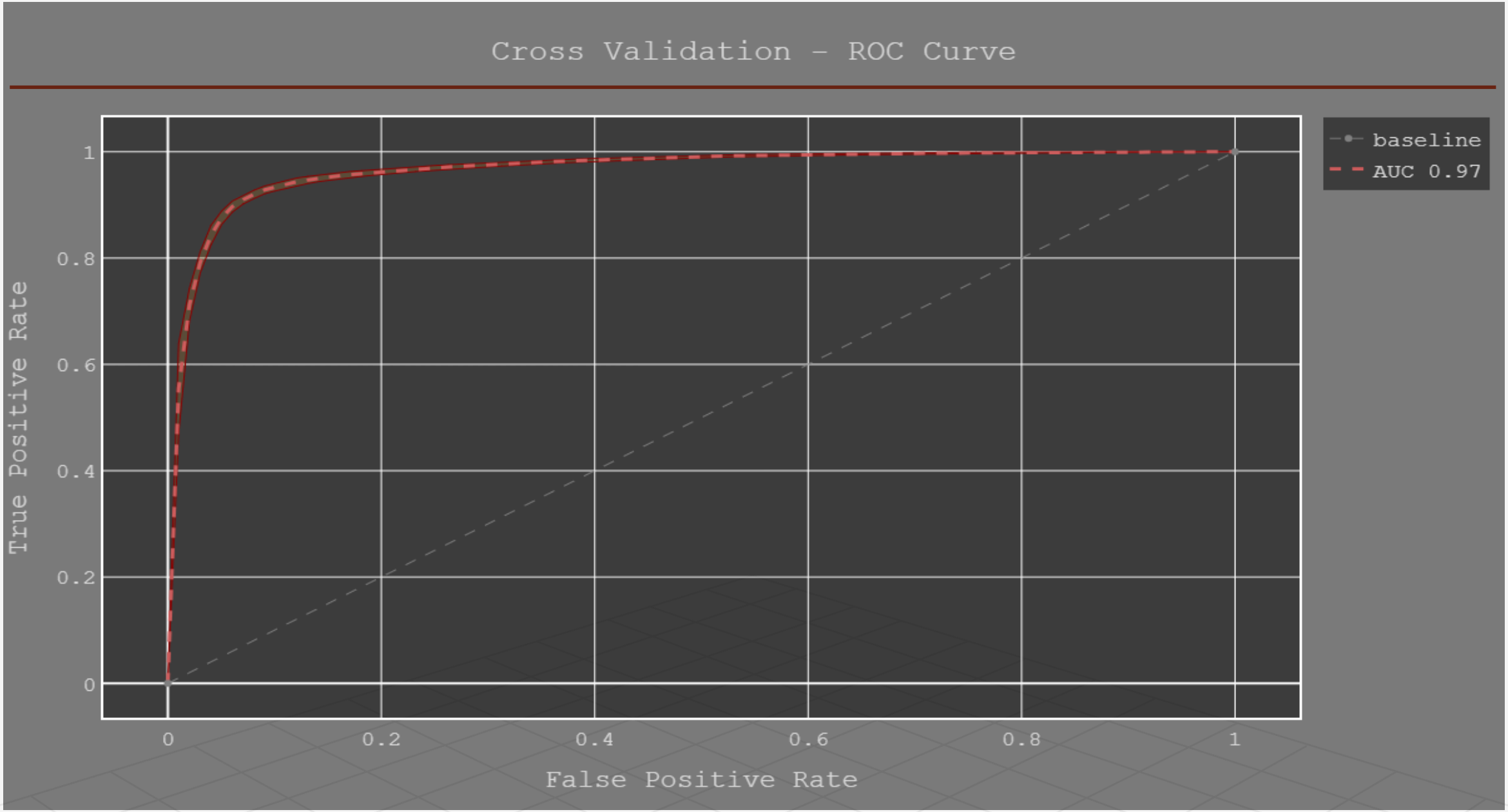
} Cross Validation

Mean F1: 0.918 | Sigma F1: 0.003

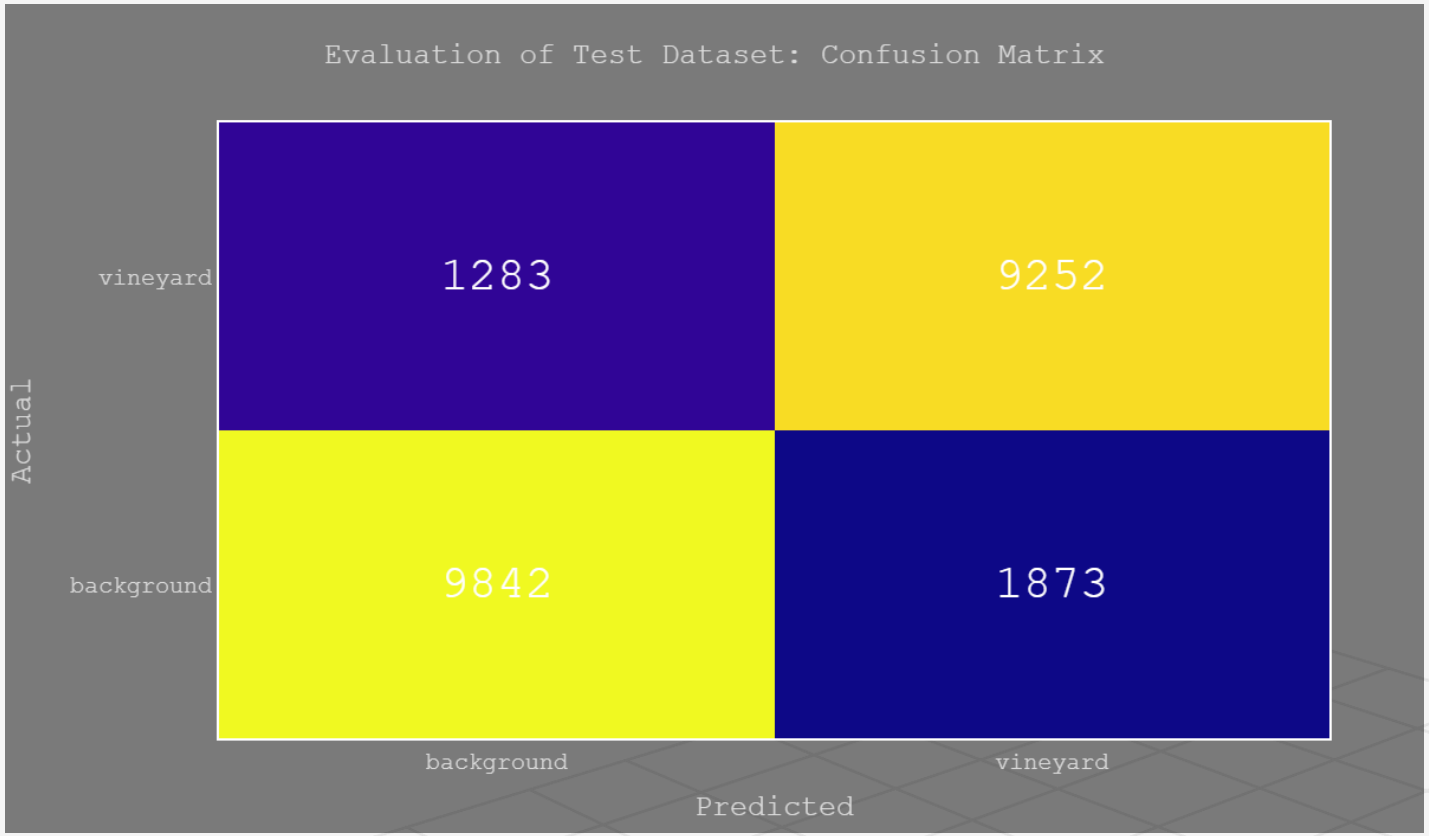
Selection of decision threshold → max(**TPR−FPR**)

Finally, the entire dataset used to train the model for the last time.

