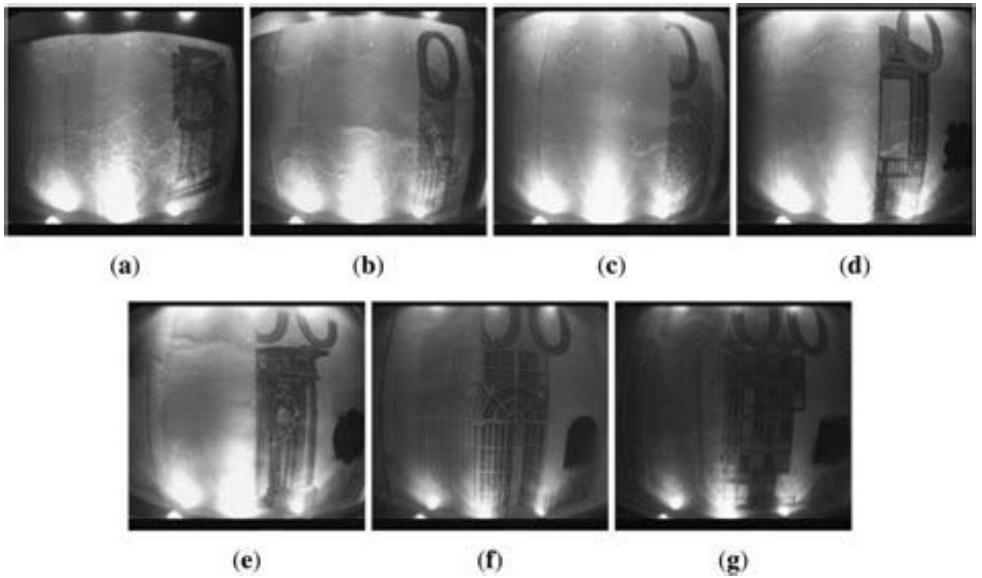
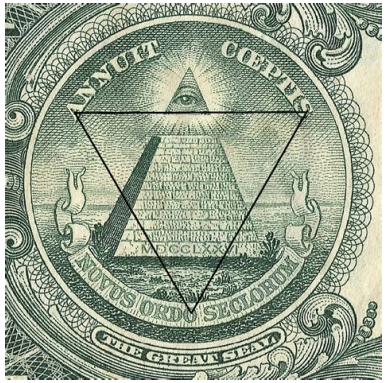
# Banknote authentication specimens

from UCI database

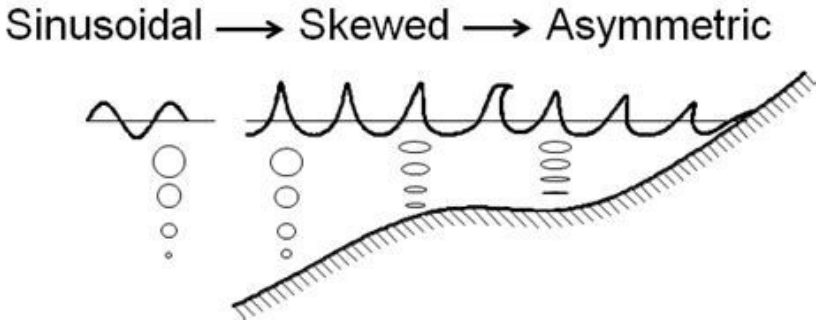



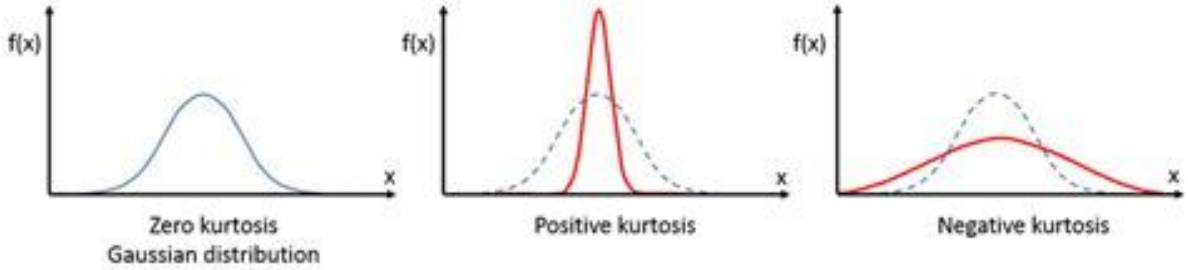
# Data extracted from 1372 examples Of both real and forged notes.

