Dataset Split

The set contains 6000 samples of which 2125 are Defective and 3875 are Non Defective samples.
I randomly removed 1750 Non Defective samples so that the dataset has 2125 samples from each class.

20% of 2125 = 425
Thus the test set now contains 425 samples from each class.
We are left with 1700 samples from each class

10% of 1700 = 170
Thus the validation set now contains 170 samples from each class.
We are left with 1530 samples from each class

Thus the training set now contains 1530 samples from each class.

	Defective	Non Defective
Train	1530	1530
Validation	170	170
Test	425	425