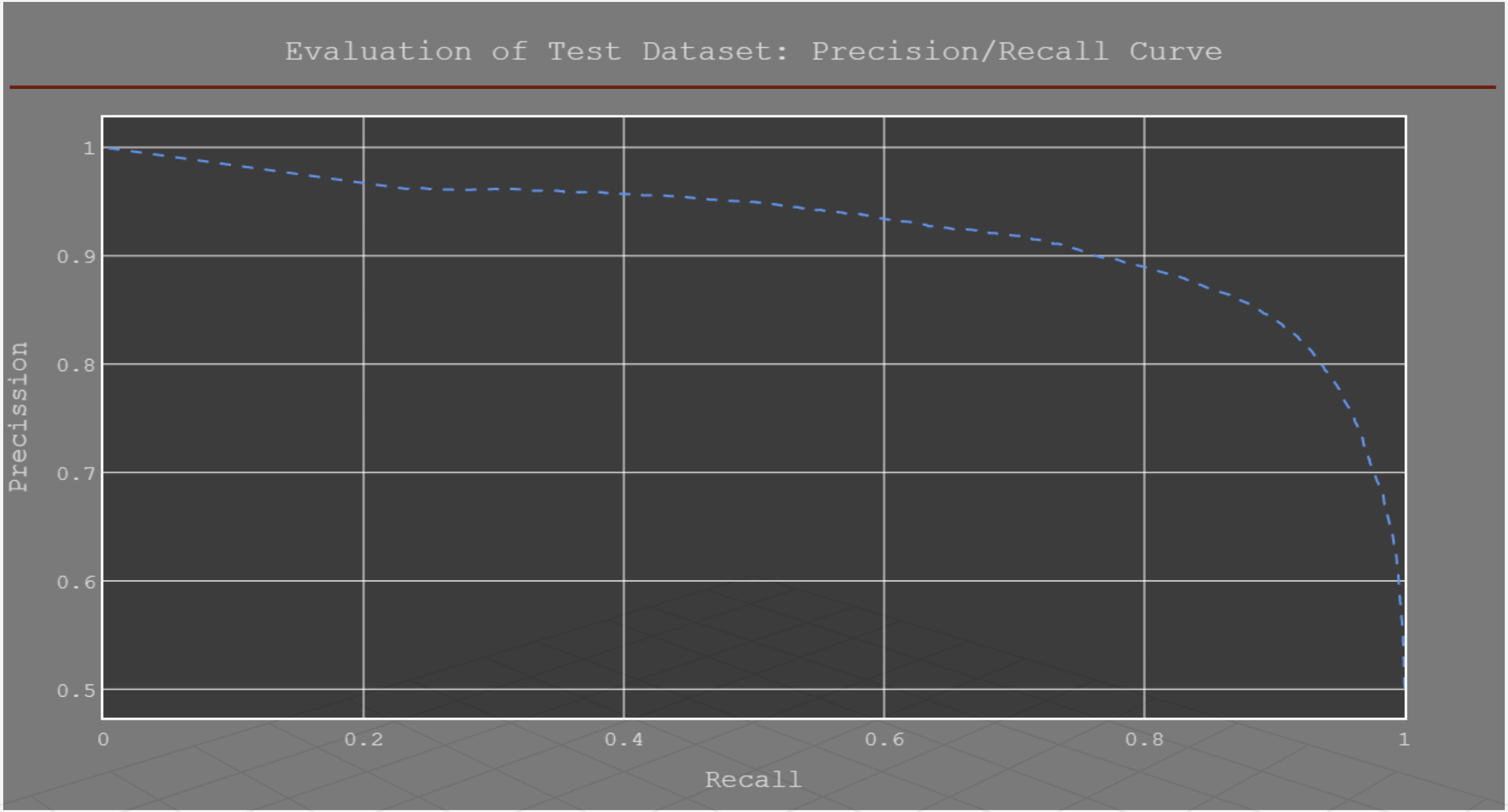

Cross Validation - Precision/Recall Curve
Precission
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
Recall

