Fake Banknote Detection

Monday, April 29, 2024
Team 9

Spencer Michalke, Zain Raza, Ryotaro Takehara

Table of Contents

List of Figures and Tables

Executive Summary

In this report, we explore the use of machine learning on visual inspection of real and fake Japanese banknotes. We use two datasets in our analysis: an acquired online dataset from a university repository and a created dataset using our own phone cameras and the same image transformation technique as did the acquired. Extracted features of the wavelet-transformed images included variance, skewness, curtosis, and entropy. Next, we trained and evaluated various machine learning models on the extracted features of the acquired dataset in order to classify between real and fake banknotes. Using the best model, a voting classifier ensemble, we examined its use on our created dataset. As a result, it is possible to use machine learning models to identify between genuine and fake Japanese banknotes. We also found that the model primarily focuses on the variance as the most important feature for differentiation. Further research and training on “dirty” banknotes would be necessary to apply a more realistic situation. Moreover, the use of professionally-made counterfeits may benefit the model in finding real subtle differences rather than the model potentially focusing on specific differences between real and fake banknotes. In our created dataset for example, our fake bake notes did not have a shiny anti-counterfeit sticker, which could have led the image processing tool to exhibit less variance, thus contributing to abnormally high recognition rates with our model.

Introduction

Counterfeiting money can cause a serious financial issue for those involved. It can lower the value of genuine currency and increase distrust in the money system. For businesses, it can lead to sizable financial losses. According to Japan's National Printing Bureau (2024), only a few thousand banknotes are detected annually. When performed manually, checking for genuine banknotes can be a burdensome process, often having to exercise a person's sense of touch and sight. This requires a lot of attention and a considerable amount of time. Thus simply put, it is an inefficient use of resources. If there were to be any machine learning based system that could distinguish between genuine and fake banknotes solely by images, it would reduce the burden of those who are involved in handling banknotes on a daily basis.

These issues have motivated us to use machine learning models to aid the manual process. After using image processing techniques, such as wavelet transformation, we want to train a model to identify whether or not a banknote is real based on those extracted features. Therefore, in addition to the anti-counterfeit measures employed on banknotes, machine learning models can add to the security of genuine banknotes by verifying frequency-domain level details through visual training.

Literature Review

There are several machine learning studies that have shown researchers trying to identify fake banknotes just through image investigation. The most popular methods applied in recent research include deep learning, or more specifically CNN (such as Pham et al., 2020 and Pachón et al., 2021). Some older papers in computer vision, such as Tushar (2017), applied image feature extraction techniques on the spatial domain of banknote images. Another paper leverages hardware to detect the microscopic, counterfeit details of fake banknotes (Baek, 2018).

Surprisingly enough, there is not much research done on the application of frequency domain analysis to detect fake banknotes. Analysis in frequency domains can sometimes capture details that are not apparent in the spatial domain. Some researchers applied such analysis to detect DeepFake, or realistic images generated by DeepLearning (such as Richard et al., 2020). In this report, we explore the possibility of applying frequency domain analysis to fake banknote detection tasks. Application of such analysis might reduce the complexity of detection tasks and would make it possible to detect fake banknotes without applying some complex models, such as CNN.

Data

Our project used two datasets for model training and analysis. The first dataset was used for model training, validation, and evaluation while the second dataset used the best performing model for real world application.

The first dataset was acquired at the University of California Irvine Machine Learning Repository. Created by Volker Lohweg and donated on April 15, 2013, the dataset contains four statistical values that were extracted from wavelet-transformed[1] images of both genuine and fake banknotes. Images of the specimens were taken with an industrial camera. The statistical values include variance, entropy, skewness, and curtosis. Variance, entropy, skewness, and curtosis each describe different features of distribution and it can be said that the characteristic of frequency domain distribution is adequately captured by those 4 metrics. The dataset itself contains 1,372 instances with no missing values, suggesting it has already been prepared and cleaned for immediate use. The table below shows an example of the data. Class asserts whether the banknote was real or fake.

Table 1: An example of the data received is shown below.

Variance	**Skewness**	**Curtosis**	**Entropy**	**Class**
3.6216	8.6661	-2.8073	-0.44699	0
4.5459	8.1674	-2.4586	-1.4621	0
3.866	-2.6383	1.9242	0.10645	0
3.4566	9.5228	-4.0112	-3.5944	0

The second dataset was constructed personally. First, 1200 pictures of a banknote, 600 real and 600 fake, were taken at various angles in different lighting conditions. To add to the differences, different phone cameras were used to take the pictures. All banknote pictures had a white background to remain consistent, but only the fake banknotes were newly printed. While some of the photos may not showcase the entire banknote, none of the photos were cropped to specific dimensions for transformation. Examples of these photos are shown in the figure below. Then, the wavelet transformation tool was used to extract the same features as the first dataset.

