Development of an Automated Coin Grading System: Integrating Image Preprocessing, Feature Extraction, and ML Modeling

Jianzhu Chen

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

The Bradley Department of Electrical and Computer Engineering

Creed F. Jones, Chair

Luke F. Lester

A. Lynn Abbott

Nov. 26th, 2024

Blacksburg, Virginia

Keywords: Automated Coin Grading, Machine Learning, Computer Vision, and multi-layer perceptron.

Development of an Automated Coin Grading System: Integrating Image Preprocessing, Feature Extraction, and ML Modeling

Jianzhu Chen

ABSTRACT

For more than 70 years, the Sheldon Coin Grading Scale has been essential in quantifying the value of coins within the coin collecting industry. Traditionally, coin grading has relied on human graders who may deliver inconsistent results. This inconsistency leads to variations in coin values. In this thesis, we present an automated coin grading system that uses image preprocessing, feature extraction, and advanced machine learning techniques to predict the grade across different coin types. Our system employs synthetic reference masks to identify "expected" regions, like the contours of reliefs, and "unexpected" regions, such as surface non-uniformities. All detected significant elements and tiny elements, extracted from these regions, will serve as one of the feature sets. Additionally, we extract color histograms as another feature set to analyze color and texture in detail. Both feature sets from the obverse and reverse sides of the coins are processed using a multi-layer perceptron (MLP) model and a random forest model. The best-performing model is then selected to grade the coins by analyzing their overall wear patterns and color characteristics. Our grading system has demonstrated an accuracy of up to 91.3% in predicting the Sheldon Grading Scale across five coin types, allowing for a grading tolerance of ± 4. For a single coin type (Franklin Half Dollar), it has achieved an accuracy of up to 95.1% with a tolerance of ± 1.

Development of an Automated Coin Grading System: Integrating Image Preprocessing, Feature Extraction, and ML Modeling

Jianzhu Chen

GENERAL AUDIENCE ABSTRACT

For over 70 years, the Sheldon Coin Grading Scale has been crucial for quantifying the value of coins in the large coin collecting industry. Traditionally, coin grading has depended on human graders, which may lead to inconsistent results and variations in coin values. In this thesis, we present an automated coin grading system that uses advanced image processing and advanced machine learning techniques to predict the grade of various coin types. Our system uses synthetic reference masks to identify key areas, such as the contours of the designs on the coins, and detect any irregularities on the surface. We analyze significant details and tiny elements from these areas to form one set of features. Additionally, we extract color information to examine the coin's color and texture. Both sets of features, from the front and back of the coins, are processed using a multi-layer perceptron (MLP) model and a random forest model, which grades the coins by assessing their overall wear and color. Our grading system has demonstrated an accuracy of up to 91.3% in predicting the Sheldon Grading Scale across five coin types, allowing for a grading tolerance of ± 4. For a single coin type (Franklin Half Dollar), it has achieved an accuracy of up to 95.1% with a tolerance of ± 1.

Dedicated to Virginia Tech.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The evaluation of collectible coins is vital for determining their market value and has traditionally relied on the Sheldon Coin Grading Scale (Table.1.1). The Sheldon scale is a grading scale devised to facilitate the coin trading and coin collection business [3]. The categories extend beyond what is explained in Table 1.1. This grading method depends heavily on human expertise to assess wear, luster, and toning. However, the subjective nature of this process often leads to inconsistent grades, affecting the valuation of coins. Moreover, grading services such as PCGS and NGC not only charge significant fees but also require weeks to months to complete the grading process. These issues highlight the need for a more reliable, efficient, and cost-effective automated coin grading method.

1.1 Related Work

Recent advances in computer vision and artificial intelligence have spurred research into automated systems for coin recognition and counterfeit detection [3, 4, 5, 6, 7], though fewer efforts have focused on coin grading systems. One prominent method involves the Scale Invariant Feature Transform (SIFT) algorithm, which is widely used to evaluate the wear on coins by classifying them into categories like "Good," "Fine," and "Circulated" as well as identifying counterfeit coins [3, 7]. However, while SIFT effectively identifies surface preservation, it overlooks other critical aspects, such as strike and toning, which can

Scale	Grade Label	Specifications
3	About Good (AG)	Extremely worn, with parts of the lettering, date, and legend worn smooth. The date may be difficult to read
4	Good (G)	Significantly worn, with the design still visible but faint in some areas, and many details appear flat
8	Very Good (VG)	Heavily worn, but the main features remain clear and bold, though somewhat flattened
12	Fine (F)	Moderate to significant even wear is present, but the entire design remains bold, with an overall appealing appearance
20	Very Fine (VF)	Moderate wear is visible on the high points of the design, but all major details remain clear
30	Choice Very Fine (VF)	The surface and highest points of the design show light, even wear, while all lettering and major features remain sharp
40	Extremely Fine (EF)	The design shows light wear throughout, but all features remain sharp and well-defined, with traces of luster possibly visible
45	Choice Extremely Fine (EF)	Slight overall wear on the highest points, with all design details remaining exceptionally sharp
50	About Uncirculated (AU)	Light wear is evident on several high points, but at least half of the original mint luster remains intact
55	Choice About Uncirculated (AU)	Signs of friction on the raised areas of the design, though most of the original mint luster is still present
60	Uncirculated (MS)	A coin with no signs of wear, though it may display several contact marks, and its surface could be spotted or show reduced luster
63	Choice Uncirculated (MS)	A coin featuring a few noticeable contact marks or blemishes in key focal areas, with potentially diminished luster
65	Gem Uncirculated (MS)	An uncirculated coin of above-average quality, which may be brilliant or slightly toned, with only a few minor contact marks on the surface or rim
70	Perfect Uncirculated (MS)	In pristine new condition, with no signs of wear. The highest possible quality, showing no marks from scratches, handling, or contact with other coins

Table 1.1: Sheldon Coin Grading Scale [2]

significantly impact a coin's grade. Other methods, such as color-based metrics [4], have also been proposed for grading, but they face challenges in capturing the nuanced features of toning, where visual appeal often outweighs straightforward color comparisons.

In another study, Pan and colleagues proposed using deep learning combined with handcrafted filters to quantify “unexpected elements” like scratches or dirt marks [6]. Although their approach achieved results comparable to manual grading, it focused solely on surface preservation and ignored strike quality, another vital factor in coin grading. This thesis extends these efforts by accounting for both “expected” elements, such as strike, and “unexpected” elements, like wear. The method was simplified by replacing handcrafted filters with more efficient edge detection algorithms to capture tiny features, while also integrating color histograms to assess toning. This combined approach significantly enhances the accuracy of grading, particularly in handling the intricate characteristics that influence a coin's final value.

In addition, Atighehchian's paper [3] emphasizes image preprocessing techniques, such as image normalization and background elimination, as critical steps for improving grading accuracy. These methods align with the steps taken in my research, particularly with regard to centering and background removal during preprocessing. Atighehchian's method of background elimination through normalization shares similarities with the use of Hough Circle Detection and masking techniques to ensure that only the coin is analyzed, excluding any irrelevant background. Both approaches aim to improve feature extraction by providing consistent and well-prepared images for subsequent analysis. This highlights the importance of effective image preprocessing in coin grading, where the accurate removal of noise and consistent presentation of the coin's features are essential for the reliability of machine learning models.

Moreover, previous efforts encountered challenges such as the absence of high-resolution images with consistent lighting, which are crucial for accurately identifying and classifying features, and most datasets comprised fewer than 200 samples. However, in this thesis, we utilize approximately 13,500 samples, which provides a more robust foundation for training and evaluating machine learning models. This larger dataset allows for more accurate feature extraction and model generalization, ensuring our system can better handle the nuances of coin grading across various conditions and types.

1.2 Objective

Mint state difference between MS60 to MS70 will be primary focus in this thesis since they represent the most valuable coins and have the highest distribution in the dataset. Grading Mint State coins involves both artistic judgment and scientific methodology, especially at PCGS [8]. The distinctions between grades, such as MS60 and MS61, or even MS60 and MS70, are significant. Unlike the grading system for circulated coins, which primarily considers wear differences to distinguish grades like XF40 from XF45, the Mint State grading system focuses on additional factors. There are four fundamental components that contribute to the grading of Mint State coins [8]:

1. **Strike**: The completeness/incompleteness of a coin's intended detail when originally struck.
2. **Luster**: The strength and pattern of light reflected off a coin.
3. **Surface preservation**: The condition of the surface of a coin, notably marks and/or scratches.
4. **Eye appeal**: The most subjective aspect of coin grading, reflecting the overall visual impression, often manifested as "toning."

This thesis introduces an automated coin grading system designed to address these fundamental elements using an image processing framework. This framework can identify both major and minor wear features on different types of coins. It also uses synthetic reference masks to classify “expected” and “unexpected” elements which can evaluate a coin’s strike and surface preservation. The proposed system divides each coin into eighteen areas for detailed examination of factors that affect the coin’s grade. Moreover, this grading system focuses on strike, surface preservation, and eye appeal, with the color histogram serving as a key element to evaluate toning. Unlike luster, which is challenging to quantify, these aspects are measurable and integral to the evaluation. Additionally, the proposed coin grading system improves upon previous methods by utilizing advanced machine learning techniques. This system is robust and provides a more detailed evaluation, working effectively with various types of coins, whether they are made of copper, nickel, or silver.

Chapter 2

Image Acquisition And General Image Preprocessing

2.1 Source of the Dataset

David Lawrence Rare Coins (DLRC) was chosen as the main source for our dataset because of its strong reputation in the numismatic community. DLRC offers a wide range of collectible coins, known for the high quality of its images, typically around 1000 x 1000 pixels. This resolution reveals critical details such as scratches, environmental damage, dirt marks, and toning—key for accurate coin analysis and grading by automated systems. Most coins from DLRC also feature superior focus and illumination. DLRC also provides detailed information for past auction coins, including proof status, Sheldon Coin Grading Scale, toning degree, eye appeal, grading service details, any signs of environmental damage, cleaning details, and more [9]. We selected a diverse range of coin types from DLRC for our proposed system, including the Franklin Half Dollar (FHD), Mercury Dime, Buffalo Nickel, Washington Quarters, and Morgan Silver Dollar. All these types of coins were scraped using Python from DLRC in March 2024. These coin types were chosen due to the extensive number of available images, their varied materials, and their widespread recognition and popularity in the market.

In table 2.1, the total number of images, approximately 13,500, provides a robust foundation for training a machine learning model, especially when considering that this dataset spans various coin types and year ranges. The diversity and quantity of the images should facilitate effective feature extraction, allowing the model to learn subtle differences in grading across different coins and time periods. However, given the constraints of available sources, the current dataset represents the maximum number of high-quality images that can be obtained from DLRC. Other sources of coin images are limited both in quantity and resolution, making DLRC the most viable and reliable option for this study.

Type of Coin	**Number of Images**	**Year Range**	**Material**
Franklin Half Dollar	2754	1948-1963	90% Silver, 10% Copper
Mercury Dime	2308	1916-1945	90% Silver, 10% Copper
Buffalo Nickel	3061	1913-1938	75% Copper, 25% Nickel
Washington Quarter	2268	1932-1964	90% Silver, 10% Copper
	125	1965-1998	75% Copper, 25% Nickel
Morgan Silver Dollar	2989	1878-1904	90% Silver, 10% Copper

Table 2.1: Coin Specifications of Our Dataset

2.2 Proof & Non-Proof Coins

Proof coins are made with a special, mirror-like finish for collectors, not for everyday use. They are made using polished dies and struck multiple times to bring out more detail, giving them outstanding visual quality that is usually well-preserved. On the other hand, non-proof coins are made for daily use and are struck only once with standard dies, giving them a simpler finish [9]. These coins often vary in how worn they are, from like-new to very worn, which makes grading them automatically a more complex task [10].

Non-proof coins are chosen for grading because they exhibit a wide range of conditions due to circulation, making grading essential to assess their value, whereas proof coins are typically in uniformly high condition and less reliant on grading for differentiation. This focus helps us develop a system that can accurately grade the wide range of wear found on these coins.

(a) Non-proof coin

(b) Proof coin

Figure 2.1: Sample of Non-proof and proof coins

2.3 Image Preprocessing

Image preprocessing is the first important step in developing our automated grading system. The main purpose of this step is to make sure all input images are standardized and are set up well for the next steps. This uniformity is particularly important for later generating a reference mask from well-oriented, top-grade coins. In our research, image preprocessing involves several crucial tasks:

The standardized approach involved the following:

1. **Consistent Image Size:** All images are resized to a consistent resolution of 1000 x 1000 using the Lanczos resampling method.

