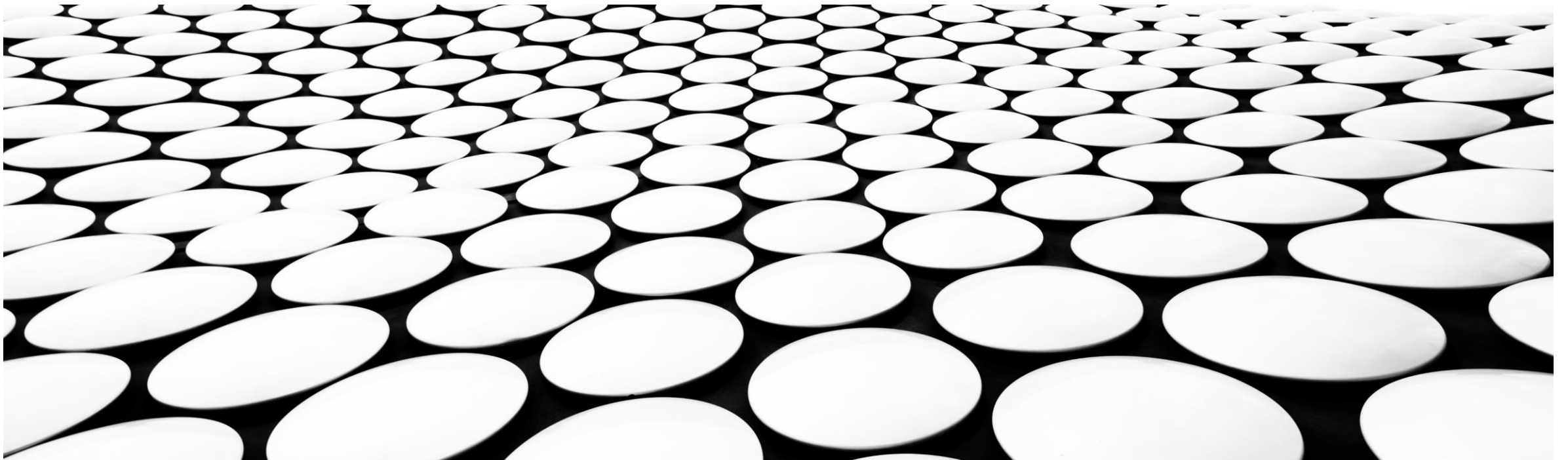
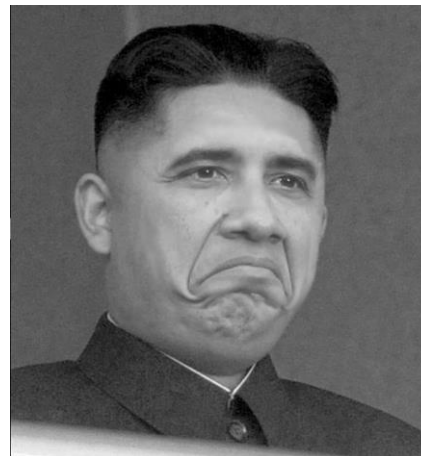
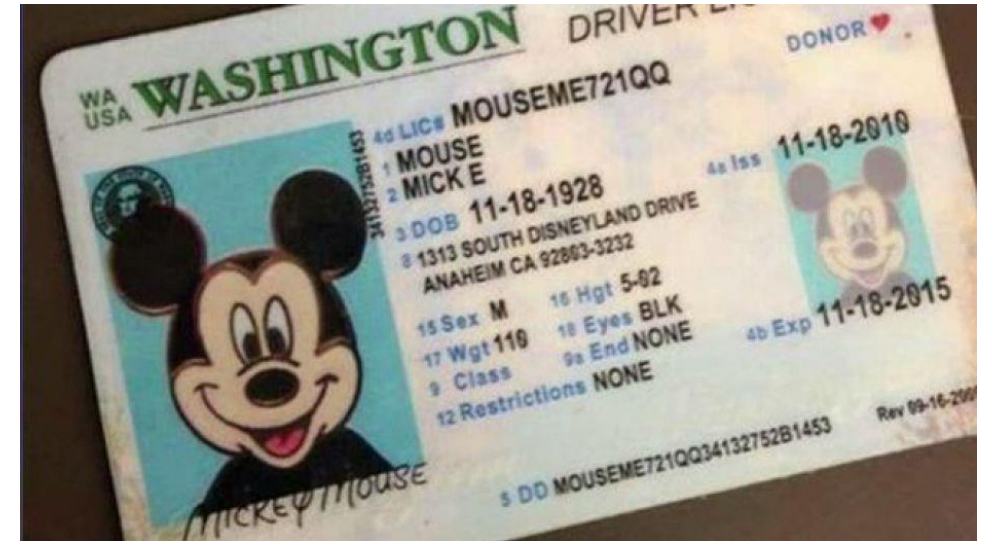

PHOTOSHOP OR NOT?