- Industrial camera
- 600 X 600 pixels
- Wavelet transform tool used to acquire features
- 4 base criteria used for inspection:

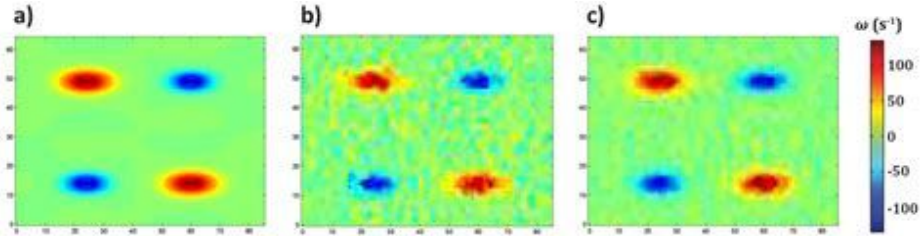


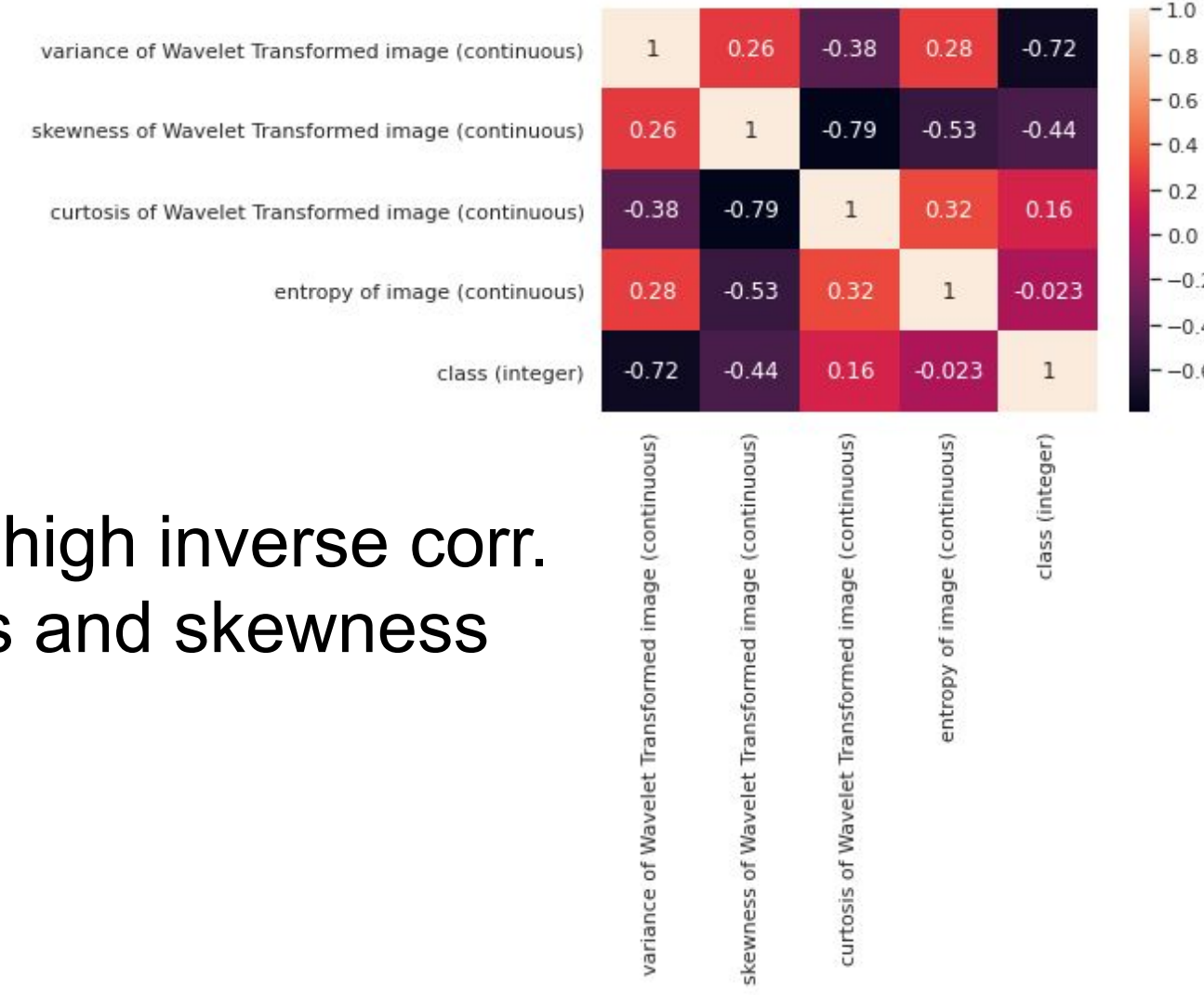
- Variance of wavelet
- Skewness
- Curtosis
- entropy