2. **Hough Circle Detection:** The segmentation is achieved through a robust image preprocessing technique known as Hough Circle Detection [11]. Each image from the dataset is converted to grayscale and blurred to smooth out high-frequency noise. Sobel edge detection is then applied to the blurred image to highlight the edges of the coin. Subsequently, an experimental threshold value is applied to the images to enhance segmentation. The Hough Circle Detection algorithm is then used on the thresholded images to identify potential circles. In Fig.2.2a, the outer edge of the coin is detected.

3. **Image Centering and Background Removal:** Once a circle representing the coin is detected, a mask is created to obscure everything except the coin itself. The image is then adjusted to center the coin.

4. **Scaling and Final Adjustments:** The centered coin is resized to maintain a uniform radius across all images using bicubic interpolation. We ensure that each processed image has the coin perfectly centered with a standardized diameter (900 pixels), making it easier for the grading model to assess the coin's features accurately with matched masks.

(a) Hough Circle Detection (b) Coin Image After Image Preprocessing

Figure 2.2: Image Preprocessing

Chapter 3

Reference Masks and Sub-Regions Masks

3.1 “Expected” Element and “Unexpected” Element Masks

“Expected” Element should be the overall relief of coins which represent the strike element. The anatomy of a coin can be described by the following components: [1]:

- **Obverse & Reverse**: The front side, known as the ”obverse” or ”heads,” and the back side, referred to as the ”reverse” or ”tails,” of a coin.
- **Edge**: The outer border of a coin is called the edge, which can be plain, reeded, lettered, or decorated.
- **Rim**: The raised part of the edge on both sides of a coin that helps protect the coin’s design from wear.
- **Inscription**: The primary words or lettering on a coin, also known as the legend.
- **Mint Mark**: A small letter or symbol on a coin that indicates where it was minted. Current U.S. mint marks include P for Philadelphia, D for Denver, S for San Francisco, and W for West Point.

- **Relief**: The part of a coin's design that is raised above the surface.
- **Field**: The flat area of a coin's surface that is not occupied by design or inscription is called the field.

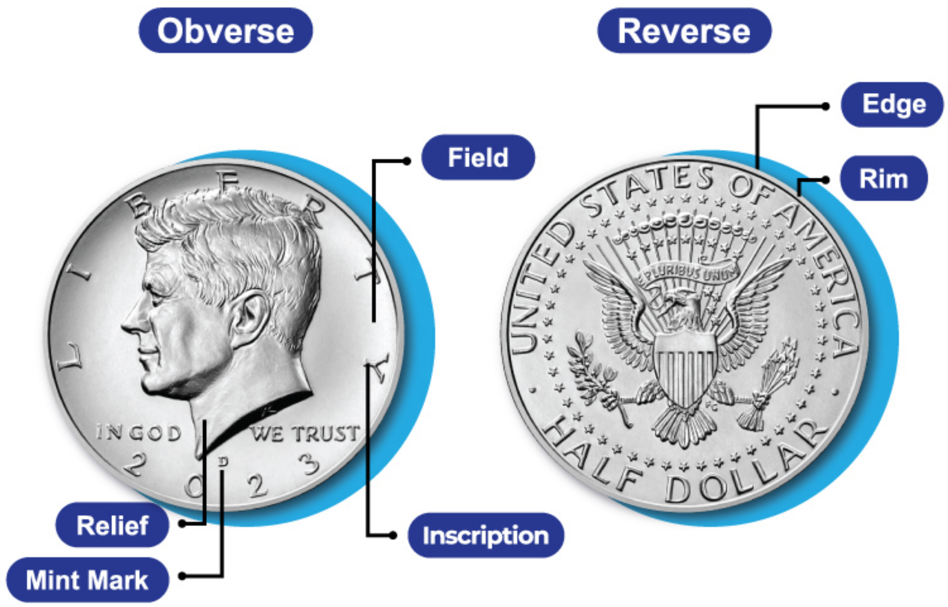


Figure 3.1: Anatomy of a Coin [1]

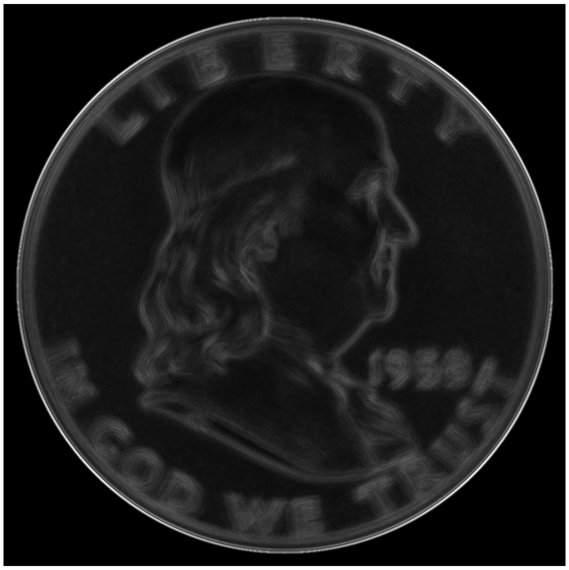
In the pursuit of automating coin grading, constructing accurate reference masks can significantly impact the performance of the system, as it can evaluate two fundamental components: strike and surface preservation. These masks assist in identifying and quantifying "unexpected elements" such as scratches, dirty marks, and wear that can represent a coin's condition of surface preservation, as well as "expected elements" like the relief of contours that represnt the coin's strike, which clearly distinguish between top-grade and lower-grade coins. Inspired by the techniques discussed by Pan and Tougne [6], we have refined their methods, using high-grade and well-oriented coins to create synthetic reference masks. Our approach introduces several methodological refinements aimed at enhancing the precision of these masks. All top-grade, well-oriented coins from each type were selected since they represent the ideal state of surface preservation, thereby serving as a benchmark for grading other coins. All reference coin images underwent a series of preprocessing steps:

1. **Gaussian Blurring:** To minimize noise and eliminate minor imperfections like micro scratches, Gaussian blurring is applied using a (9, 9) kernel and sigma values of 1.5. The six-sigma rule (kernel size $\approx 6 \times \sigma$), a statistical principle defining the range within which most data points in a normal distribution lie, ensures that significant features are retained while high-frequency components are filtered out. This optimizes the blur's effectiveness by capturing nearly all important coin features within the chosen kernel size [12].

2. **Mean Gradient Map Images and Mean Thresholded Images:** Unlike the Laplacian operator used in the referenced study [6], we employed the Sobel edge detection method using a 3x3 kernel size. The Sobel operator provides a better approximation of gradient vectors in an image, which are essential for accurately outlining the primary contours of the coin [13]. Following edge detection, we applied Otsu's thresholding to convert the gradient images into binary images. This method automatically determines the optimal threshold value to distinguish between the foreground (contours of relief) and the background, thus enhancing the visibility of contours of the coin [14]. After processing all the images, we generated mean thresholded images and mean gradient map images, as shown in Fig. 3.2

3. **Combination of Mean Gradient Map and Thresholded Images:** A composite image is then created by combining the mean thresholded and gradient map images. We applied Otsu's thresholding again to convert the composite image into a binary image, as shown in Fig. 3.3b

4. **Morphological Closing:** We applied the morphological closing function to connect



(a) Mean Gradient Map

(b) Mean Thresholded Image

Figure 3.2: Edge Detection

(a) The Composite Image

(b) The Thresholded Image

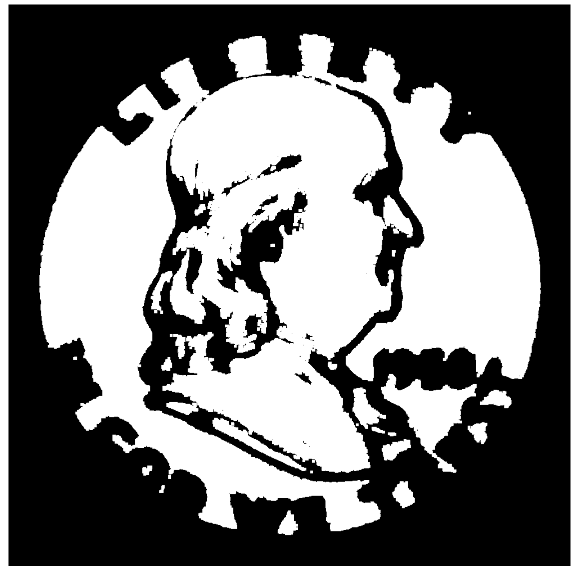
Figure 3.3: The composite Image and Its Thresholded Image

closely lying elements in the previous binary image using a 5x5 rectangular kernel, which is crucial for creating a continuous and cohesive mask that includes all relevant contours without fragmentation.

5. **Final "Expected" Element Mask & "Unexpected" Element Mask:** The DLRC designs specific holders to secure the coins, which results in the prongs also being captured in the images. To reduce noise from the prongs, we adjust the diameter of the mask, effectively eliminating this noise. This adjustment produces the final "expected" element mask. After obtaining this mask, we generate the "unexpected" element mask by inverting the "expected" element mask. The resulting "unexpected" mask highlights non-uniformities on the coin's surface, identifying areas that deviate from the "expected" contours.

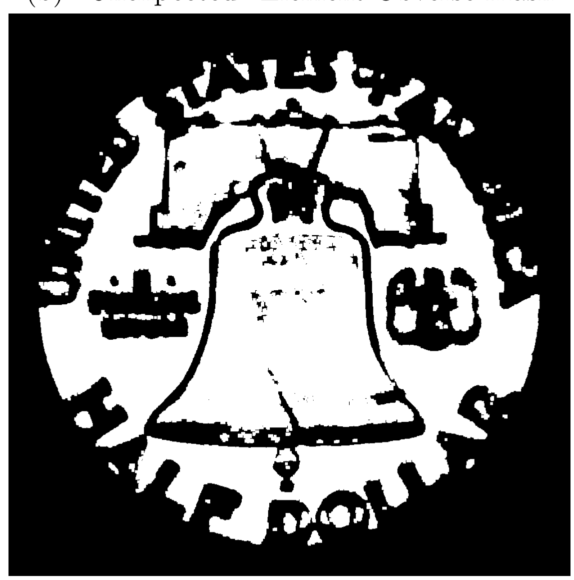
The procedure described above is almost uniformly applied to all coin types on obverse and reverse in our dataset to generate both "expected" and "unexpected" element masks. Below, we present the element masks on the Morgan Silver Dollar, as shown in Fig. 3.5

(a) “Expected” Element Obverse-Mask



(b) “Unexpected” Element Obverse-Mask

(c) “Expected” Element Reverse-Mask



(d) “Unexpected” Element Obverse-Mask

Figure 3.4: Element Masks on Franklin Half Dollar

(a) "Expected" Element Obverse-Mask

(b) "Unexpected" Element Obverse-Mask

(c) "Expected" Element Reverse-Mask

(d) "Unexpected" Element Reverse-Mask

Figure 3.5: Element Masks on Morgan Silver Dollar

3.2 Eighteen Sub-Region Masks

Different areas of a coin can have varying impacts on its final grade due to distinct features and conditions they may exhibit. For example, the portrait of Benjamin Franklin and the date are the most important areas for grading a Franklin Half Dollar (FHD) coin. In contrast, the edge of almost every type of coin is generally considered to be of lesser importance [15]. To address this, we use three concentric circles with radii of 254, 360, and 440, along with six radial lines intersecting at the center. This creates a total of 18 wedge-shaped sections,as shown in Fig. 3.6, ensuring that each sub-region has a similar area and uniformly covers relevant parts of the coin. This segmentation ensures comprehensive coverage of all areas of the coin, allowing for detailed analysis and feature extraction from each specific region. By treating each region individually, the system can apply distinct weights to the features detected in each area based on their significance to the overall grade of the coin. Each sub-region will serve as an individual mask for later analysis. This design has several advantages:

- The outer regions can cover the rim, edge, and inscription (legend) on the coin, such as *"LIBERTY"*, and *"IN GOD WE TRUST"*, as shown in Fig. 3.7a, allowing the system to classify the principal words or lettering on a coin separately.
- The middle regions often capture non-uniformities (field) unrelated to the contours of the coin, making them ideal for analyzing surface preservation.
- The composition of the middle and inner regions often captures the portrait and relief. For example, some regions focus on the sharpness and detail of the portrait's hair, and facial features, providing detailed information on a coin's strike condition.