JOEY CHEW, GA-DSI-13



PROBLEM STATEMENT

- Identify Modified Digital Images
 - Combat Criminal Fraud
 - Academic Falsification
 - Common Mischief

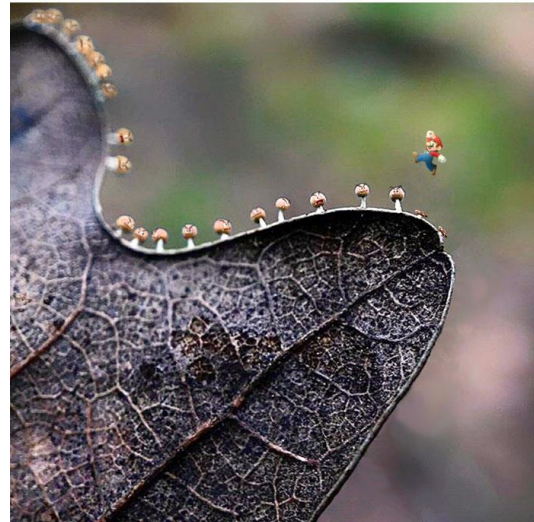


GOAL - MODEL PREDICTING BINARY CLASSIFICATION

- Class 0: Unmodified Images



- Class 1: Modified Images

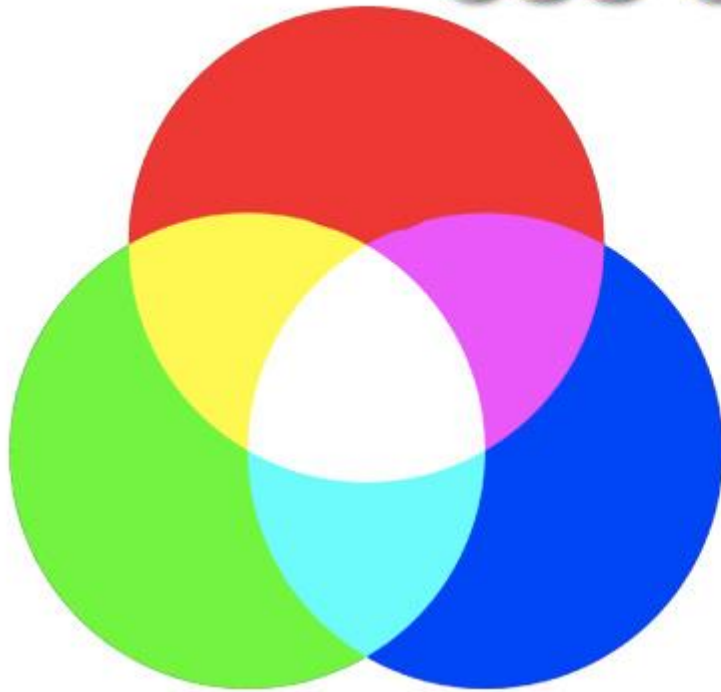


DATA SOURCE

- <https://github.com/dbisUnibas/PS-Battles>,
- Databases and Information Systems Research Group
 - Department of Mathematics and Computer Science, University of Basel, Switzerland.
- 10, 000 Original Images
- 90, 000 Photoshopped Images

COLOR CHANNELS

CSS COLORS



RED : #FF0000 or rgb(255,0,0)

GREEN : #00FF00 or rgb(0,255,0)

BLUE : #0000FF or rgb(0,0,255)

Red Value. Green Value, Blue Value

TOP 4 IMAGE FORMATS FOR THE WEB [2018]

| | JPG | GIF | PNG | SVG |
|--------------|-----|-----|-----|-----|
| VECTOR | | | | ✓ |
| RASTER | ✓ | ✓ | ✓ | |
| TRANSPARENCY | | | ✓ | ✓ |
| ANIMATION | | ✓ | ✓ | ✓ |
| LOSSY | ✓ | | | |