[1] Wavelet transform is a technique that transforms spatial domain information (such as signal or image) into frequency domain using a combination of simple wavelets of different scales.

Figure 1: Two real and fake banknotes are shown on the left and right respectively.

Hypothesis and Exploratory Analysis

Before diving into model training, we first performed some initial data analysis techniques on the acquired dataset to find any underlying patterns. A box plot was used to show the distribution of data for the acquired dataset. While there isn't enough difference between the real and fake banknote features, entropy and curtosis seem to have a large amount of outliers.

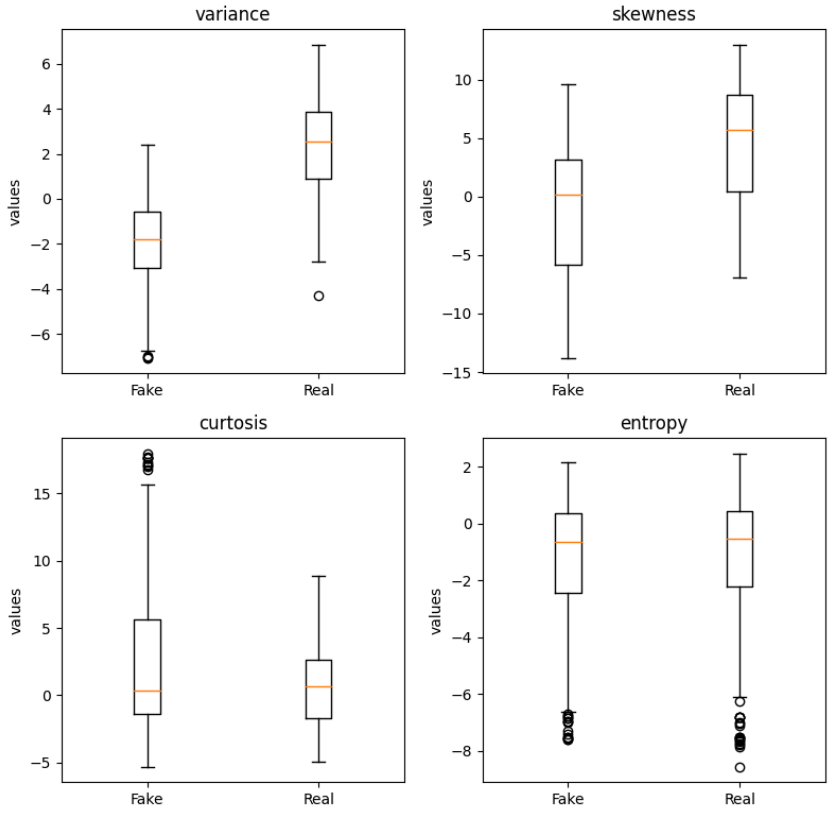

Figure 2: Boxplot of the acquired dataset features.

