Dataset (GAPSv2) summary:

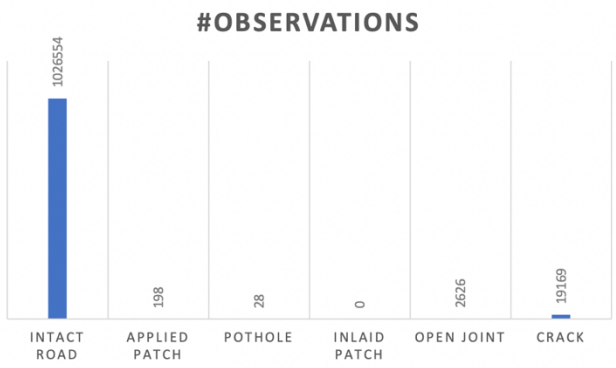
There are a total of 2468 images with shape `(1080, 1920, 3)`

Dataset size = `(2468, 1080, 1920, 3)`

Training-split summary

There are a total of 330 images. The summary of the dataset is provided below:

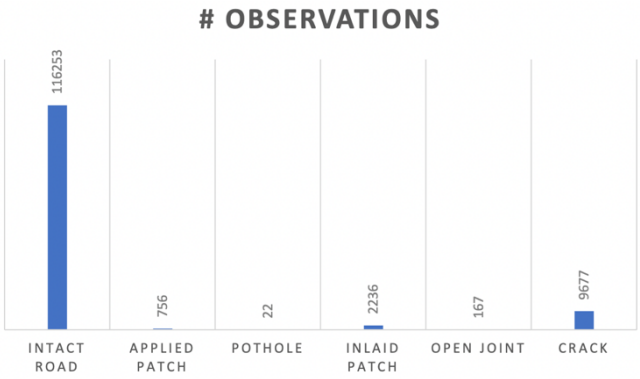
Class	Observations
Intact road	1,026,554
Applied patch	198
Pothole	28
Inlaid patch	0
Open joint	2626
Crack	19169



Test-split summary

There are a total of 500 images. The summary of the dataset is provided below:

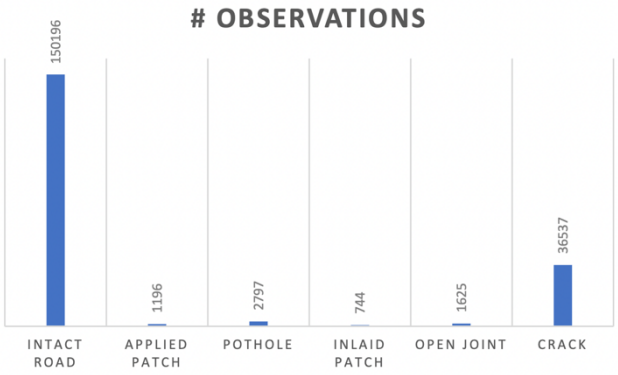
Class	Observations
Intact road	116253
Applied patch	756
Pothole	22
Inlaid patch	2236
Open joint	167
Crack	9677



Validation-split summary

There are a total of 51 images. The summary of the dataset is provided below:

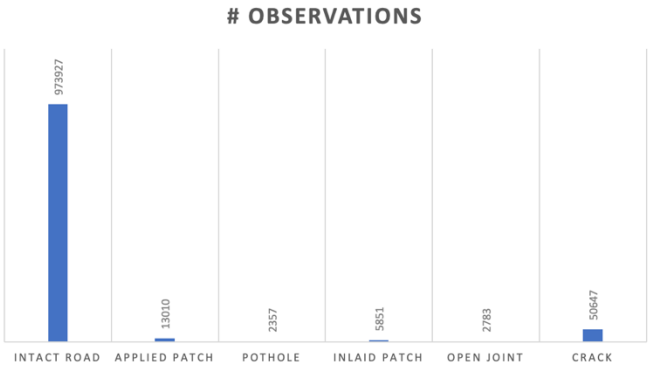
Class	Observations
Intact road	150,196
Applied patch	1,196
Pothole	2,797
Inlaid patch	744
Open joint	1,625
Crack	36,537



Validation-test summary

There are a total of 334 images. The summary of the dataset is provided below:

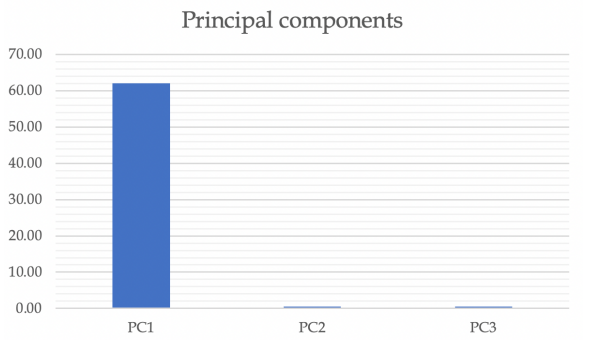
Class	Observations
Intact road	973,927
Applied patch	13,010
Pothole	2,357
Inlaid patch	5,851
Open joint	2,783
Crack	50,647



Data visualization

A part of the training data i.e., about 48,000 images belonging to three major class (i.e., intact pavement, applied patch, crack) is considered to reduce the dimensions of data to two, to visualize the difference between the classes. Some crack images (green) lies in the region of intact images (white points), these are tough to classify and would require special attention while training the classification model.

Principal components	% Variance explained
PC1	62.07
PC2	0.58
PC3	0.49
PC4	0.32
PC5	0.27
PC6	0.25
PC7	0.24
PC8	0.24
PC9	0.23
PC10	0.23



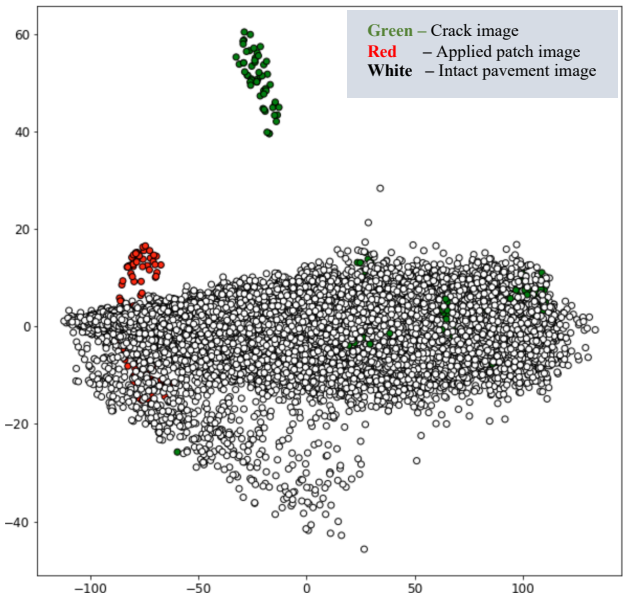


Fig. PC1 – PC2 plot