Furthermore, this approach allows the system to be adaptable to different types of coins without the need for coin-specific masks. Instead of creating and weighting masks for specific areas which vary from one coin type to another, the system utilizes a uniform approach where the machine learning model learns to recognize and appropriately weight the importance of features across various coin types. This method not only enhances the system's flexibility and applicability across diverse coins but also improves the accuracy and performance of the grading process.

The decision to use 18 regions instead of 24 was driven by the need to balance granularity with model performance. Increasing to 24 regions would add 48 more features, risking overfitting given the dataset's average size of around 2,600 images per coin type. While 24 regions might capture finer details, this would require either a larger dataset or advanced regularization techniques to prevent overfitting. Future research could explore this approach to evaluate its potential benefits for grading accuracy.

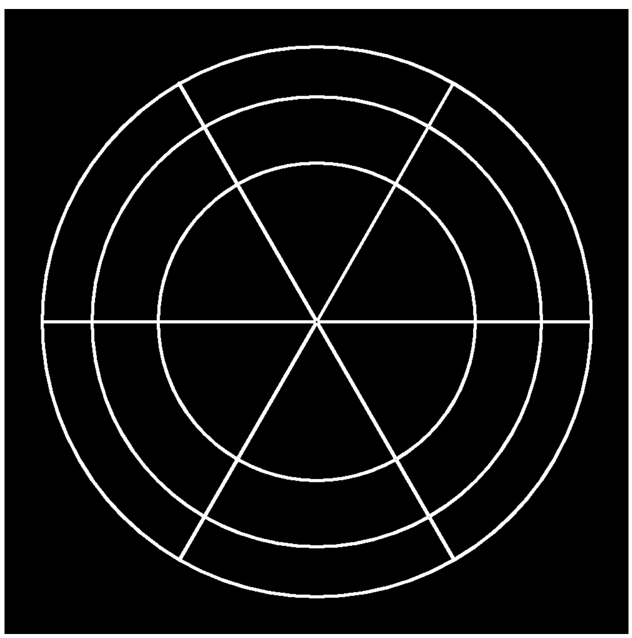
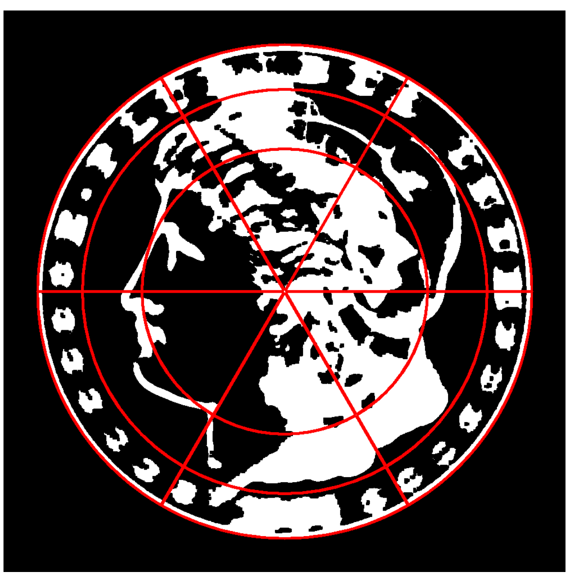


Figure 3.6: Eighteen Sub-Regions

(a) Sub-Regions on Expected Element Obverse Mask

(b) Sub-Regions on Expected Element Reverse Mask



(c) Sub Regions On Obverse-MSD

(d) Sub-Regions on Reverse-MSD

Figure 3.7: Eighteen Regions On Element Masks

Chapter 4

Significant and Tiny Element

The grading process involves not only recognizing the general condition of the coin but also identifying features that significantly influence its value. These features are categorized into "significant elements" (SE) and "tiny elements" (TE), which have distinct impacts on the grading outcome of the coins' strike and surface preservation. Edge detection is an ideal method to detect and classify significant and tiny elements separately. However, the gradient map technique can only identify the external contours of a large dirty mark but may miss the entire area, whereas thresholding methods might fail to detect minor scratches that aren't sufficiently dark [6]. To address this, both techniques are integrated to achieve comprehensive detection, similar to Pan and Tougne's method.

4.1 Significant Elements (SE)

Significant elements are prominent and easily observable relief of coins or noticeable imperfections, such as large scratches, shock marks, and dirty marks. These elements are generally more apparent and can drastically affect a coin's grade because they are significant deviations from the coin's expected strike or ideal surface preservation. Below are the steps for detecting the significant elements:

1. The process begins with applying Gaussian blur to each image with a kernel size

of $(7,7)$ to obtain the blur image $I_{(7,7)}$. This specific kernel size is selected because it effectively smooths the image, filtering out micro features that do not count as significant elements while preserving large scratches and important relief details of the coin. We also applied the six-sigma rule—by choosing a 7x7 kernel and using a sigma of 1.17 (since kernel size $\approx 6 \times \sigma$), ensuring that the Gaussian distribution is well captured within the kernel size.

2. Next, gradient images I_G are obtained by applying Sobel edge detection. The resulting gradient magnitudes highlight the edges and transitions in intensity.

3. The gradient magnitude image I_G is then thresholded using Otsu's method to create a binary image I_{SE}, which represents significant elements where it can isolate areas with large and noticeable features.

4.2 Tiny Element (TE)

Tiny elements are small scratches and subtle changes in texture that blend into the coin's surface, making parts of it look grainy [6]. While less impactful than significant elements (SE), they still affect the coin's grade. Tiny elements are hard to spot individually because they mix with the coin's texture. This texture is complex due to the metal, the minting process, mild abrasions, and patina. These features vary between coins and even different areas on the same coin. Graders say that areas with tiny elements look more "grainy" compared to smooth areas without these tiny scratches[6]. Below are the steps for detecting the tiny elements:

1. A similar method is used with slight modifications. Instead of using a $(7,7)$ kernel size for the Gaussian blur, a smaller kernel size of $(5,5)$ and and a sigma of 0.83 are

applied to obtain the blurred image $I_{(5,5)}$.

2. Following the same steps as for significant elements, gradient images I_{G2} are obtained by applying Sobel edge detection to the blurred image.

3. The gradient magnitude image I_{G2} is then thresholded using Otsu's method to create a binary image I_R.

4. The final tiny element image $I_{(TE)}$ is derived by subtracting the significant element image $I_{(SE)}$ from the binary image I_R.

4.3 Elements on Higher-Grade Coins and Lower-Grade Coins

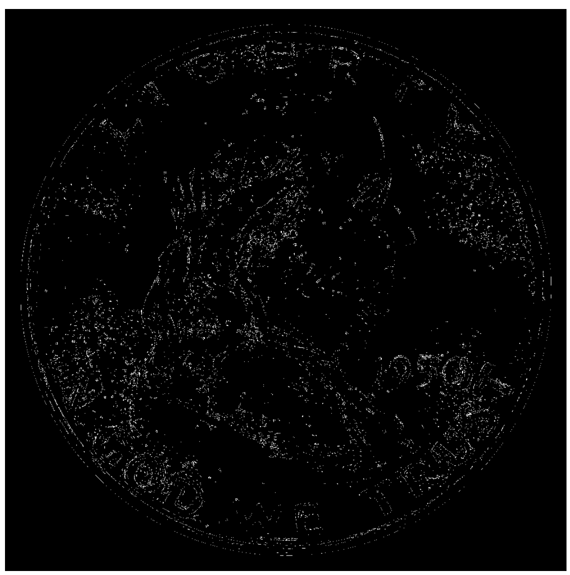
A higher-grade coin (MS67 coin in Fig. 4.1a) often exhibit a well-preserved surface with minimal visible wear, maintaining its original luster and detailed features, while lower-grade coin (AU50 coin in Fig. 4.2a) usually show noticeable wear and loss of contours, with evident scratches and marks on its surface [16].

The SE image (in Fig. 4.1b) of the MS67 coin shows prominent and clear contours of the coin's relief features, including the edges and fine details. There are few significant scratches or marks, indicating the coin's high grade. However, the SE image also reveals areas with minimal edges, likely due to glare and reflections caused by the coin's highly reflective surface. This may introduce bias, as these regions appear smoother than they are due to the lighting setup [17]. The SE image (in Fig. 4.2b) of the AU50 coin reveals numerous significant scratches and marks. The contours of the relief features are less pronounced compared to the MS67 coin, and there are more interruptions in the surface details.

The TE image (in Fig. 4.1c) for the MS67 coin highlights minimal micro-scratches and abrasions. The surface appears mostly smooth with little graininess, confirming the coin's high grade. Similar to the SE, the TE image may also be affected by lighting bias, showing fewer tiny elements in highly reflective areas. The TE image (in Fig. 4.2c) for the AU50 coin displays a higher density of micro-scratches and abrasions. The surface is much grainier, reflecting more extensive wear and lower grade.

(a) MS67 Coin

(b) SE

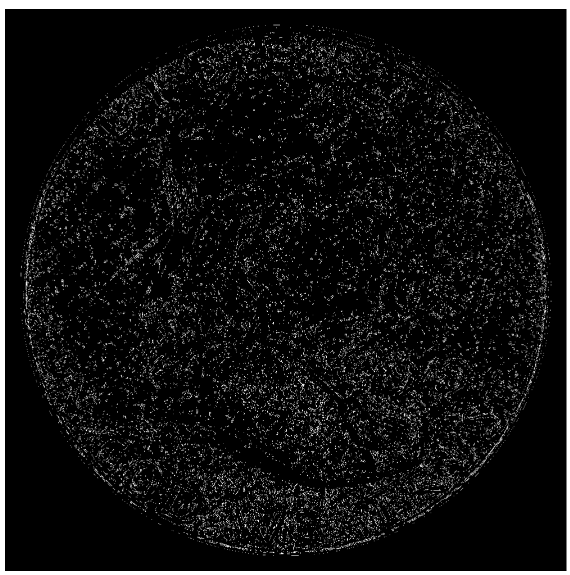


(c) TE

Figure 4.1: MS67 Coin - SE and TE

(a) AU50 Coin

(b) SE



(c) TE

Figure 4.2: AU50 Coin - SE and TE

4.4 Masks on the Significant & Tiny Elements

After obtaining the significant elements and tiny elements through edge detection and image segmentation, the next step involves applying the "expected" element mask and "unexpected" element mask to extract four key features: significant "expected" element, significant "unexpected" element, tiny "expected" element, and tiny "unexpected" element, as shown in Fig. 4.3.

- **Significant "Expected" Element (SEE)**: The significant "expected" element refers to the prominent strike of the coin that are supposed to be present in a top-grade coin. By applying the "expected" element mask to the significant element image, we isolate these intended features, ensuring that only the large, expected contours and relief details are highlighted.

- **Significant "Unexpected" Element (SUE)**: The significant "unexpected" element consists of the prominent defects and wear marks that should not be present in a top-grade coin. These include large scratches, shock marks, and significant dirty marks. By applying the "unexpected" element mask to the significant element image, we isolate these undesirable features. This helps in quantifying the amount of significant damage on the coin.

- **Tiny "Expected" Element (TEE)**: The tiny "expected" element might be small relief details and textures that are part of the coin's original design. By applying the "expected" element mask to the tiny element image, we isolate these finer details, ensuring they are accounted for in the grading process.

- **Tiny "Unexpected" Element (TUE)**: The tiny "unexpected" element refers to small defects and wear marks that are less noticeable but still affect the coin's grade.

These include micro scratches and subtle abrasions that make the surface grainy. By applying the “unexpected” element mask to the tiny element image, we isolate these minor defects, which helps in assessing the overall wear and tear on the coin.

4.5 Dividing Elements into Regions and Feature Calculation

Once the four types of elements (SEE, SUE, TEE, TUE on the obverse and reverse of the coin) are identified, each element is divided into 18 sub-regions to capture the distribution of features across different areas of the coin, as shown in Fig. 4.3. For each sub-region, the number of while pixels—representing the number of intensity corresponding to SEE, SUE, TEE, and TUE elements—are calculated. The number of while pixels from all 18 sub-regions for each element (SEE, SUE, TEE, TUE on obverse and reverse of the coin) form one of the feature sets for our machine learning model. This feature set evaluates the trace of wear and how well the coin is preserved. This comprehensive feature set ensures accurate assessment of the coin’s grade based on both significant and tiny, expected and unexpected elements. Additionally, another feature set used in our model is the HSV histogram, which captures the color information of the coin. This will be introduced in detail in Chapter 5.