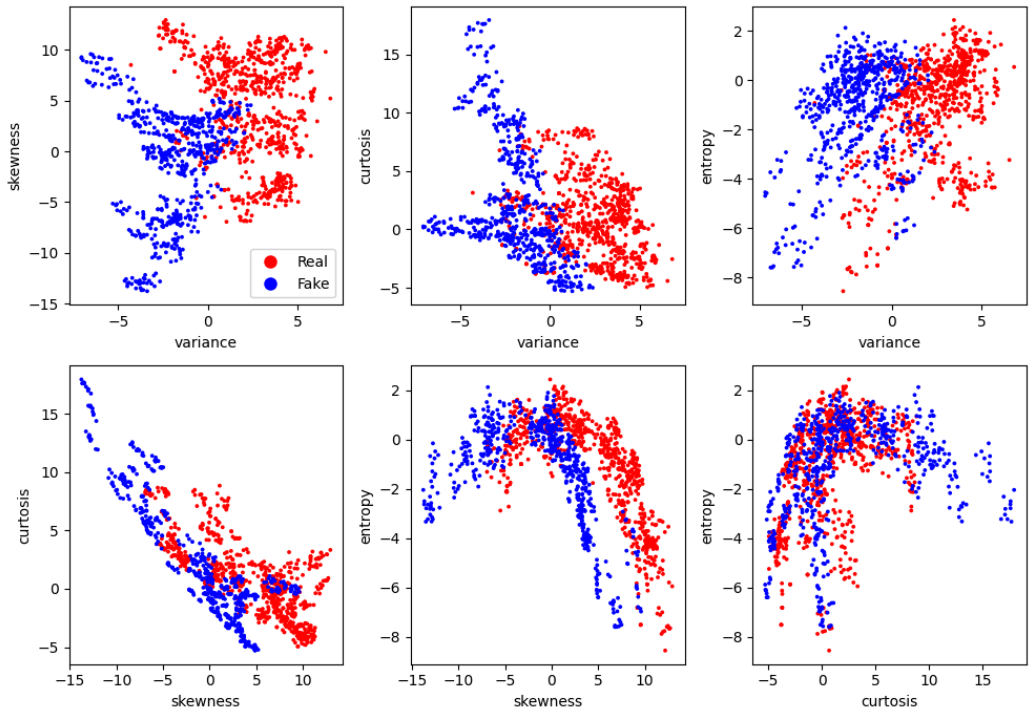

Figure 3: Variable combination scatter plots from online dataset

Next, we graphed a scatter plot of different combinations of the features, as shown above in **Figure 3**. It's interesting to note that the acquired dataset had negative entropy and variance. This doesn't usually occur. However, when we applied the scatter plot to our created dataset as shown below, we found the same general topological patterns. We believe that the negative variance & entropy found in the acquired dataset might be due to some sort of standardization technique not specified by the original creator. After observing the patterns, we decided to apply models that can draw non-linear decision boundaries since a linear one does not seem blatantly apparent. Therefore, our hypothesis is that real and fake banknotes can be separable from those features using machine learning models that can draw non-linear decision boundaries.

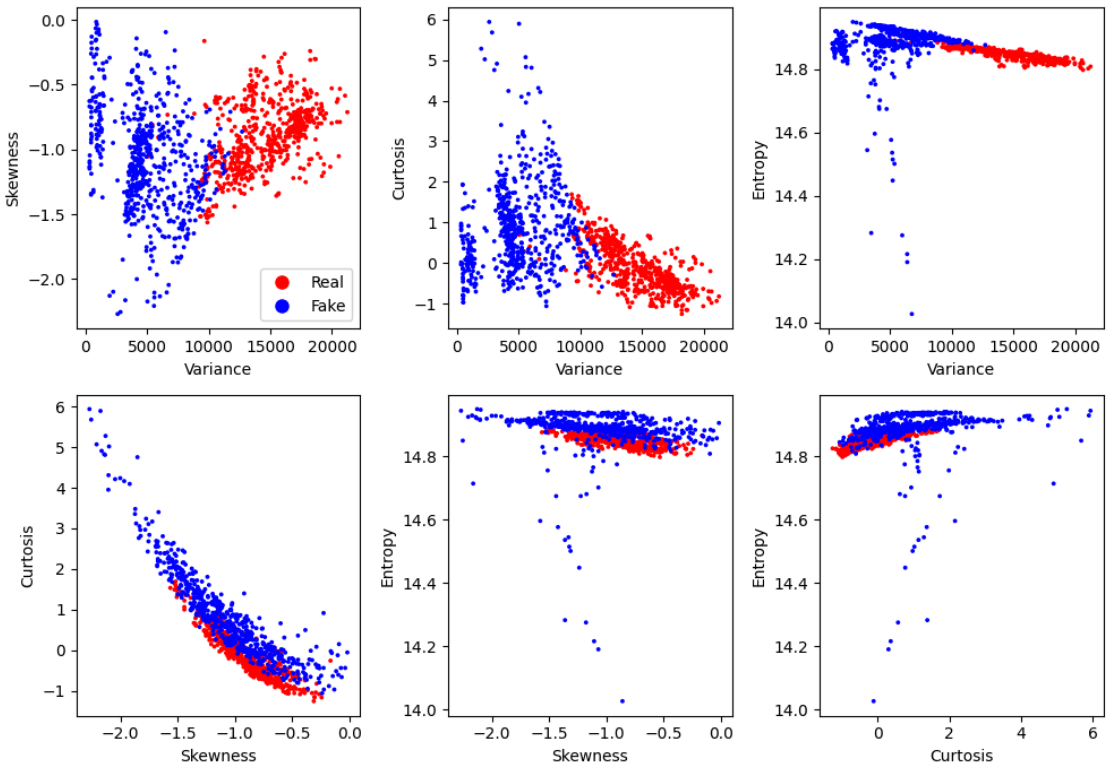

Figure 4: Variable combination scatterplots from our dataset.