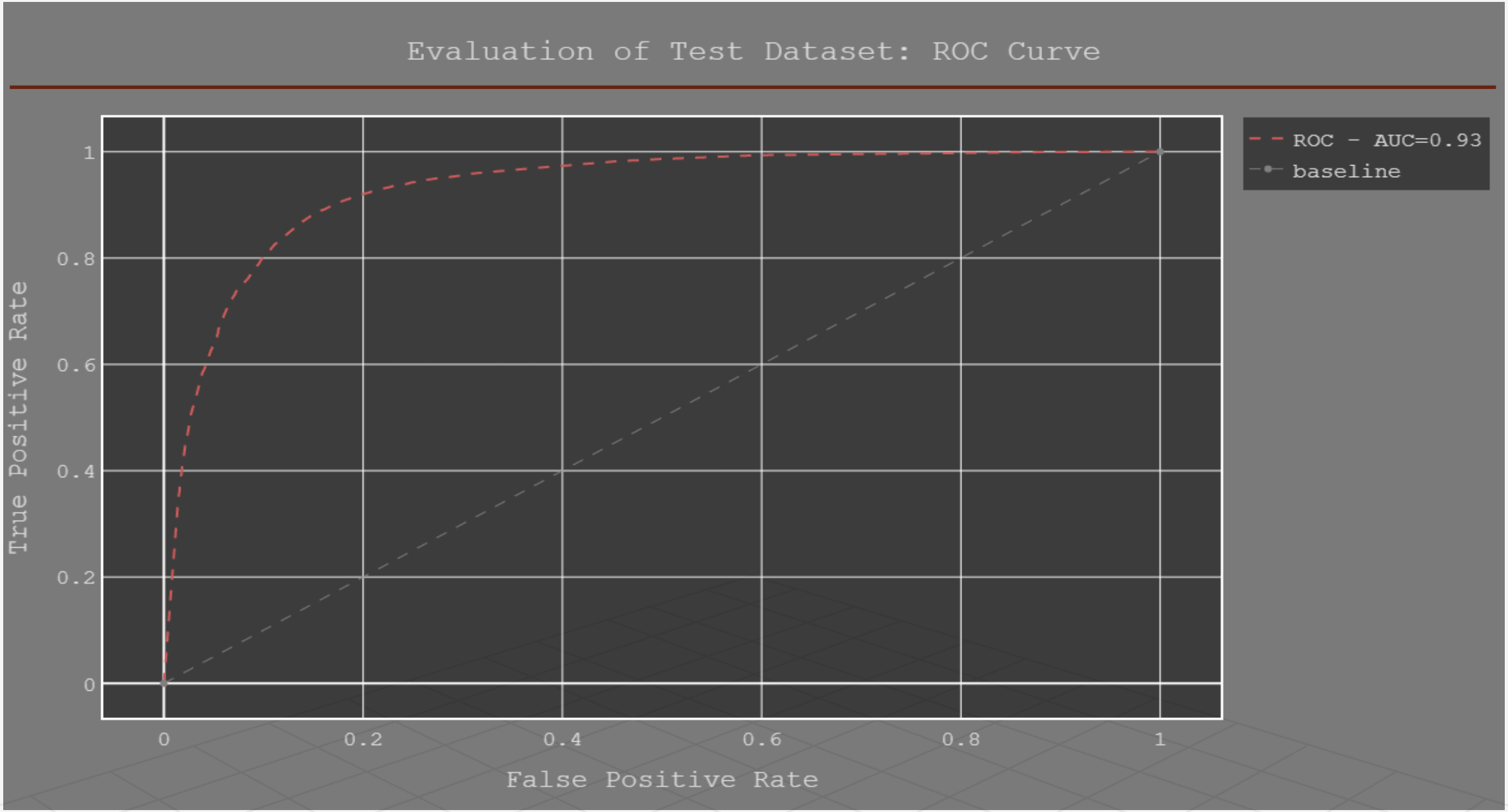

Cross Validation - ROC Curve
True Positive Rate
False Positive Rate
0
0.2
0.4
0.6
0.8
1
baseline
AUC 0.97

Using Classifier to predict unknown data (Area 3)