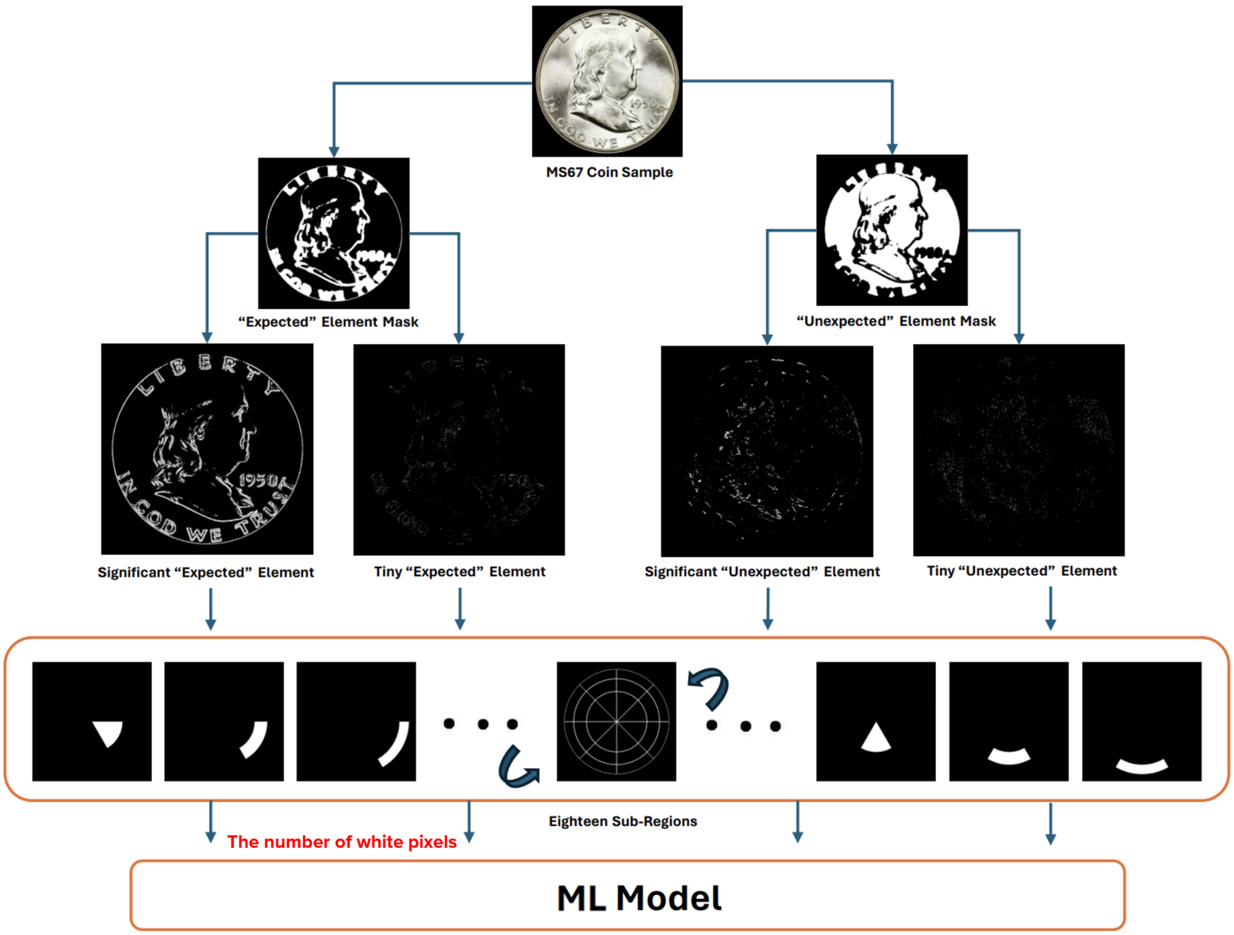


Figure 4.3: Feature Extraction Framework For the Grading System

4.6 Rotation and Alignment

Most images in our dataset do not have optimal orientation, which poses a challenge for accurate analysis. Both the "expected" element mask and the "unexpected" element mask are highly sensitive to the orientation of the images. To ensure consistent and reliable results, we implemented a transformation process to adjust the orientation of the images based on the SUE detected.

1. **Generating Rotated Images**: Each image was rotated in steps from -5 degrees to

+5 degrees, with increments of 0.1 degrees, resulting in 100 rotated versions of each image.

2. **Detection of SUE Element:** For each rotated image, we applied the “unexpected” element mask to each significant element image to isolate the SUE. This resulted in a SUE element binary image for each rotation.

3. **Calculating Intensity Values**: The number of white pixels in the SUE element binary image was calculated for each rotated version. The orientation with the least number of intensity pixels was considered the optimal orientation, as it indicated minimal unexpected elements due to proper alignment.

In Fig. 4.4, the image on the left (a) shows a coin before the transformation, while the image on the right (b) shows the same coin after the transformation. This process adjusts the coin's orientation, making details like the date more legible and consistent. Standardizing the orientation of the images significantly improves the accuracy of our feature extraction and grading. By aligning each coin image uniformly, we ensure that the subsequent analysis is based on a consistent reference point, which is crucial for the reliability of the automated grading system.

(a) Before Rotation

(b) After 4.3 Degree Rotation

Figure 4.4: Sample of Coin Rotation

Chapter 5

Color Histogram & Toning

Color affects the coin's grade in several ways. Toning, which refers to the color changes that occur on a coin's surface due to oxidation and other environmental factors, can either enhance or diminish its appeal and value [18]. The toning and the uniformity of the color distribution are critical in assessing the coin's overall condition [15]. Therefore, exploring the color histograms (hue, saturation, and value) of coins can help our proposed automated grading system evaluate a coin's Sheldon Grading Scale. In this chapter, we explore the extraction of color features from coin images—focusing on the HSV histogram—and their correlation with the Sheldon Coin Grading Scale and degree of toning.

5.1 Hue, Saturation, and Value (HSV)

Hue represents the type of color (red, blue, green) and is essential for identifying specific color characteristics of the coin. Different hues can indicate various states of preservation and toning. Saturation measures the intensity or purity of the color. High saturation indicates vibrant colors, while low saturation indicates dull or muted colors [19]. The saturation of a coin's color can reflect its exposure to elements and overall wear.

Value refers to the brightness or darkness of a color and is an important aspect of a coin's luster. However, value can introduce bias in coin grading due to varying lighting conditions during image capture. In Fig. 5.1, the areas highlighted in red represent glare or white pixels caused by lighting, which do not accurately reflect the coin's true color features. Pixels with low saturation and high intensity typically indicate glare, while those with low saturation and low intensity suggest dark shadows. To address these challenges, we filter out pixels with a saturation level below 15, which helps reduce the impact of both glare and shadows, ensuring that the color features used for grading are more accurate and consistent. However, this approach presents a particular challenge for untoned silver coins, as they naturally exhibit low saturation across the surface.

Figure 5.1: Sample of Glare on Coin Images

After filtering out the pixels, a comparison between higher-grade and lower-grade coins histogram is generated. In Fig. 5.2, it illustrates the differences in color histograms between all MS67 coins and AU58 coins on Franklin Half Dollar, highlighting how these variations correspond to the coin's condition (degree of toning) and grading.

- **Hue:** The higher-grade and lower-grade coins show distinct distributions in their hue values. Higher-grade coins have a higher frequency of lower hue bins compared to lower-grade coins. This suggests that higher-grade coins, typically in better condition, may have a more pronounced and consistent toning pattern, indicating less exposure to elements or handling. In contrast, the wider distribution of hues in lower-grade coins could imply more variation in toning, often due to wear or environmental exposure.

- **Saturation:** The saturation histograms reveal a significant difference in color intensity between higher-grade and lower-grade coins. Higher-grade coins exhibit higher frequencies in lower saturation bins, indicating more vibrant colors. This vibrancy is associated with coins in better condition. On the other hand, lower-grade coins show higher frequencies in higher saturation bins, indicating more muted and less intense colors, which can be a result of oxidation or wear over time.

- **Value:** The value histograms for higher-grade and lower-grade coins show overlapping distributions, but with some differences in frequency across the bins. Higher-grade coins tend to have a slightly higher frequency in mid-range value bins, suggesting a balance in brightness that reflects good preservation conditions. Lower-grade coins show a broader distribution, implying that lighting conditions and potential wear might affect the perceived brightness or darkness of these coins.

Hue and saturation provide more critical information about a coin's condition compared to value. Subtle variations in hue and saturation can indicate different levels of toning and wear, which are crucial for grading [15]. However, lower-grade and higher-grade coins show overlapping distributions in the value color, suggesting it might be less correlated with coin grading. Therefore, we decided to use more bins of hue and saturation than value in our feature set.

Moreover, utilizing all bins from the hue, saturation, and value histograms would excessively complicate the task and lead to overfitting, especially given the limited dataset [20]. Consequently, we selected 36 bins for hue, 26 bins for saturation, and 18 bins for value (Fig. 5.3)

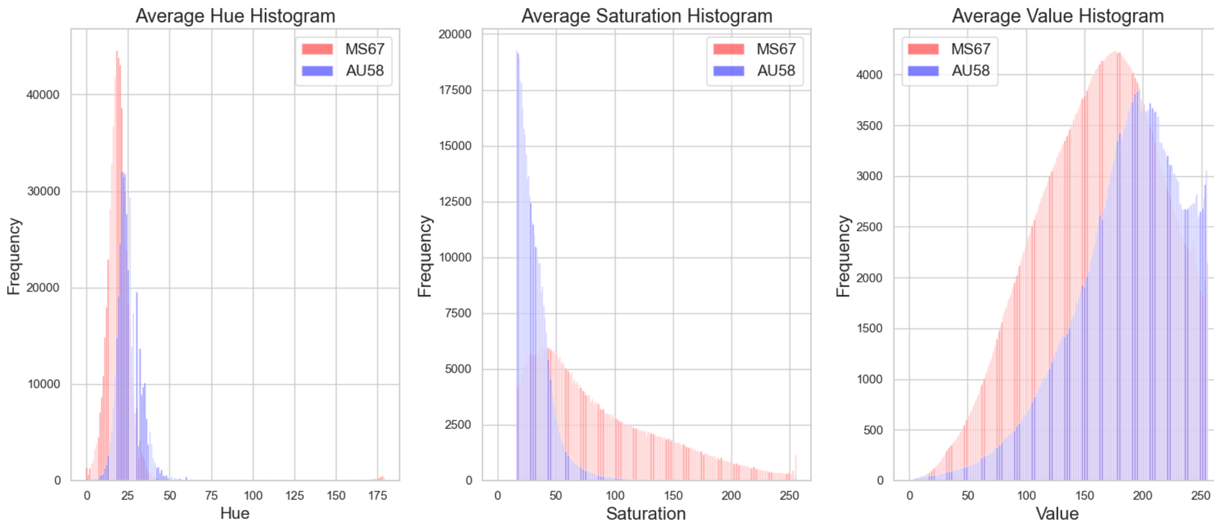


Figure 5.2: Average HSV Histograms of All MS67 and AU58 FHD Coins

as the color feature set for the input of the models. This feature set is sufficient to assess a coin's overall degree of toning and wear. By carefully choosing the number of features, we reduced the model's complexity and enhanced the MLP model's ability to generalize to unseen data.

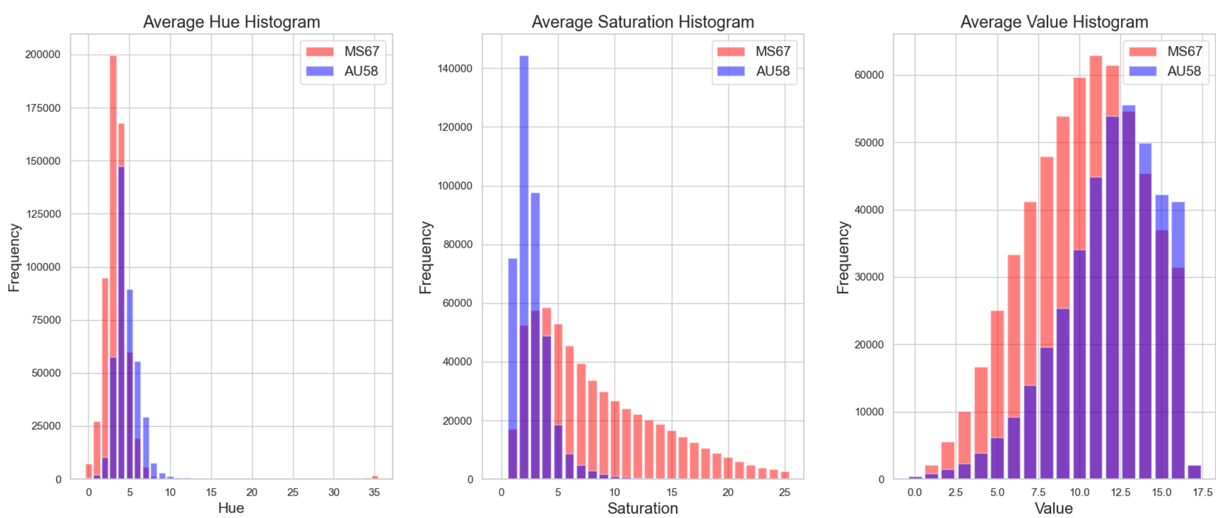


Figure 5.3: Average HSV Histograms of All MS67 and AU58 Franklin Half Dollar Coins With Fixed Bins

5.2 Degree of Toning

Eye appeal, often manifested as "toning," is one of the fundamental elements of coin grading. It refers to the overall aesthetic quality of a coin. [21]. Eye appeal contributes about 10% to a coin's overall grade during the standard grading process. Thus, one of its aspects—the toning— also plays a crucial role in determining final market value upon sale significantly [21].