Model Training and Validation

After the initial exploratory analysis, our next task was to train various models on the acquired data in order to have a calibrated model for use on our created dataset. The acquired dataset was split for training, validation, and testing 55%, 20%, and 25% respectively. We trained three separate models on the acquired data for comparison. The first model was a soft voting classifier ensemble (VCE), consisting of Support Vector Classification (SVC), k-nearest neighbors (KNN), and Decision Tree classifiers. **Table 2**, shown below, displays the various parameters chosen for tuning the models. The other two models included Random Forest Regressor (RFR) and XGBoost Regressor (XGBoost). Various maximum depths, between 2 and 20, was the only parameter tested for RFR and XGBoost.

Table 2: Various Parameter for Tuning the VCE Model

SVC	'svc__C': [0.1, 1, 10] 'svc__gamma': [0.1, 1, 10]
KNN	'knn__n_neighbors': [3, 5, 7] 'knn__weights': ['uniform', 'distance']
Decision Tree	'dt__max_depth': [2, 5, 10] 'dt__min_samples_split': [2, 5, 10]

To find the best hyperparameters for the VCE model, GridSearchCV was fitted to the training data. GridSearchCV tests every possible combination of the parameter set we gave it and measures the performance of the specific combination. Recall score, which measures the ratio of true positives to true positives and false negatives, was used as the performance metric. It is important that we use the recall score as our performance metric because it penalizes the false negative classifications. Other parameters specified within GridSearchCV included choosing a cross validation of 5 and verbose value of 1. We then performed scoring on the validation dataset using the best parameters. **Table 3** shows the best parameters chosen.

Table 3: Best Parameters for VCE Model

SVC	'svc__C': 0.1 'svc__gamma': 0.1
KNN	'knn__n_neighbors': 3 'knn__weights': 'uniform'
Decision Tree	'dt__max_depth': 2 'dt__min_samples_split': 2

Tuning the hyperparameters for the RFR and XGBoost models required multiple iterations, each in a different random state, so that we can get a confidence interval of the performance at each maximum depth. At each iteration, the models were fitted to the training dataset, predicted using the validation dataset, and then renormalized to have a probability between 0 and 1. We can then classify the prediction using a decision threshold, which was chosen as 0.5 because a calibrated model will obtain a decision

threshold near this value. Again, we use the recall score metric to measure the performance between the validated dataset and our predicted values. This completes one iteration for each model, which was repeated 10 times at each maximum depth. **Figure 5** shows a graphical representation of the recall scores at different maximum depths. We chose a maximum depth of 2 for XGBoost and a maximum depth of 6 for RFR. We ultimately chose a maximum depth of 6 for RFR because the difference in subsequent recall scores is marginal and we wanted to prevent overfitting.

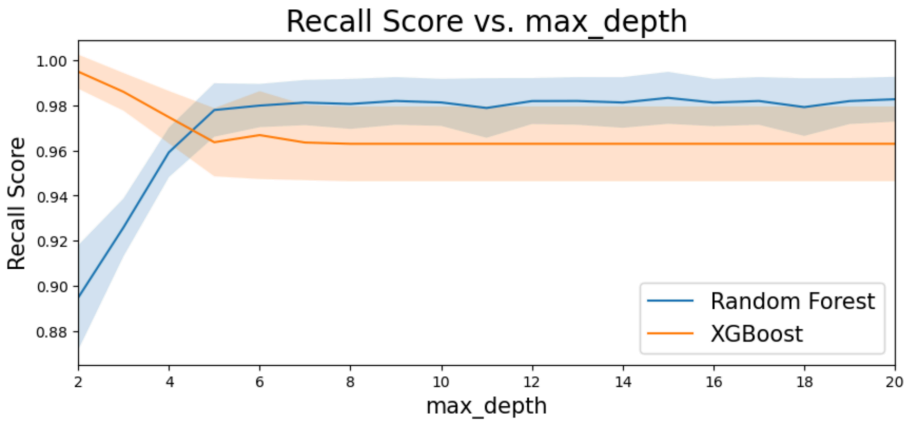

Figure 5: Recall Score vs Maximum Depth of each Regressor

Finally, we evaluated the three models by using their best parameters and averaged recall scores. For each iteration, we performed bootstrapping on the test portion of the dataset, fitted the best hyperparameters, predicted the test scores, renormalized the test scores, and classified the scores using a decision threshold of 0.5. We performed 25 iterations, each being in a random state, so that we could gain a confidence interval of each model's performance.