INSUFFICIENT COMPUTING POWER



INSUFFICIENT RAM



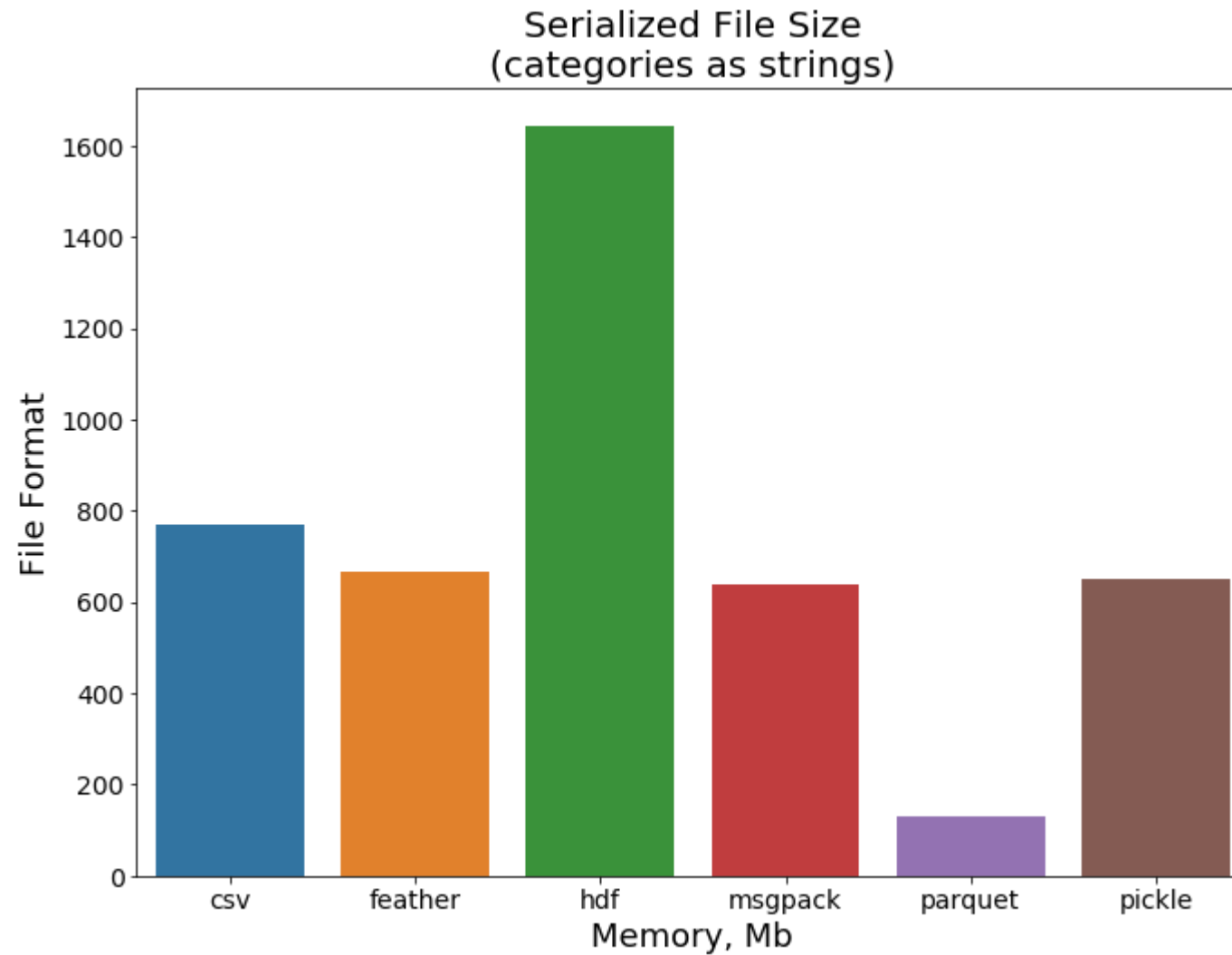
- 32 GB RAM
- Thumbnail (256, 256, 3)
- 100,000 >> 3,000 Images
- Batch Processing