The degree of toning on a coin, also called its color rating, shows how much the coin's appearance has changed since it was minted. This degree can be rated on a scale from one to five [22]:

- **Degree 1 toning:** They usually shows the original color of the coin at time of minting or pure "blast white" silver coins (Fig. 5.5a).
- **Degree 2 toning:** They show at least 75% white for silver coins. Coins with copper typically starts at a degree 2 toning (Fig. 5.5b).
- **Degree 3 toning:** They show 25 - 50% toning. Coins with copper usually display red-brown color (Fig. 5.5c).
- **Degree 4 toning:** They show more than 50% toned. Coins with copper show brown color (Fig. 5.5d).
- **Degree 5 toning:** They shows very strong and noticeable toning. Coins with copper show deep brown color (Fig. 5.5e).

Even if two coins share the same degree of toning, the coin with more attractive and even toning may receive a higher grade within the same wear category, as toning can be classified as either positive or negative [23]. Positive toning—coins with beautiful, desirable toning, such as vibrant rainbow colors or natural patinas—are often highly sought after by collectors and can command higher prices (Fig. 5.4b). On the other hand, negative toning—coins with unattractive toning, such as dark, splotchy, or uneven toning—may be less appealing and therefore fetch lower prices (Fig. 5.4a). If the toning is seen as unattractive, it can decrease both the grade and the value [23]. However, there is only one metric to measure the degree of toning without distinguishing between positive and negative toning. Thus, we decided to use the color histogram as the input feature set to predict the degree of toning alone, and to serve as one of the feature sets for predicting the Sheldon Grading Scale.

(a) Degree 4 Toning; Grade: 50

(b) Degree 4 Toning; Grade: 65

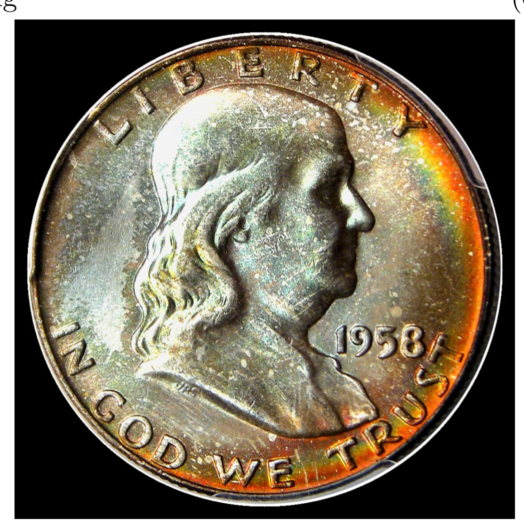
Figure 5.4: Sample of Positive (Right) And Negative (Left) Toning

(a) Degree 1 Toning

(b) Degree 2 Toning

(c) Degree 3 Toning

(d) Degree 4 Toning



(e) Degree 5 Toning

Figure 5.5: Samples of Different Degrees of Toning Rated on Franklin Half Dollar Coins

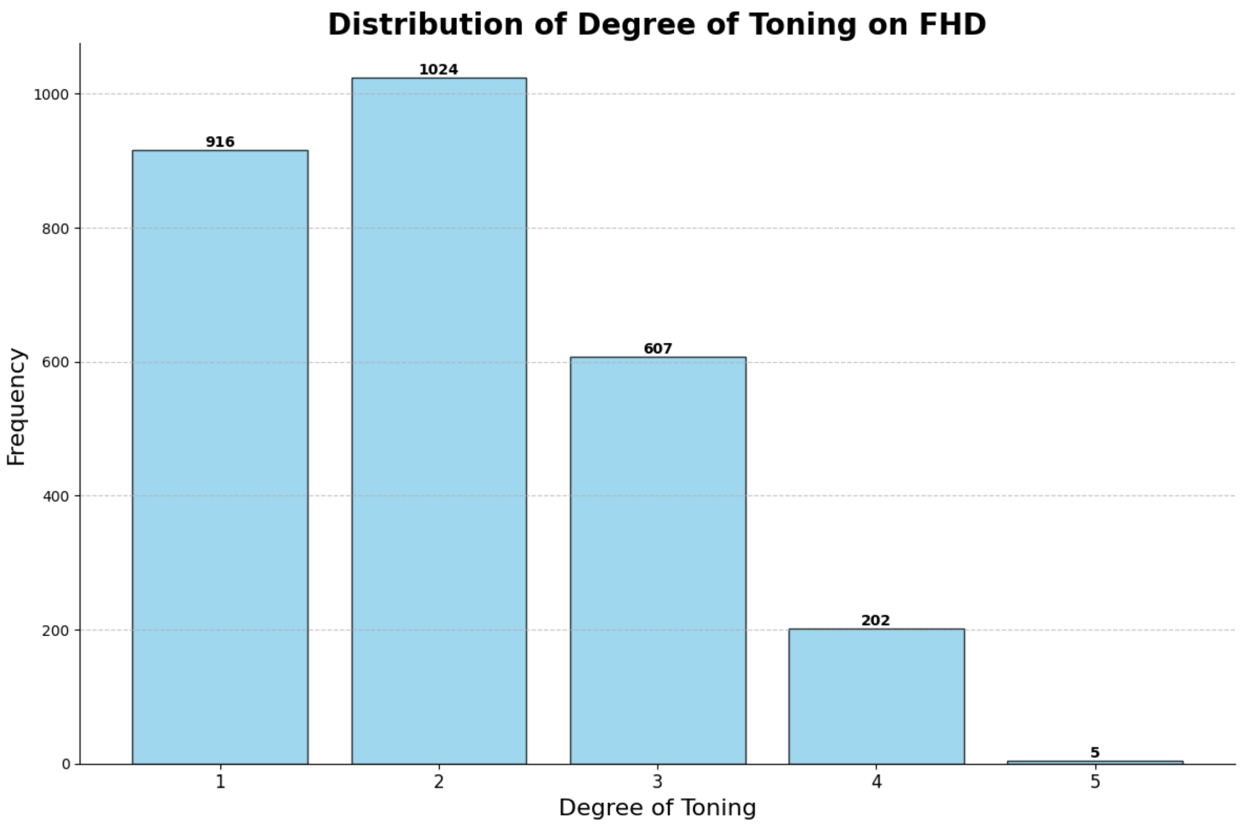


Figure 5.6: Distribution of Degree of Toning on Franklin Half Dollar

Chapter 6

Multi-Layer Perceptron (MLP) & Random Forest Model Design

6.1 MLP for Sheldon Coin Grading Scale Prediction

The Sheldon Coin Grading Scale is a numerical grading system ranging from 1 to 70, which assesses the quality and condition of coins. Accurately predicting coin grades using this scale requires a model capable of understanding subtle variations and patterns in the data. The Multi-Layer Perceptron (MLP) model is particularly well-suited for this task. A MLP model is a type of feedforward artificial neural network that consists of multiple layers of nodes, or neurons, with each layer fully connected to the next. These layers include an input layer, one or more hidden layers, and an output layer. The primary function of an MLP is to map sets of input data onto a set of appropriate outputs. MLPs are particularly adept at solving problems that are not linearly separable due to their ability to learn complex, non-linear relationships between inputs and outputs [24]. The MLP model is especially appropriate for this task for some important reasons:

- **Handling High-Dimensional Data**: The dataset used for coin grading consists of over 2000 samples for every type of coin, each with 304 features. MLPs are effective at handling high-dimensional data and can process large amounts of information simul-

Category	**Important Features**	**Number of Feature Sets**
18 Sub-Regions	Significant "Unexpected" Elements	18
	Significant "Expected" Elements	18
	Tiny "Unexpected" Elements	18
	Tiny "Expected" Elements	18
Color Histograms	Hue	36
	Saturation	26
	Value	18
One Side of the Coin		152
Total (Both Sides of the Coin)		304

Table 6.1: Distribution of The Feature Sets

taneously. This capability allows the model to capture the intricate details necessary for accurate grading [25].

- **Multi-Class Classification & Regression Capability**: We developed both multi-class and regression MLP models for our grading system. Coin grading is inherently a multi-class classification problem, as the model needs to classify coins into one of the many possible grades on the Sheldon Coin Grading Scale. The output layer of the MLP uses a softmax activation function. Each output node represents a class, and softmax ensures that the sum of all probabilities equals 1, making it easier to interpret the predicted probabilities for each class [26]. This property allows the model to assign confidence scores to each possible grade, aiding in accurate classification. In addition to classification, predicting coin grades can also be approached as a regression problem, where the model predicts a continuous grade value. The flexibility of MLPs allows them to be adapted for regression tasks as well by using appropriate loss functions (e.g., mean squared error) and activation functions in the output layer [27]. This dual capability of handling both classification and regression makes MLPs highly versatile for coin grading applications.

6.2 Addressing Class Imbalance with SMOTE

The distribution of the Sheldon Coin Grading Scale in the dataset is highly imbalanced (Fig. 6.1 and Fig. 6.2), with certain grades, particularly higher ones, being overrepresented. This imbalance can lead to biased predictions, where the model is more accurate for the frequent classes and less so for the underrepresented ones [28]. To mitigate this issue, the Synthetic Minority Over-sampling Technique (SMOTE) was initially applied on MLP and Random Forest model. Specifically, a partial SMOTE approach was used: minority classes were balanced to 50% of the majority class, and only those minority classes with more than 10 samples were adjusted.

When dealing with five different types of coins, stratification will be employed to ensure that the SMOTE method is effectively applied to each coin type, thereby maintaining balanced representation across the dataset. In addition, we will evaluate the model's performance both with and without the application of the SMOTE method to determine its impact on prediction accuracy

6.3 Data Preprocessing

Before training the MLP and random forest for the multi-class and regression models, the dataset (feature sets obtained from feature extraction and color histogram) undergoes several preprocessing steps: Features are standardized to have a mean of 0 and a standard deviation of 1, which is crucial for ensuring that the model converges efficiently during training. Labels are then encoded into categorical format for multi-class classification tasks or continuous numerical values for regression tasks. When comparing models with and without partial SMOTE, SMOTE is only applied to the training set to balance the dataset by generating

