Training dataset: 37750 samples