Table 4: Statistical metrics after 25 iterations

VCE: min: 1.0 ; max: 1.0 ; mean: 1.0 ; std: 0.0 ; 95%: 1.0
RFR: min: 0.91390; max: 0.98717; mean: 0.95685; std: 0.01796; 95%: 0.98146
XGB: min: 0.96407; max: 1.0 ; mean: 0.99691; std: 0.00783; 95%: 1.0

Every time 25 iterations were performed, the metrics for RFR and XGBoost would change. However, the VCE model remained consistent every time, resulting in 1.0's across every metric. As a result, we decided to use the VCE model for detecting fake banknotes on our created dataset.

Results

We split our real dataset into two different parts: the first was a 50/50 real to fake banknote split. The second part was a 90/10 real to fake banknote split to simulate a more realistic situation[2]. For each type of split, we checked the recall score for 25 different bootstrapped samples from our created data to see how well the model performed. Below shows the recall results for each split.

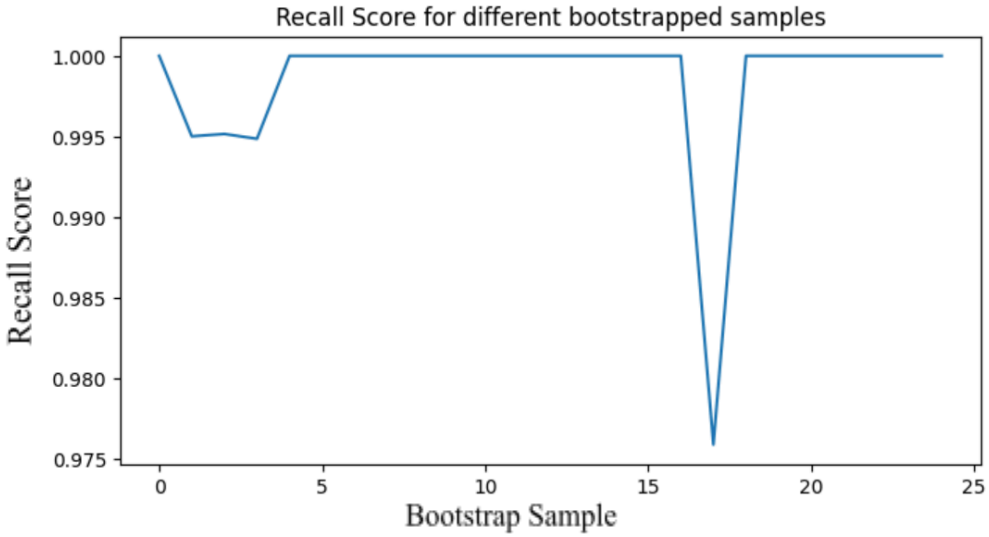

Figure 6: Recall results for 50/50 split.

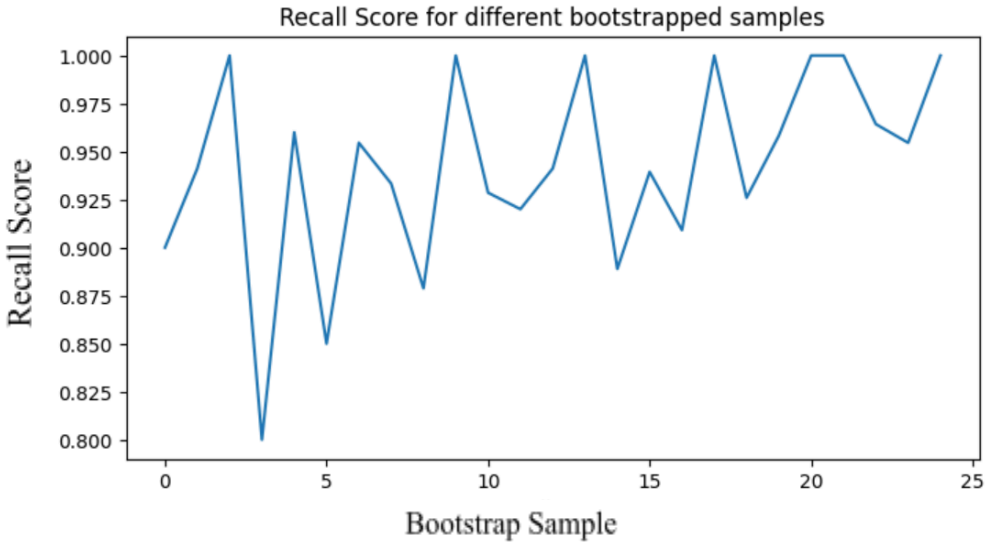

Figure 7: Recall results for 90/10 split.

[2] Usually, people would not confront fake banknotes in the odds of 1. 10% still does not necessarily reflect real life situations, but lowering the ratio further would seriously impact model performance.