DATA TRANSFORMATION AND ENGINEERING

- Balance Classes
- More Base Reference Images
- Padding to Uniform Size

FILE TRANSFER FORMATS



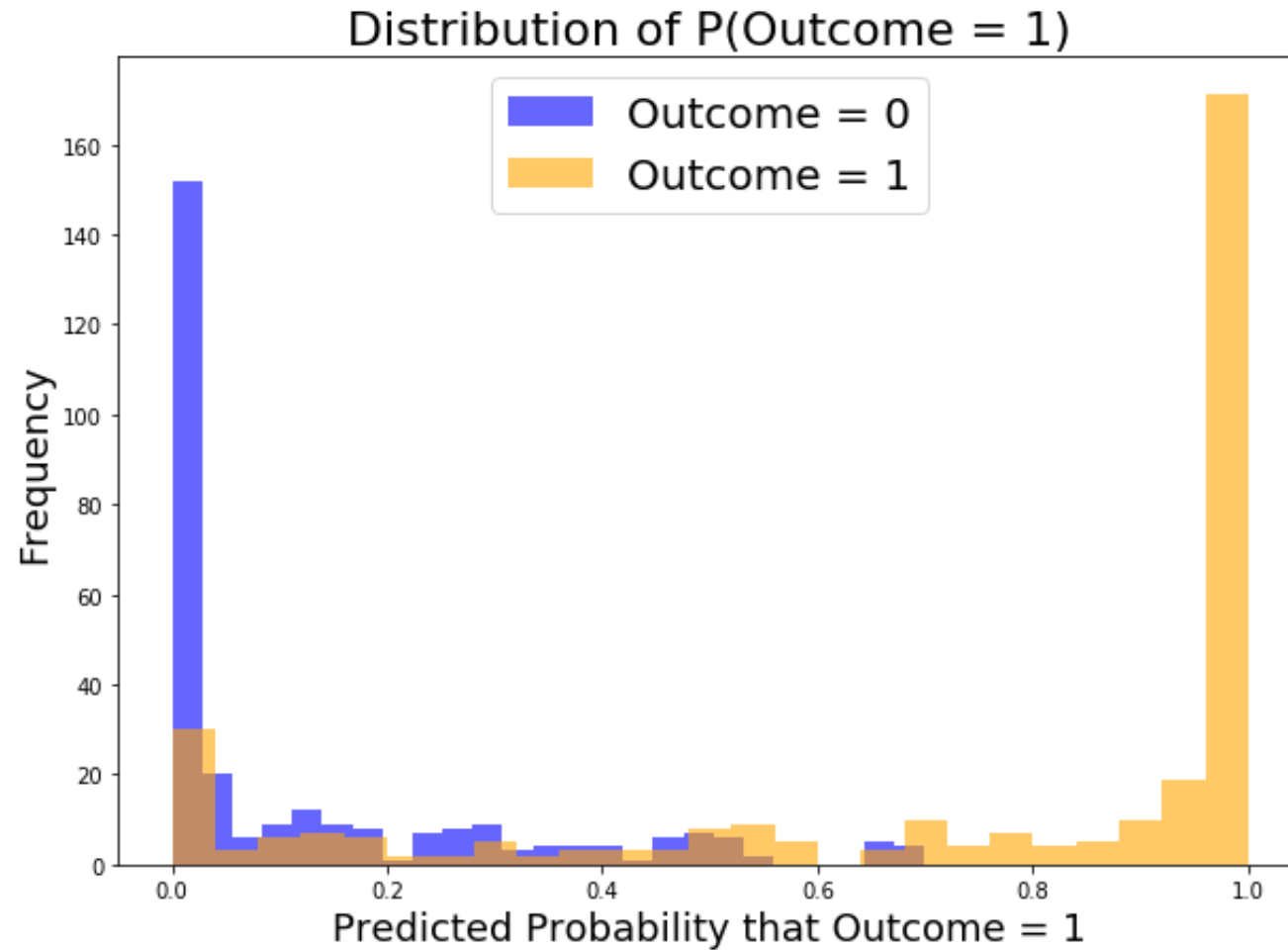


MODEL RESULTS

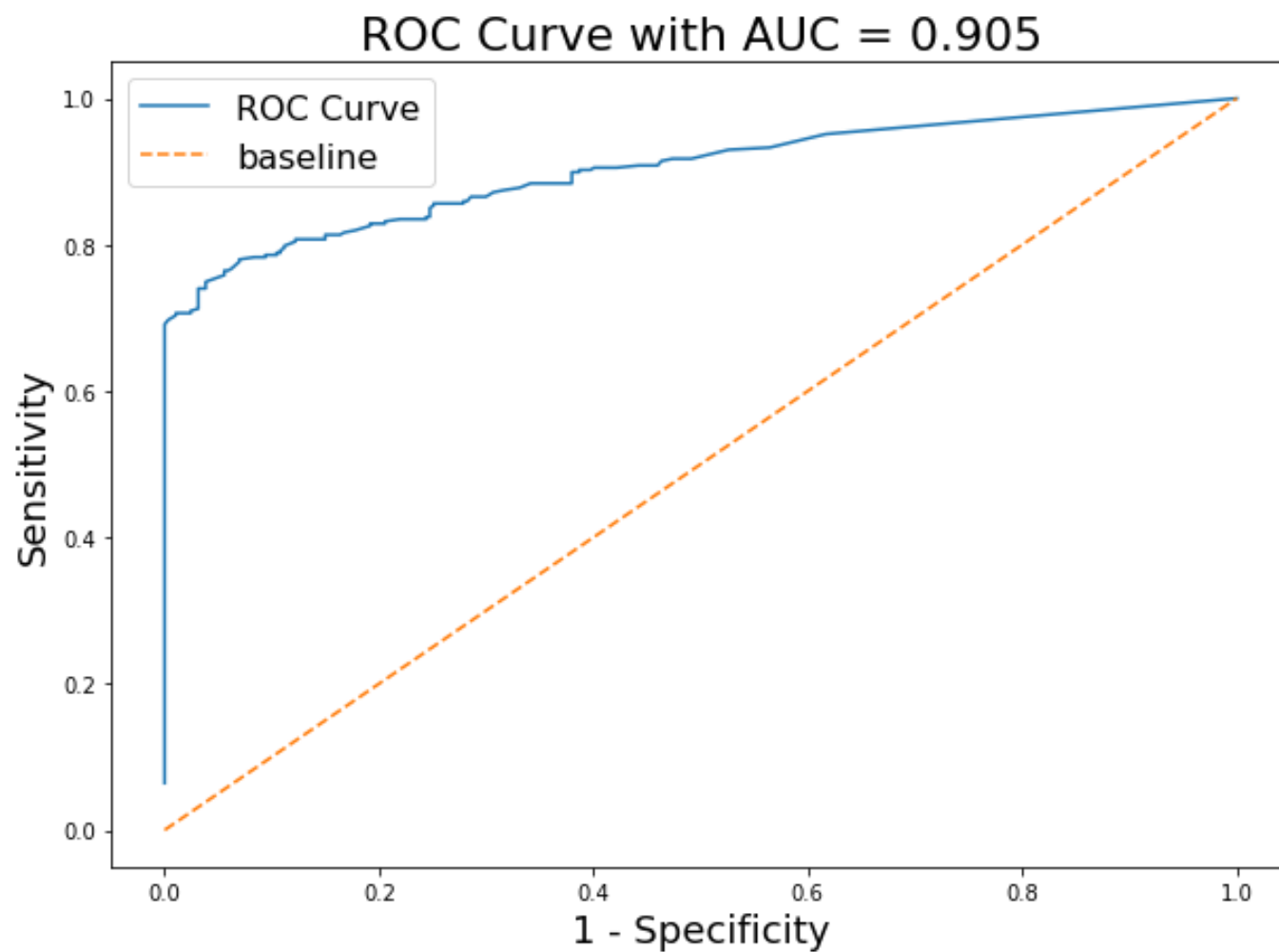
- Baseline Majority Class of 53%

| Model | Train Accuracy | Test Accuracy | Sensitivity | Specificity | Precision | F1-Score | ROC AUC Score |
|-------|----------------|---------------|-------------|-------------|-----------|----------|---------------|
| CNN | 0.95 | 0.85 | 0.93 | 0.78 | 0.79 | 0.85 | 0.91 |