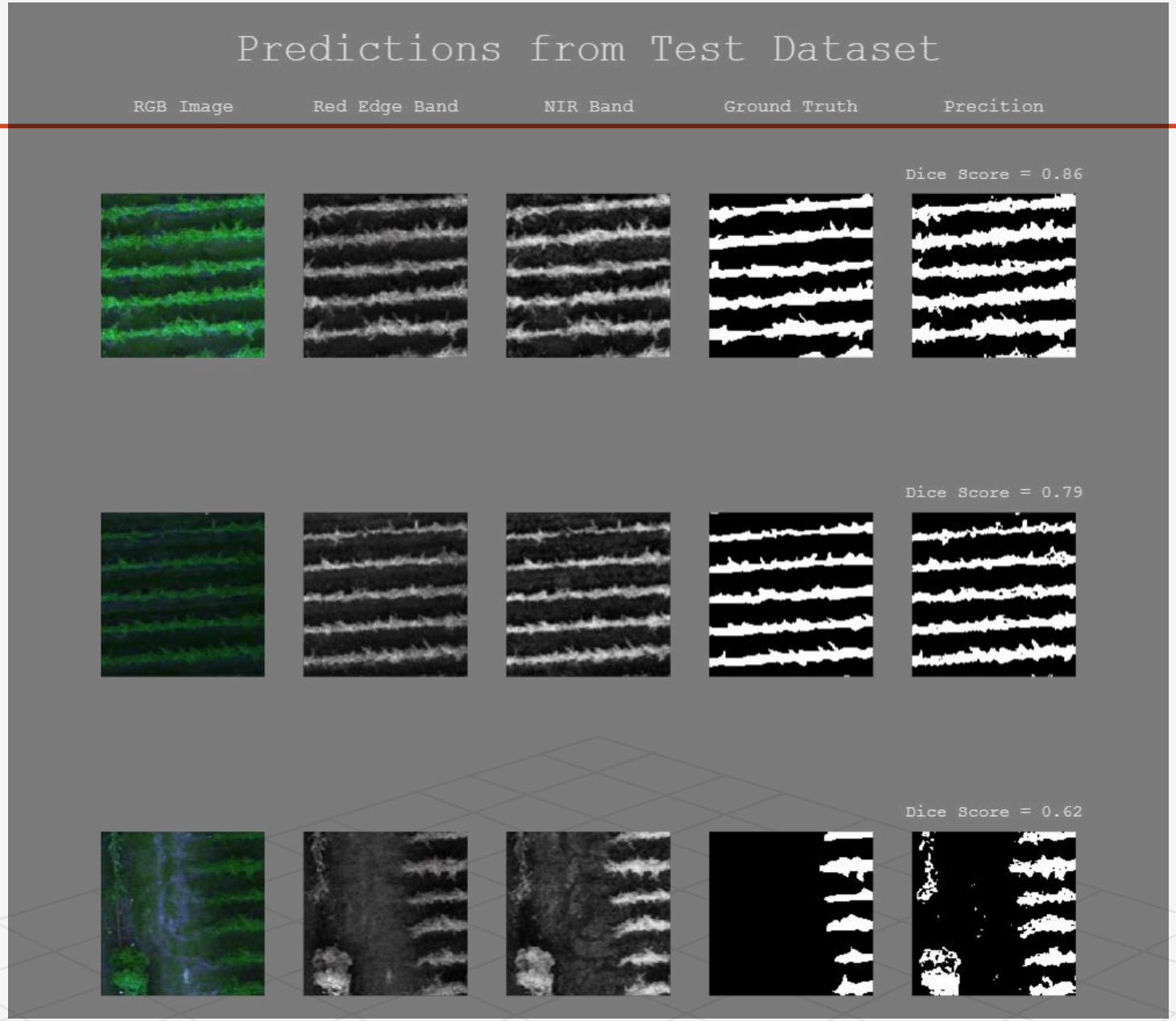


	precision	recall	f1-score
0	0.84	0.88	0.86
1	0.87	0.83	0.85
accuracy			0.85
macro avg	0.85	0.85	0.85
weighted avg	0.85	0.85	0.85