The recall score for the 50/50 split yielded nearly perfect 1.0 scores, with the exception of a dip that only went as low as 0.975. For the 90/10 split, there is a lot more variance between the recall scores, dropping to 0.800 in one sample.

The 90/10 model was then calibrated to make the prediction more reliable. The calibration curve below shows that the model is now closer to a perfectly calibrated model after the treatment.

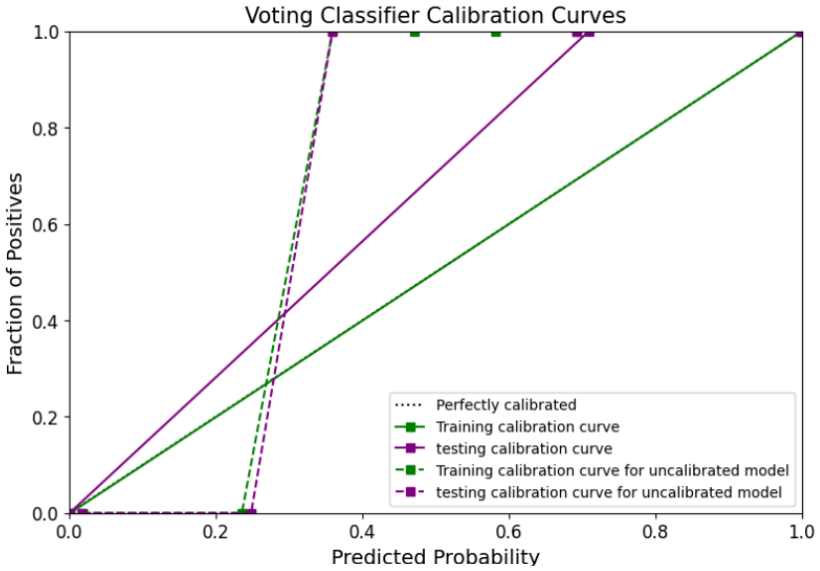

Figure 8: Calibration Curve for the 90/10 split.

Though it was not necessarily expected, better results were obtained after recalibrating the 90/10 model. This might be because the model's threshold was set to 0.5 throughout the experiment and it has been discussed that, for a calibrated classifier, an optimal threshold is indeed 0.5 (Cohen & Goldszmidt, 2004).

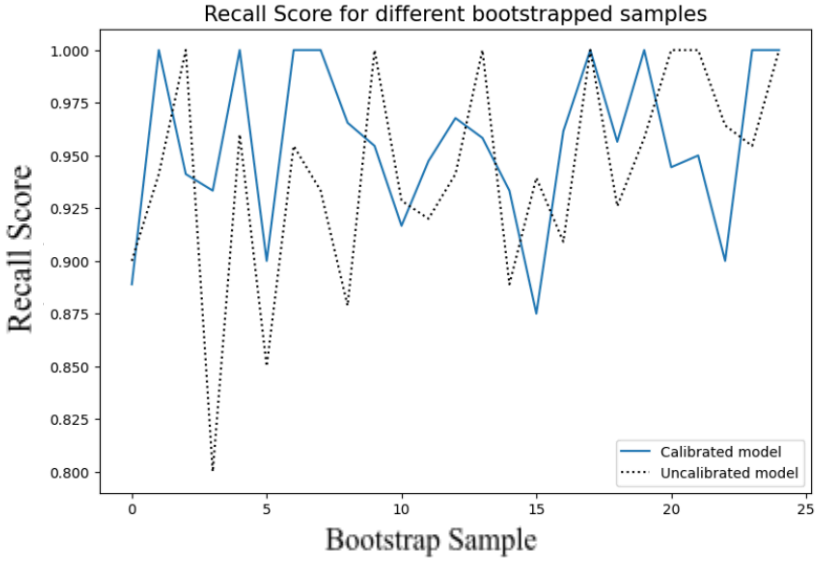

Figure 9: Calibrated recall results for 90/10 split.

Discussion

First, it is interesting that the best model in our research was the voting classifier ensemble. We were expecting the XGBoost Regressor to perform the best due to its adaptability. The calibrated VCE model performed well for our research. It yielded recall scores between 0.90 and 1.00. Although this is high, for real world applications, these scores would need to be averaged higher.

After testing the model, we wanted to know which features the model was using the most for prediction. A partial dependence plot, or PDP, can be used to show the effect a feature has on the outcome of a prediction model. **Figure 10** shows the PDP. Interestingly, the variance is the most important feature when it comes to prediction. Skewness, curtosis, and entropy were nearly consistent throughout. However, this is actually consistent with the visual information achieved through the exploratory analysis. We could really only separate based on the variance.