DISTRIBUTION OF PROBABILITIES

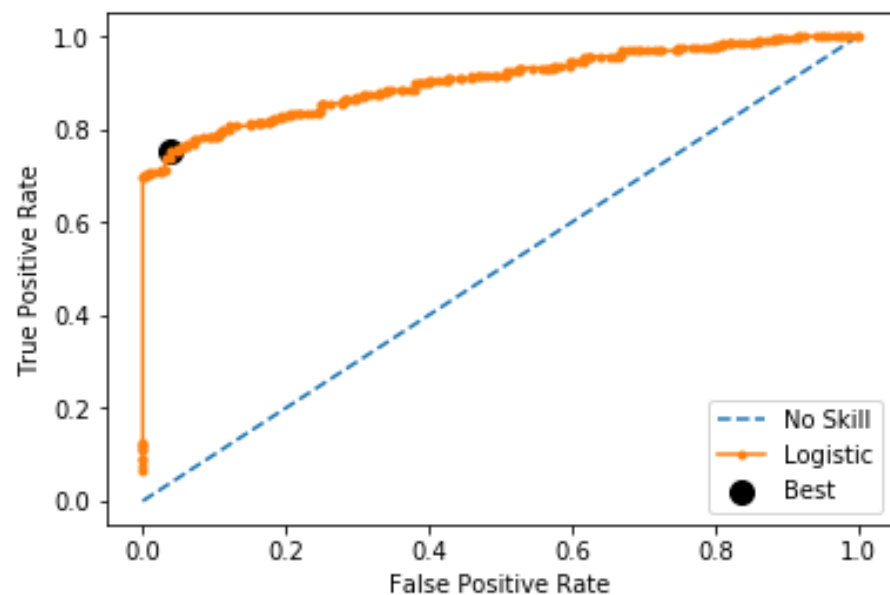


ROC AUC

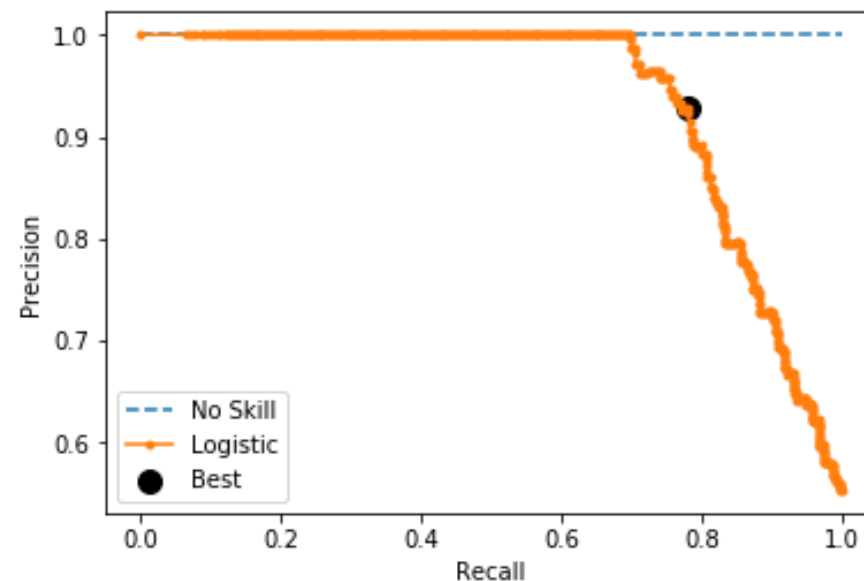


ADDITIONAL TWEAKING

Best Threshold=0.521476, G-Mean=0.852



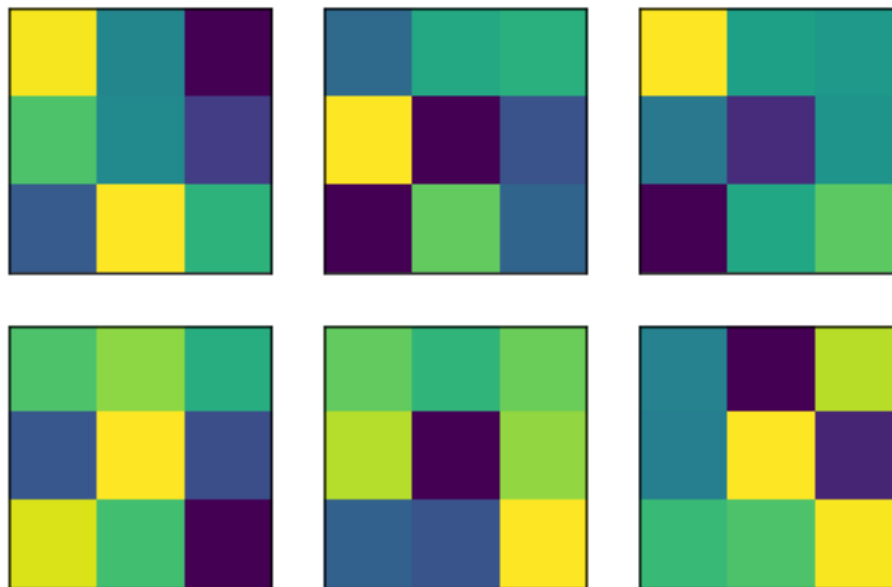
Best Threshold=0.484920, F1-Score=0.847



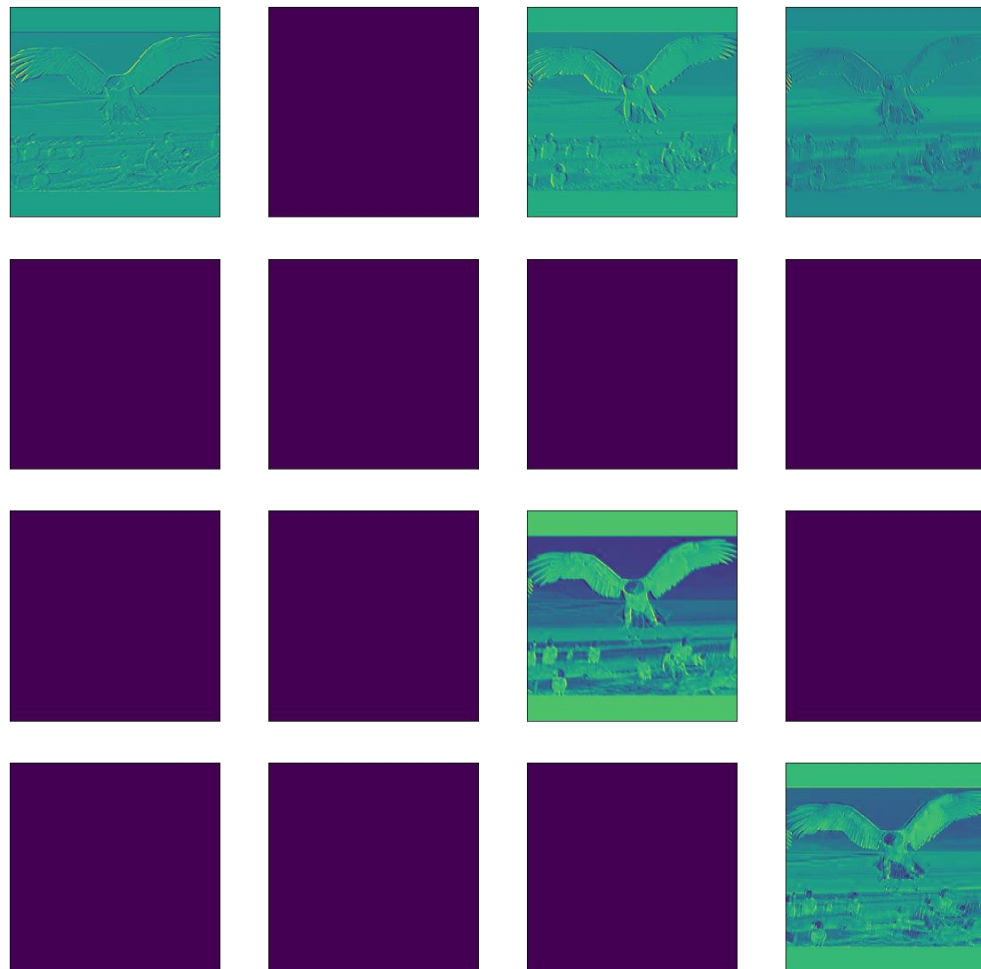
VISUALISATION IN CNN NETWORK – SAMPLE IMAGE



VISUALISATION IN CNN NETWORK – 1ST LAYER FILTERS



VISUALISATION IN CNN NETWORK – FEATURE MAPS



CONCLUSION

- 85% Accuracy – Good but not Great!
- Launchpad for Improvement and Further Development
- Need for More Base Reference Images
- Videos (SARIMAX, KNN)
- Masking, Partial Extraction
- Pre-defined Filters