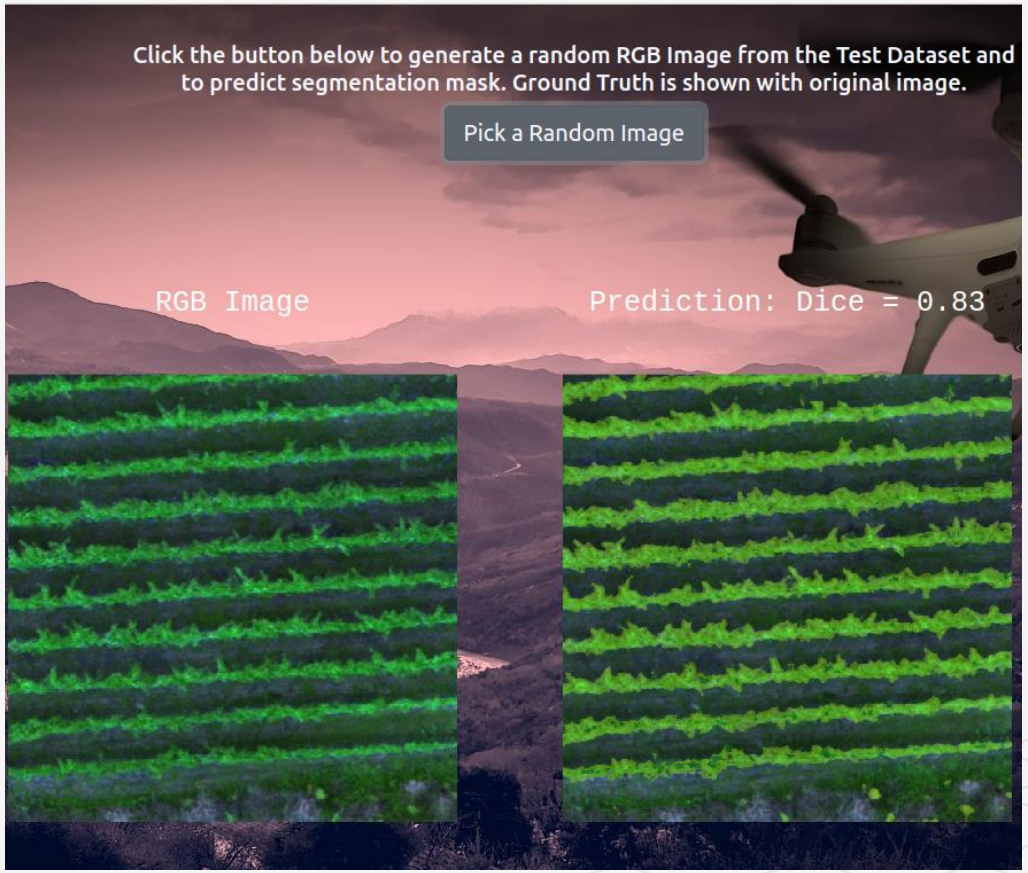

Evaluation of Test Dataset: Precision/Recall Curve
Precission
1
0.9
0.8
0.7
0.6
0.5
0
0.2
0.4
0.6
0.8
1
Recall