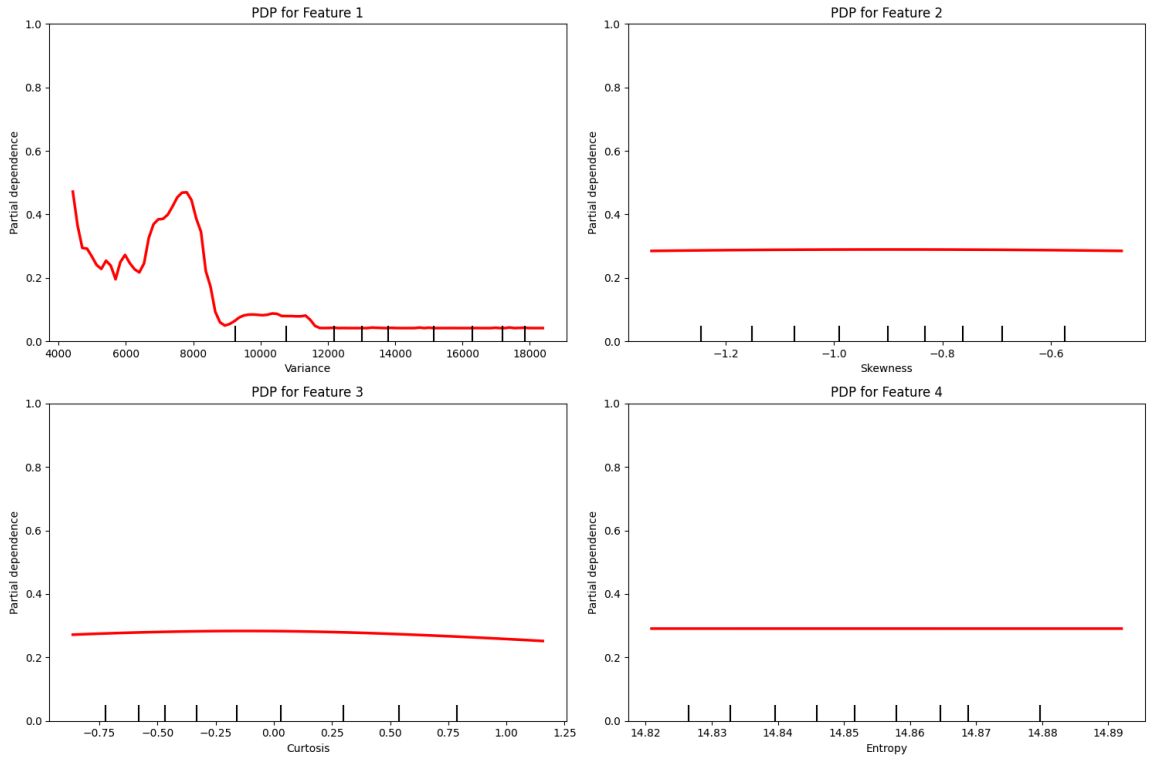

Figure 10: Feature contribution to model performance.

Limitations

The research presented does exhibit some limitations. First, when creating our dataset from the pictures we took, the banknotes were perfectly clean. They did not show any signs of use, like dirt, grime, or small cuts that could happen from being handled in the environment. Secondly, our model could have focused on specific aspects of the banknote. For instance, all genuine banknotes have an anti-counterfeit sticker that is shiny and reflects light easily. The fake banknotes used did not contain such a measure, which can

be seen below in **Figure 11** This difference might be the leading cause of the difference in variance. It is suspected because, during the exploratory analysis, cropping our photos to 400 pixels by 400 pixels, like **Figure 12**, made frequency domain information almost indistinguishable between real and fake banknotes (**Figure 13**). While the true reason behind this phenomenon is unknown, it is possible that making the shiny sticker out of frame might have affected the variance of the image. This phenomenon certainly requires further investigations.

Figure 11: Left image shows an anti-counterfeit sticker on a real banknote while the right image does not.

Figure 12: Left and right image is a closeup of a genuine and fake banknote respectively.