-Data was all float except for 1 integer
-No cleaning was needed
-No remarkable outliers
-Histogram plot indicated left and right skewness






Heatmap shows high inverse corr. Between curtosis and skewness

# Before and after train test split

| | VOWTI | SOWTI | COWTI | EOI | Class |
|---|---|---|---|---|---|
| **0** | 3.62160 | 8.66610 | -2.8073 | -0.44699 | 0 |
| **1** | 4.54590 | 8.16740 | -2.4586 | -1.46210 | 0 |
| **2** | 3.86600 | -2.63830 | 1.9242 | 0.10645 | 0 |
| **3** | 3.45660 | 9.52280 | -4.0112 | -3.59440 | 0 |
| **4** | 0.32924 | -4.45520 | 4.5718 | -0.98880 | 0 |
| **...** | ... | ... | ... | ... | ... |
| **1367** | 0.40614 | 1.34920 | -1.4501 | -0.55949 | 1 |
| **1368** | -1.38870 | -4.87730 | 6.4774 | 0.34179 | 1 |
| **1369** | -3.75030 | -13.45860 | 17.5932 | -2.77710 | 1 |
| **1370** | -3.56370 | -8.38270 | 12.3930 | -1.28230 | 1 |
| **1371** | -2.54190 | -0.65804 | 2.6842 | 1.19520 | 1 |

1372 rows × 5 columns

| | VOWTI | SOWTI | COWTI | EOI |
|---|---|---|---|---|
| **430** | 1.569100 | 6.34650 | -0.18280 | -2.409900 |
| **588** | -0.278020 | 8.18810 | -3.13380 | -2.527600 |
| **296** | 0.051979 | 7.05210 | -2.05410 | -3.150800 |
| **184** | -1.755900 | 11.94590 | 3.09460 | -4.897800 |
| **244** | 2.428700 | 9.38210 | -3.24770 | -1.454300 |
| **...** | ... | ... | ... | ... |
| **843** | -0.526450 | -0.24832 | -0.45613 | 0.419380 |
| **494** | 2.569800 | -4.40760 | 5.98560 | 0.078002 |
| **1032** | 0.163580 | -3.35840 | 1.37490 | 1.356900 |
| **710** | 2.401200 | 1.62230 | 3.03120 | 0.716790 |
| **333** | 3.000900 | 5.81260 | -2.23060 | -0.665530 |

412 rows × 4 columns

# 4 models with confusion matrix


- Logistic regression--------- `array([[226, 3], [ 2, 181]])`
- KNeighborsClassifier---------- `array([[229, 0], [ 0, 183]])`
- DecisionTreeClassifier---------- `array([[228, 1], [ 7, 176]])`
- RandomForestClassifier---------- `array([[229, 0], [ 2, 181]])`

| | | | | | |
|---|---|---|---|---|---|
| **Accuracy :** | 0.9878640776699029 | **Recall :** | 0.9890710382513661 | **Prec Score:** | 0.9836956521739131 |
| | 1.0 | | 1.0 | | 1.0 |
| | 0.9830097087378641 | | 0.9672131147540983 | | 0.9943820224719101 |
| | 0.9975728155339806 | | 0.994535519125683 | | 1.0 |

Conclusions:

- Clean data set. Would be good to explore with a set with noise.
- Found topic to be interesting will look for more data sets.
- Test for SVM and Naive Bayes
- Verify balance of class