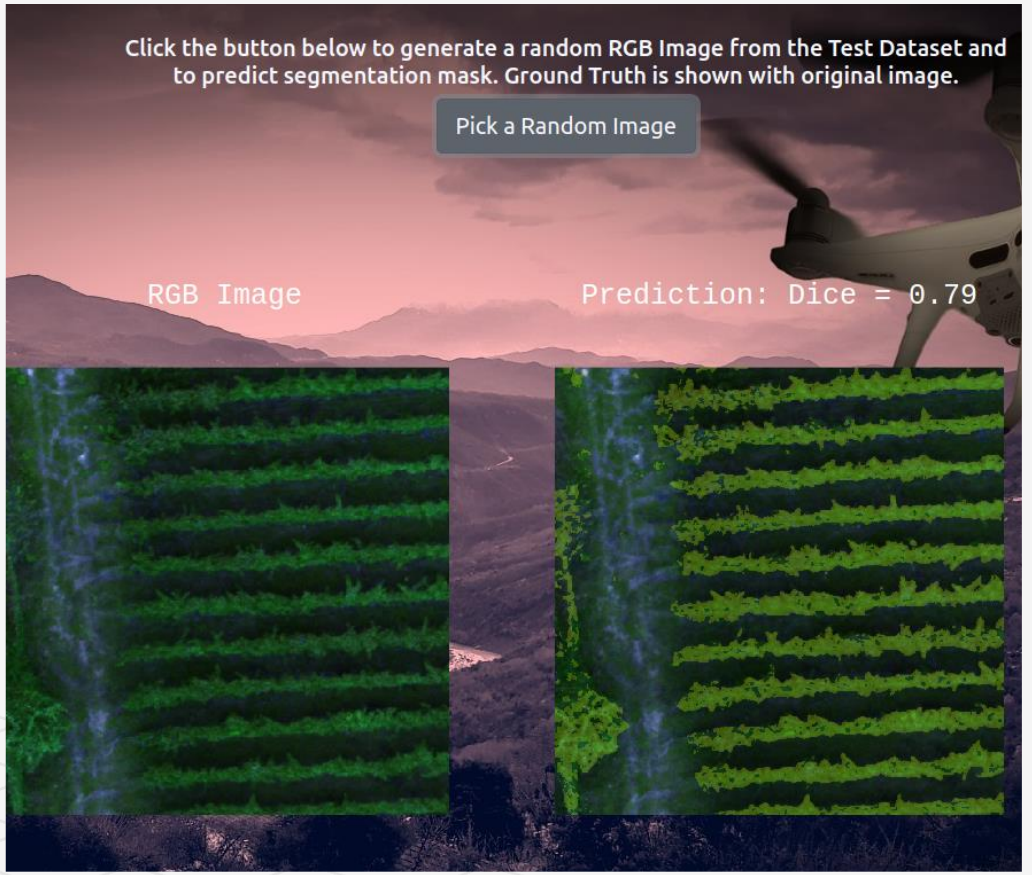

Evaluation of Test Dataset: ROC Curve
True Positive Rate
False Positive Rate
0
0.2
0.4
0.6
0.8
1
ROC - AUC=0.93
baseline