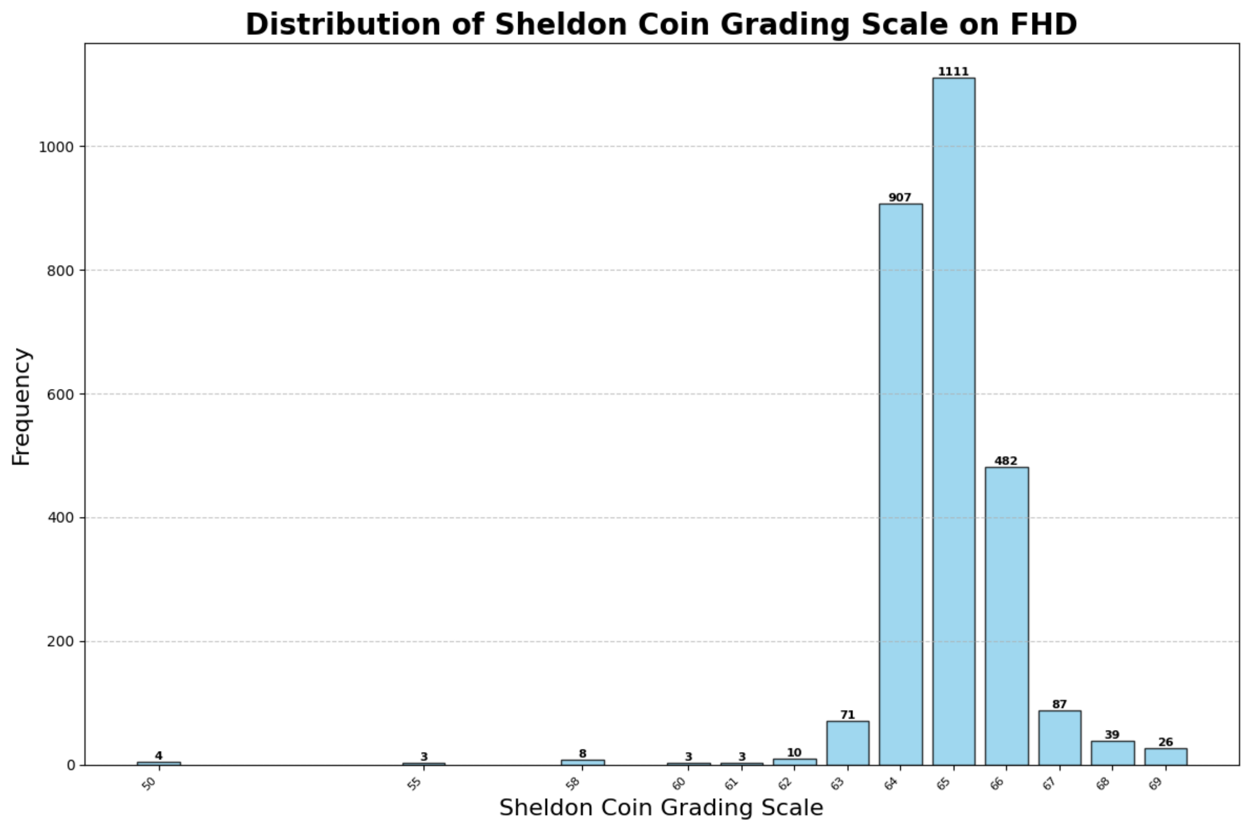


Figure 6.1: Distribution of Sheldon Coin Grading Scale on Franklin Half Dollar

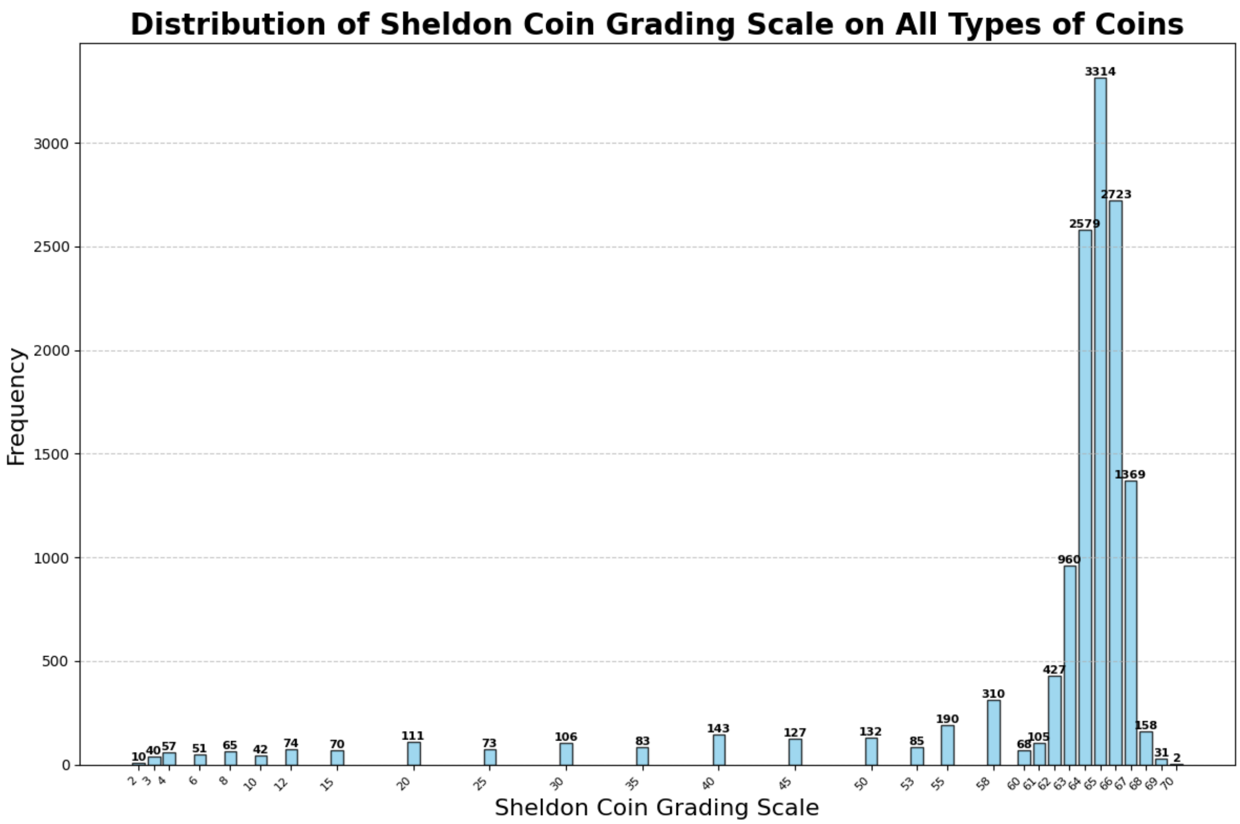


Figure 6.2: Distribution of Sheldon Coin Grading Scale on All 5 Types of Coins

synthetic samples, thereby avoiding biases towards the majority classes. The dataset is split into 80/20 training and testing sets, with 80% used for training and 20% reserved for testing, ensuring the model is trained and validated effectively. During training, 20% of the training data is further split into a validation set to monitor model performance and fine-tune hyperparameters. After each epoch, the validation loss is calculated, and if the validation loss does not improve over 10 consecutive epochs, EarlyStopping terminates the training process to prevent overfitting. A fixed random seed ensures consistent train-test splits across different runs, maintaining the reproducibility of results.

6.4 MLP Model Architecture For Sheldon Grading Scale

To develop an automated coin grading system, we implemented both MLP and Random Forest model with regression and multi-class classification mode on Fanklin Half Dollar(FHD) individually and all five types of coins to predict the Sheldon Coin Grading Scale. We then compared the performance of these models to choose the best one. The purpose of implementing a model on FHD is to evaluate how well the extracted features align with a specific type of coin. The architecture for these models is detailed below:

6.4.1 MLP Model On Franklin Half Dollar

This regression model begins with an input layer consisting of a Dense layer with 128 units, using ReLU activation and L2 regularization with a factor of 0.001. This is followed by a Batch Normalization layer and a Dropout layer with a rate of 0.3. The next layer is a Dense layer with 64 units, also using ReLU activation and L2 regularization with a factor of 0.001, followed by another Batch Normalization layer and a Dropout layer with a rate of 0.2.

- **For regression tasks**: The final layer is a Dense layer with 1 unit, utilizing a custom activation function for regression.
- **For multi-class classificaition tasks**: The final layer is a Dense layer with a number of units equal to the unique classes in the target variable, using softmax activation. The model is compiled with the Adam optimizer, categorical cross-entropy loss, and accuracy as the evaluation metric.

6.4.2 MLP Regression Model On Five Types of Coins

This model begins with an input layer consisting of a Dense layer with 256 units, using L1 and L2 regularization with factors of 0.01 each. This layer is followed by a LeakyReLU activation function with an alpha of 0.1, a Batch Normalization layer, and a Dropout layer with a rate of 0.5. The next layer is a Dense layer with 128 units, also using L1 and L2 regularization with factors of 0.01 each, followed by a LeakyReLU activation function with an alpha of 0.1, a Batch Normalization layer, and a Dropout layer with a rate of 0.4. This is followed by another Dense layer with 64 units, using the same regularization and activation functions, a Batch Normalization layer, and a Dropout layer with a rate of 0.3. The final layer is a Dense layer with 1 unit, followed by a Lambda layer applying a custom activation function. The model is compiled with the Adam optimizer, a learning rate of 0.001, mean squared error as the loss function, and mean absolute error as the evaluation metric.

6.4.3 MLP Multi-Class Model On Five Types of Coins

This model starts with an input layer consisting of a Dense layer with 128 units, using ReLU activation and L2 regularization with a factor of 0.001. This is followed by a Batch Normalization layer and a Dropout layer with a rate of 0.4. The next layer is a Dense layer with 64 units, using ReLU activation, followed by another Batch Normalization layer and a Dropout layer with a rate of 0.3. The final layer is a Dense layer with a number of units equal to the unique classes in the target variable, using softmax activation. The model is compiled with the Adam optimizer, categorical cross-entropy loss, and accuracy as the evaluation metric.

6.5 MLP Model for Degree of Toning

We also used the MLP model for predicting the degree of toning. The model architecture is similar than the previous ones. The model begins with an input layer consisting of a Dense layer with 64 units, using ReLU activation and L2 regularization with a factor of 0.01. This is followed by a Batch Normalization layer and a Dropout layer with a rate of 0.5. The next layer is a Dense layer with 32 units, also using ReLU activation and L2 regularization with a factor of 0.01, followed by another Batch Normalization layer and a Dropout layer with a rate of 0.5.

- **For regression tasks**: The final layer is a Dense layer with 1 unit, utilizing a linear activation function to predict a continuous output.

- **For multi-class classificaition tasks**: The final layer is a Dense layer with a number of units equal to the number of unique classes in the target variable, utilizing a softmax activation function for multi-class classification.

6.6 Random Forest

The Random Forest (RF) model is another powerful algorithm that was explored for the task of coin grading within this study. Random Forest is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputting either the mode of the classes (classification) or the mean prediction (regression) of the individual trees [29]. This ensemble approach helps mitigate the risk of overfitting and generally improves the model's performance by averaging out the predictions of multiple decision trees.

To ensure that the Random Forest model was optimally configured for the coin grading task, an extensive Grid Search was conducted. Grid Search is a systematic approach to hyperparameter tuning that involves exhaustively evaluating all possible combinations of a predefined set of hyperparameters [30]. For the Random Forest model, the hyperparameters explored included: number of trees, maximum depth of trees, minimum samples split, minimum samples per leaf, and maximum features.

6.7 Implementation

I conducted all model training and evaluation using Google Colab with a T4 GPU, and I wrote and executed the code within the Google Colab Python notebook environment. For the best model predicting the degree of toning on Franklin Half Dollar coins, the training time was 21.6 seconds, and the inference time was 0.66 seconds. For the best model predicting the Sheldon Grading Scale on Franklin Half Dollar coins, the training took 19.8 seconds, with an inference time of 0.03 seconds. When predicting the Sheldon Grading Scale across all five coin types, the training took 166.7 seconds, and inference took 1.1 seconds.