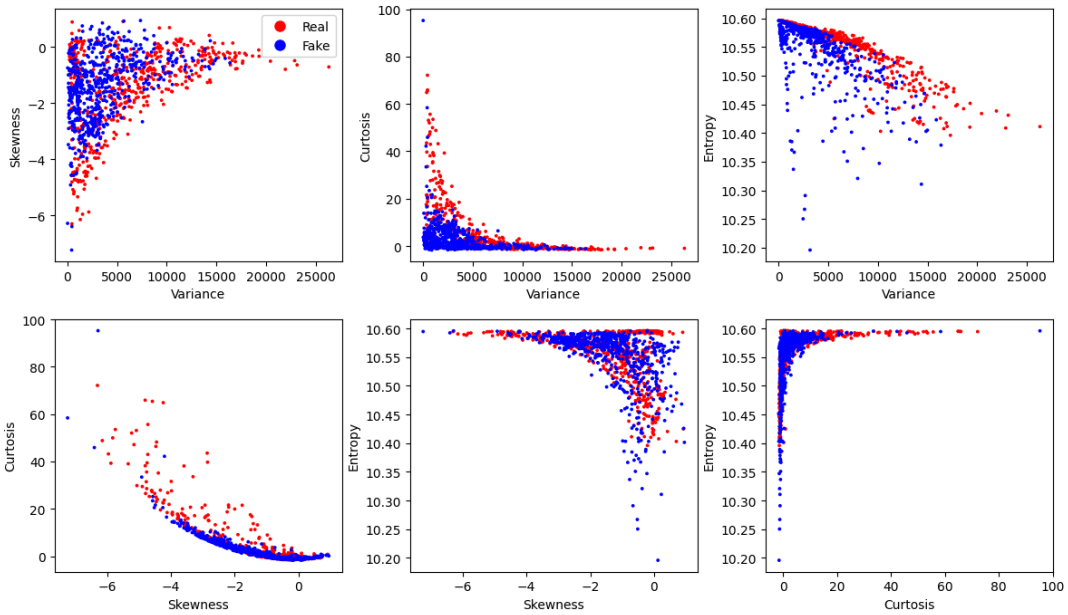

Figure 13: Variable combination scatterplots for cropped images.

Conclusion

This research project aimed to evaluate machine learning models for aiding banknote authenticity. Using the wavelet transformation tool and training various models, we achieved our goal of building a banknote classifier through low dimensional statistical values from an image. After evaluating a voting ensemble model, Random Forest and XGBoost Regressor, we found the best classifier was the voting classifier ensemble model, which took the vote of SVC, Decision Tree, and KNN. This came as a surprise to the researchers because XGBoost usually performs better in non-linear situations. Lastly, we found that the most informative feature in our model was variance.

Future Work

This research report can take many directions in order to improve its outcome and use. A potential direction could utilize more advanced image processing techniques in order to extract more precise features. Another direction could try and include banknotes that have been handled, or dirty, in the real world to make for more realistic situations. Lastly, similar exploratory analysis could be done again with fake banknotes using better quality counterfeits, such that they include shiny seals, as to see the actual effectiveness in a real world situation.

Acknowledgements

The authors would like to recognize Dr. Arya Farahi for their influence, guidance, and support during this project and semester.

Team Contribution

Ryotaro: 100%	**Spencer: 100%**	**Zain: 100%**
● **Data collection** ● **Data preprocessing / visualization** ● **Test with real data** ● **Model interpretation** ● **Project management**	● **Data collection** ● **Model evaluation** ● **Model selection** ● **Report writing / management**	● **Data collection** ● **Hyper-parameter tuning** ● **ROC calculation** ● **Model calibration**

Bibliography

Agasti, T., Burand, G., Wade, P., & Chitra, P. (2017, November). Fake currency detection using image processing. In *IOP Conference Series: Materials Science and Engineering* (Vol. 263, No. 5, p. 052047). IOP Publishing.

Baek, S., Choi, E., Baek, Y., & Lee, C. (2018). Detection of counterfeit banknotes using multispectral images. *Digital Signal Processing*, *78*, 294-304.

Bureau, N. P. (2024, February 14). *Characteristics of banknotes*. National Printing Bureau. https://www.npb.go.jp/en/products/intro/tokutyou.html

Cohen, I., Goldszmidt, M. (2004). Properties and Benefits of Calibrated Classifiers. In: Boulicaut, JF., Esposito, F., Giannotti, F., Pedreschi, D. (eds) Knowledge Discovery in Databases: PKDD 2004. PKDD 2004. Lecture Notes in Computer Science(), vol 3202. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-30116-5_14

Durall, R., Keuper, M., & Keuper, J. (2020). Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7890-7899).

Pachón, C. G., Ballesteros, D. M., & Renza, D. (2021). Fake banknote recognition using deep learning. *Applied Sciences*, *11*(3), 1281.

Pham, T. D., Park, C., Nguyen, D. T., Batchuluun, G., & Park, K. R. (2020). Deep learning-based fake-banknote detection for the visually impaired people using visible-light images captured by smartphone cameras. *IEEE Access*, *8*, 63144-63161.