Predictions from Test Dataset

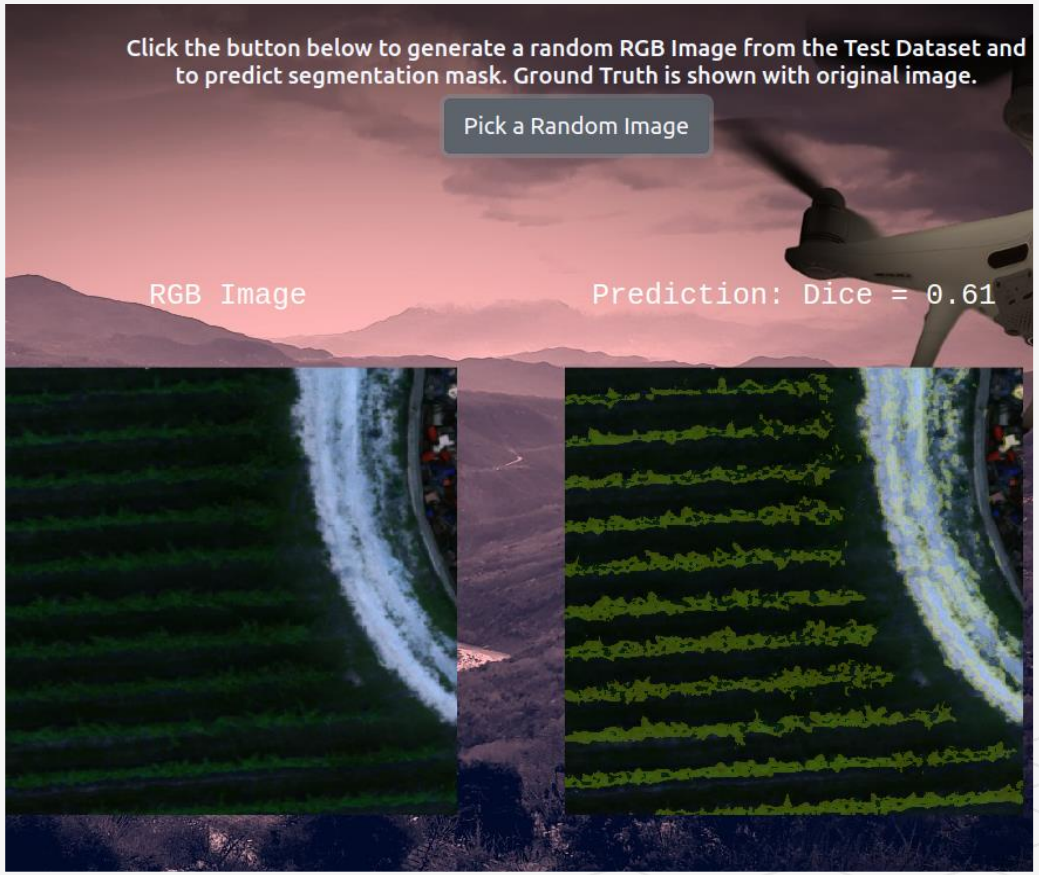


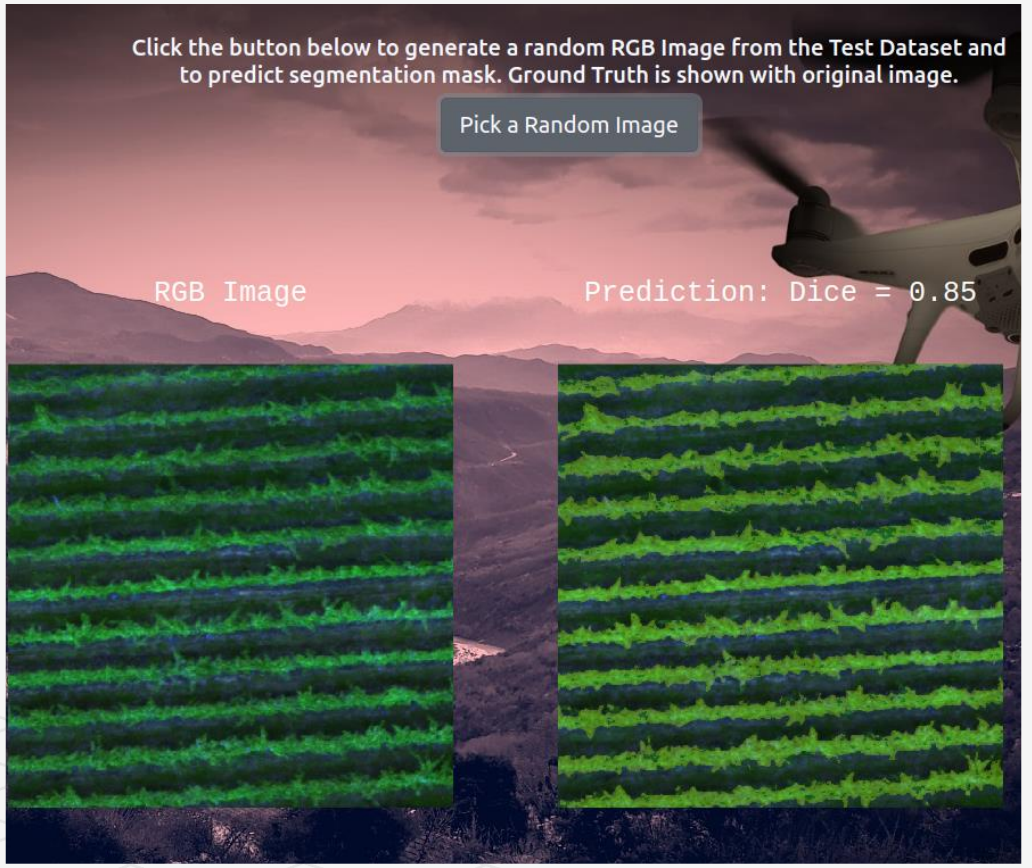
Application





Application





Thank you!