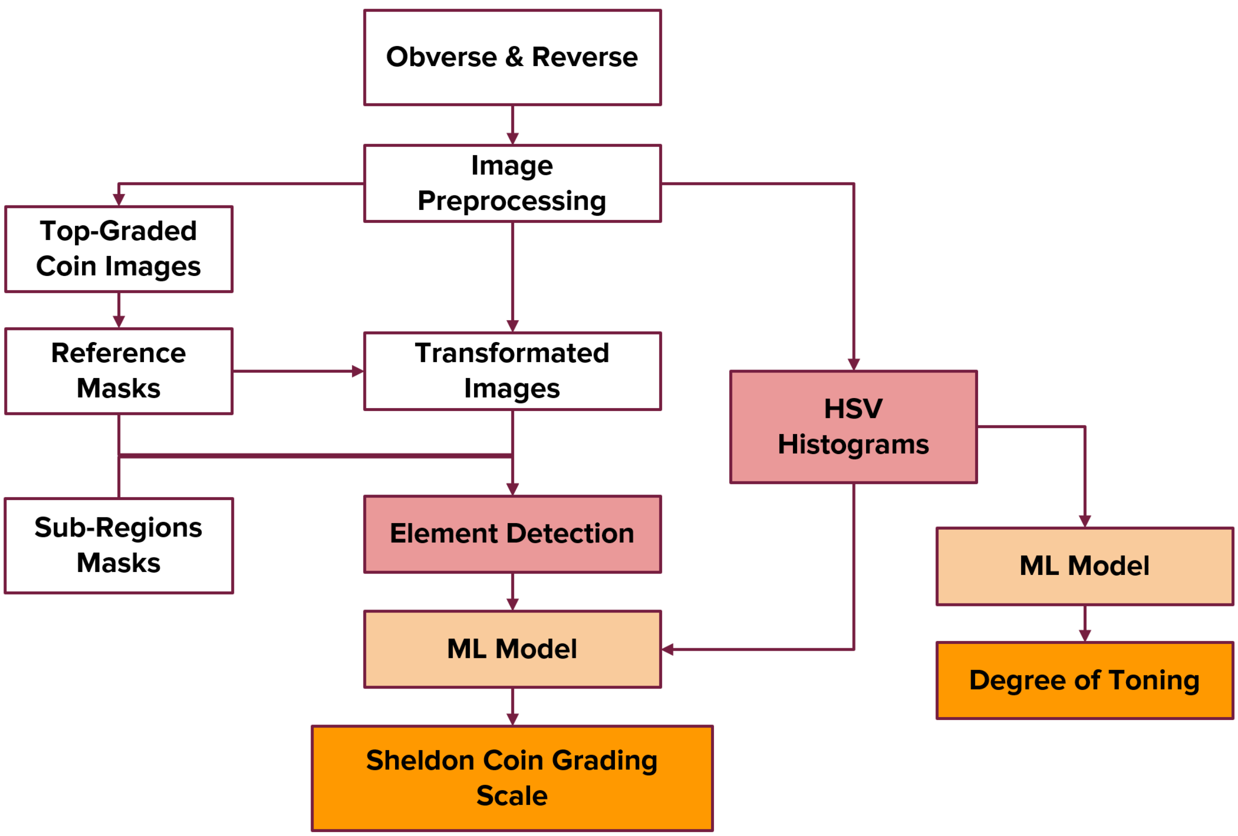


Figure 6.3: Feature Extraction Framework of the Automated Coin Grading System

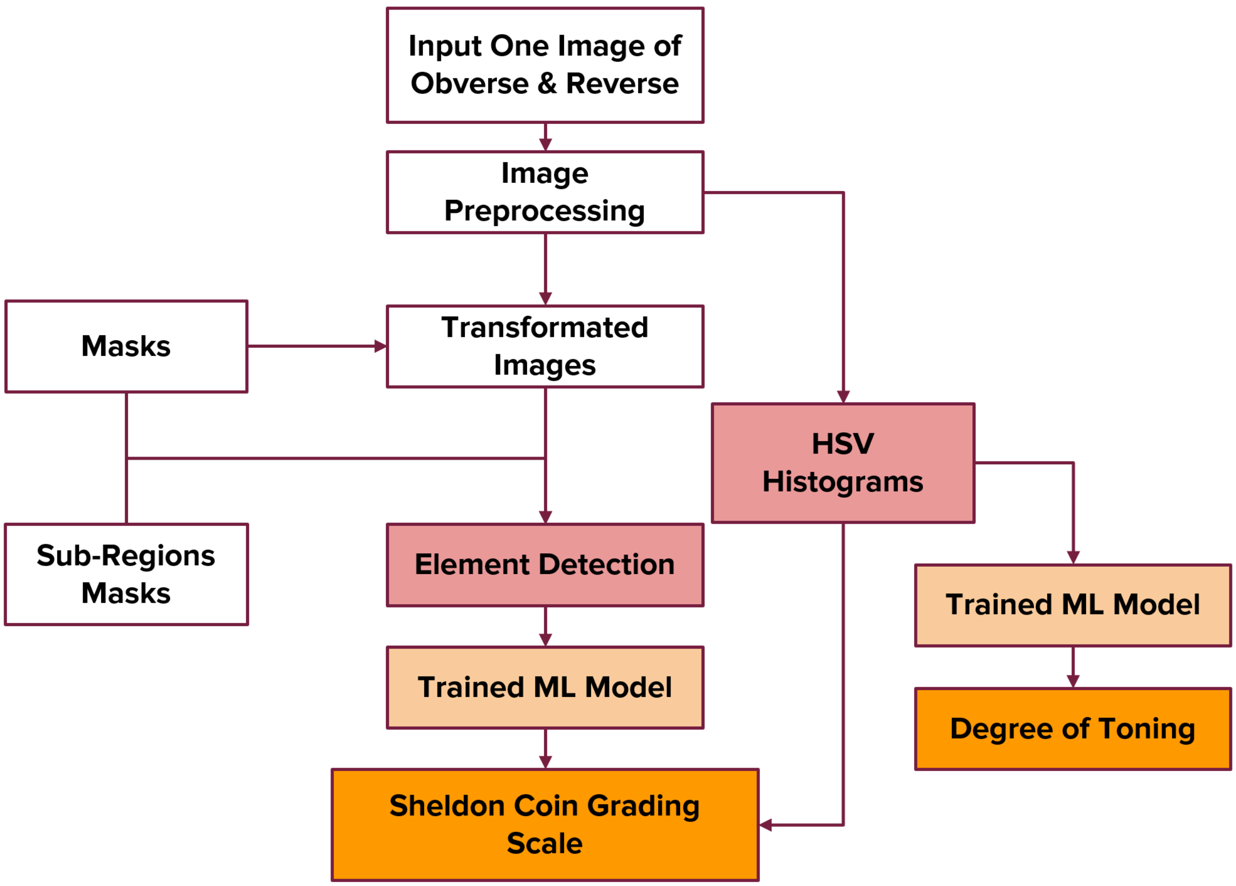


Figure 6.4: Framework for Applying a Coin to the Trained Model

Chapter 7

Experimental Result & Discussion

7.1 Degree of Toning Prediction on Franklin Half Dollar (FHD)

Model	**Degree of Toning Model Performance**			
MLP	**Regression**		**Multi-Class**	
	MAE: 0.26 MSE: 0.41 R^2: 0.69	Tolerance of 1 : 96.4% Tolerance of 2 : 99.6%	68.6%	Tolerance of 1 : 99.6% Tolerance of 2 : 99.6%

Table 7.1: Degree of Toning Model Performance for MLP

The MLP model was evaluated using both regression and multi-class classification approaches to predict the degree of toning (on a scale of 1 to 5). Although the regression model achieved low MAE and MSE, the multi-class classification model is even better since it achieved 99.6% accuracy within a tolerance of 1 (highlighted in red in Table 7.1). It demonstrated remarkable performance when slight deviations were allowed.

And its loss per epoch graph (Figure 7.1) shows a steady decline in both training and validation loss, indicating that the model is learning effectively without significant overfitting. The convergence of the validation loss towards the training loss suggests a good generalization capability, supporting the robustness of the model across unseen data.

The model exhibits a strong diagonal, indicating high accuracy in predicting the correct toning levels (Fig. 7.3), particularly for the majority classes (1, 2, and 3). However, there is some degree of misclassification, particularly for classes 4 and 5, due to the limited support (i.e., fewer training examples) for these categories.

Given the nature of the problem—where predicting the exact degree of toning is important but slight deviations can be tolerated—the multi-class classification model is preferred. The high tolerance accuracy (99.6%) indicates that the model reliably predicts the toning level within a small margin of error, making it more robust and practical for real-world applications.

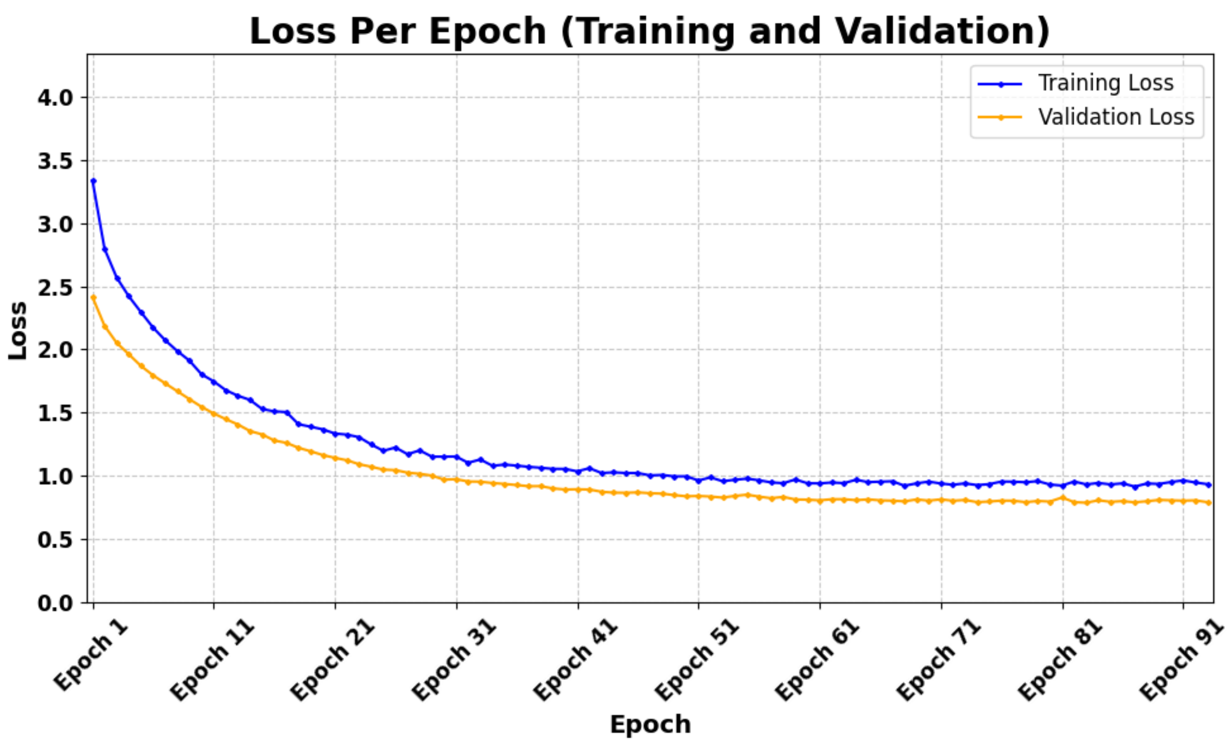


Figure 7.1: Loss Curve: MLP Multi-Class Classification Model on FHD

Degree of Toning	Precision	Recall	F1-Score	Support
1	0.73	0.74	0.73	180
2	0.68	0.68	0.68	216
3	0.68	0.61	0.64	114
4	0.50	0.62	0.56	40
5	0.50	1.00	0.67	1

Table 7.2: MLP Multi-Class Classification Model Performance metrics on FHD

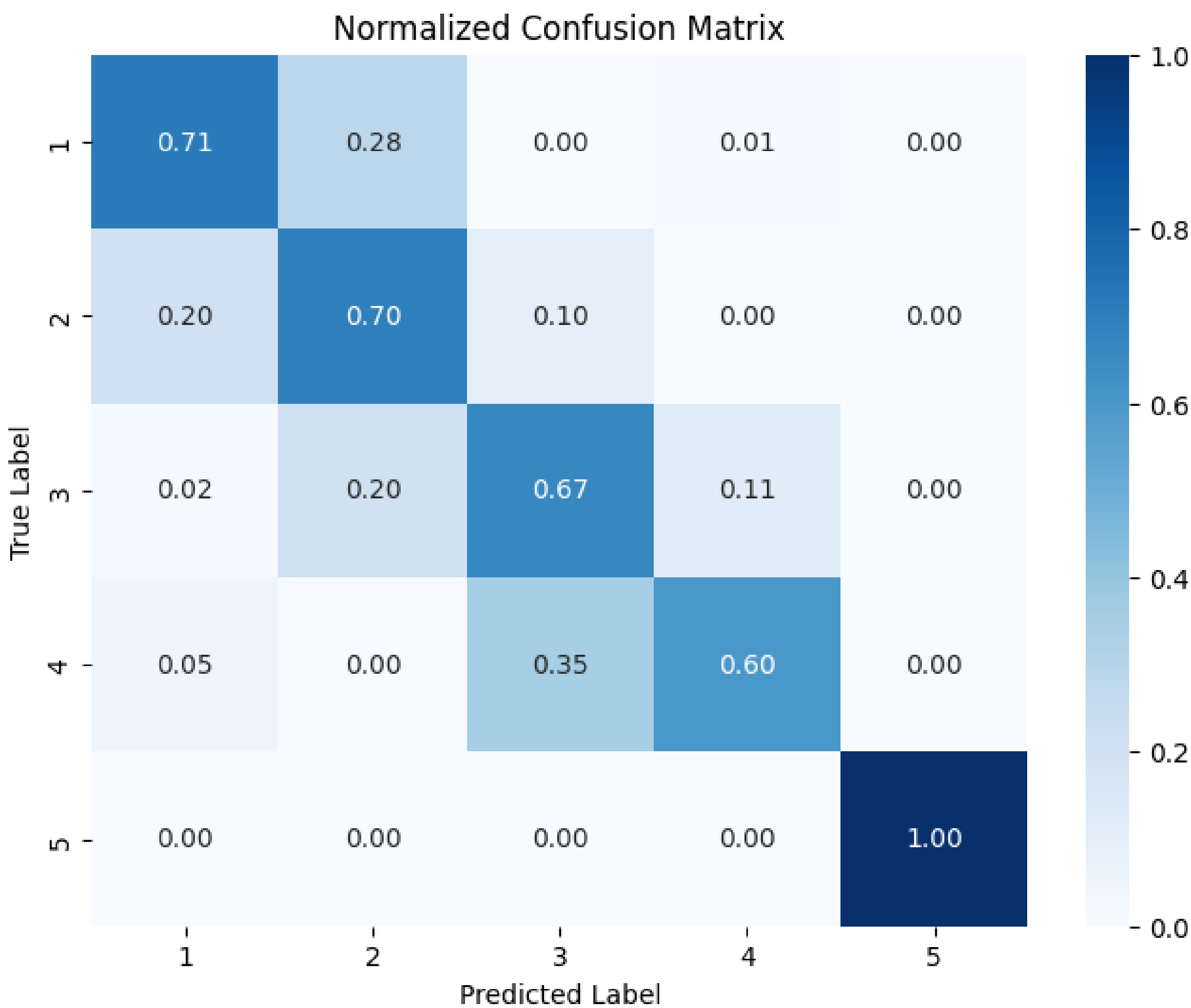


Figure 7.2: MLP Multi-Class Classification Model Normalized Confusion Metrics on FHD

7.2 Sheldon Coin Grading Scale on FHD

Model Performance on Franklin Half Dollar (Partial SMOTE)				
Model	**Regression**		**Multi-Class**	
MLP	MAE: 0.672 MSE: 1.388 R^2: 0.1885	Tolerance of 1 : 76.2% Tolerance of 2 : 92.1% Tolerance of 3 : 96.8%	50.64%	Tolerance of 1 : 91.8% Tolerance of 2 : 98.6% Tolerance of 3 : 99.2%
Random Forest	MAE: 0.568 MSE: 0.986 R^2: 0.243	Tolerance of 1 : 95.3% Tolerance of 2 : 99.3% Tolerance of 3 : 99.3%	49.73%	Tolerance of 1 : 94.7% Tolerance of 2 : 99.1% Tolerance of 3 : 99.3%

Table 7.3: Sheldon Coin Grading Scale Prediction on Franklin Half Dollar (Partial SMOTE)

Model Performance on Franklin Half Dollar (No SMOTE)				
Model	**Regression**		**Multi-Class**	
MLP	MAE: 0.655 MSE: 1.20 R^2: 0.21	Tolerance of 1 : 79.3% Tolerance of 2 : 94.3% Tolerance of 3 : 97.8%	48.28%	Tolerance of 1 : 93.7% Tolerance of 2 : 99.3% Tolerance of 3 : 99.3%
Random Forest	MAE: 0.579 MSE: 0.956 R^2: 0.265	Tolerance of 1 : 95.1% Tolerance of 2 : 99.3% Tolerance of 3 : 99.5%	47.91%	Tolerance of 1 : 94.9% Tolerance of 2 : 99.3% Tolerance of 3 : 99.5%

Table 7.4: Sheldon Coin Grading Scale Prediction on Franklin Half Dollar (No SMOTE)

- **MLP vs. Random Forest:** Random Forest consistently outperforms MLP across all regression metrics (MAE, MSE, R^2) and tolerance levels, both with and without SMOTE. This indicates that Random Forest is better at predicting the exact Sheldon Coin Grading Scale values on Franklin Half Dollar.

- **Regression vs. Multi-Class Classification:** The Sheldon Coin Grading Scale is a continuous range from 1 to 70, but when treated as a multi-class classification task (categorizing into discrete grades), regression shows slightly better performance in terms of tolerance of accuracy than multi-class classification.

- **SMOTE vs. Non-SMOTE:** The application of Partial SMOTE marginally decreases classification accuracy and tends to slightly degrade tolerance accuracy. Therefore, non-SMOTE is better for this model.

Based on the aboves tables, the Random Forest model in regression mode without SMOTE (highlighted in red in Table 7.4) appears to be the best option for predicting the Sheldon Grading Scale of Franklin Half Dollars (FHD). The model was used specifically on Franklin Half Dollars because there is a large dataset available for FHD in Mint State (Fig. 6.1). This range represents the most valuable coins for auction.

In Fig. 7.3, the Random Forest model in regression mode (without SMOTE) generally performs well, with most predictions being accurate or within a narrow margin of error. The x-axis represents the difference between the predicted grade and the actual grade (Difference = Predicted Grade - Actual Sheldon Coin Grading Scale), while the y-axis represents the predicted grade. The model exhibits a slight tendency to over-predicted, particularly at lower grades, but overall, the accuracy is high, with most errors being small.

The large clusters of data points around 0 and 1 difference indicate that the model performs exceptionally well for most predictions. The majority of examples are tightly clustered around these low difference values, which demonstrates that the model is highly accurate in predicting values close to the true outcomes. However, in an ideal scenario, a perfect model would have nearly all data points concentrated within a 0 to 3 difference range. While our model achieves this for a large portion of the data, a few examples fall outside this optimal range, indicating potential room for further improvement. Addressing these outliers could further refine the model's performance.

Consequently, implementing the best model on FHD has yielded highly accurate results, with an accuracy of 95.1% within a tolerance of 1 grade, demonstrating that the features we extracted align well with the Sheldon Grading Scale, further validating the model's effectiveness.

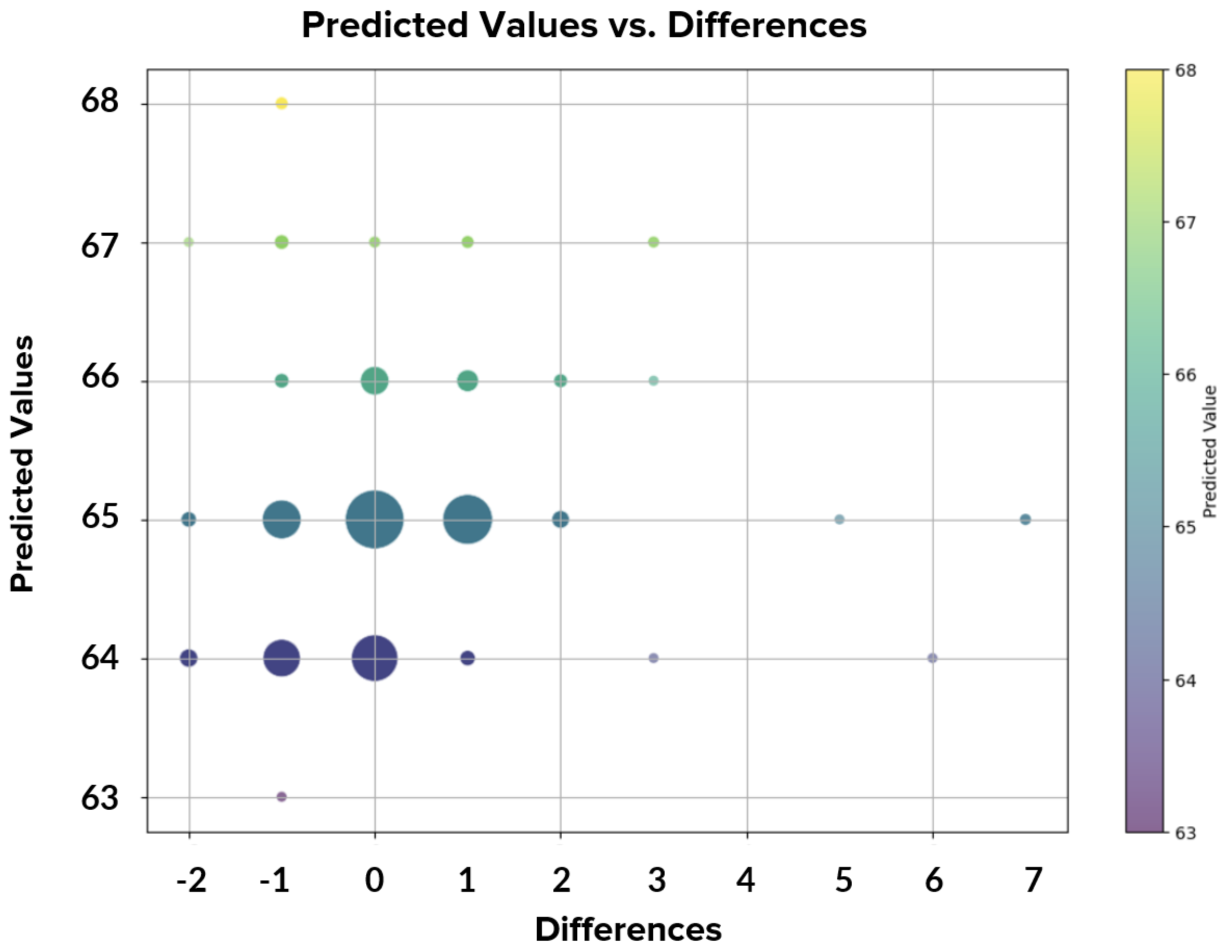


Figure 7.3: Random Forest Regression Model Without SMOTE on Franklin Half Dollar: Predicted Grade vs. Differences (Difference = Predicted Grade - Actual Sheldon Coin Grading Scale; the size of the diameters indicates the number of coins)

7.3 Sheldon Coin Grading Scale on All Five Types of Coins

Performance on All Types of Coins (Partial SMOTE)				
Model	**Regression**		**Multi-Class**	
MLP	MAE: 2.58 MSE: 19.11 R^2: 0.86	Tolerance of 1 : 45.3% Tolerance of 2 : 67.5% Tolerance of 3 : 81.0% Tolerance of 4 : 87.9%	27.66%	Tolerance of 1 : 63.9% Tolerance of 2 : 77.2% Tolerance of 3 : 82.6% Tolerance of 4 : 85.6%
Random Forest	MAE: 3.07 MSE: 29.2 R^2: 0.78	Tolerance of 1 : 49.3% Tolerance of 2 : 65.6% Tolerance of 3 : 74.8% Tolerance of 4 : 80.4%	34.00%	Tolerance of 1 : 70.9% Tolerance of 2 : 82.3% Tolerance of 3 : 86.3% Tolerance of 4 : 87.8%

Table 7.5: Sheldon Coin Grading Scale Prediction on All Types of Coins (Partial SMOTE)

Performance on All 5 Types of Coins (No SMOTE)				
Model	**Regression**		**Multi-Class**	
MLP	MAE: 1.90 MSE: 13.74 R^2: 0.9	Tolerance of 1 : 65.0% Tolerance of 2 : 81.0% Tolerance of 3 : 88.3% Tolerance of 4 : 91.3%	34.85%	Tolerance of 1 : 66.7% Tolerance of 2 : 79.9% Tolerance of 3 : 86.8% Tolerance of 4 : 88.7%
Random Forest	MAE: 2.81 MSE: 29.24 R^2: 0.78	Tolerance of 1 : 48.4% Tolerance of 2 : 68.8% Tolerance of 3 : 78.3% Tolerance of 4 : 83.5%	36.00%	Tolerance of 1 : 72.0% Tolerance of 2 : 83.4% Tolerance of 3 : 87.0% Tolerance of 4 : 88.0%

Table 7.6: Sheldon Coin Grading Scale Prediction on All Types of Coins (No SMOTE)

The below tables (Table 7.5 and 7.6) presents the performance of the MLP and random forest model applied to predict the Sheldon Coin Grading Scale across five different types of coins (Franklin Half Dollar, Mercury Dime, Buffalo Nickel, Washington Quarter, and Morgan Silver Dollar) with and without SMOTE. From the above tables, MLP consistently outperforms Random Forest in regression tasks particularly without SMOTE on all 5 types of coins. The MLP regression without SMOTE performance (highlighted in red in Table 7.6), with a Mean Absolute Error (MAE) of 1.90, is particularly noteworthy given that it predicts a grading range from 1 to 70 across five different types of coins. The high R^2 value of 0.9 further supports the model's strong predictive capabilities, making it a reliable tool for coin grading.

In Fig. 7.4, both the training and validation losses consistently decrease over the epochs, with the loss stabilizing at a low value, indicating that the model is effectively learning and generalizing well without overfitting. Thus, MLP regression model without SMOTE is the best model for predicting the Sheldon Coin Grading Scale of five types of coins.

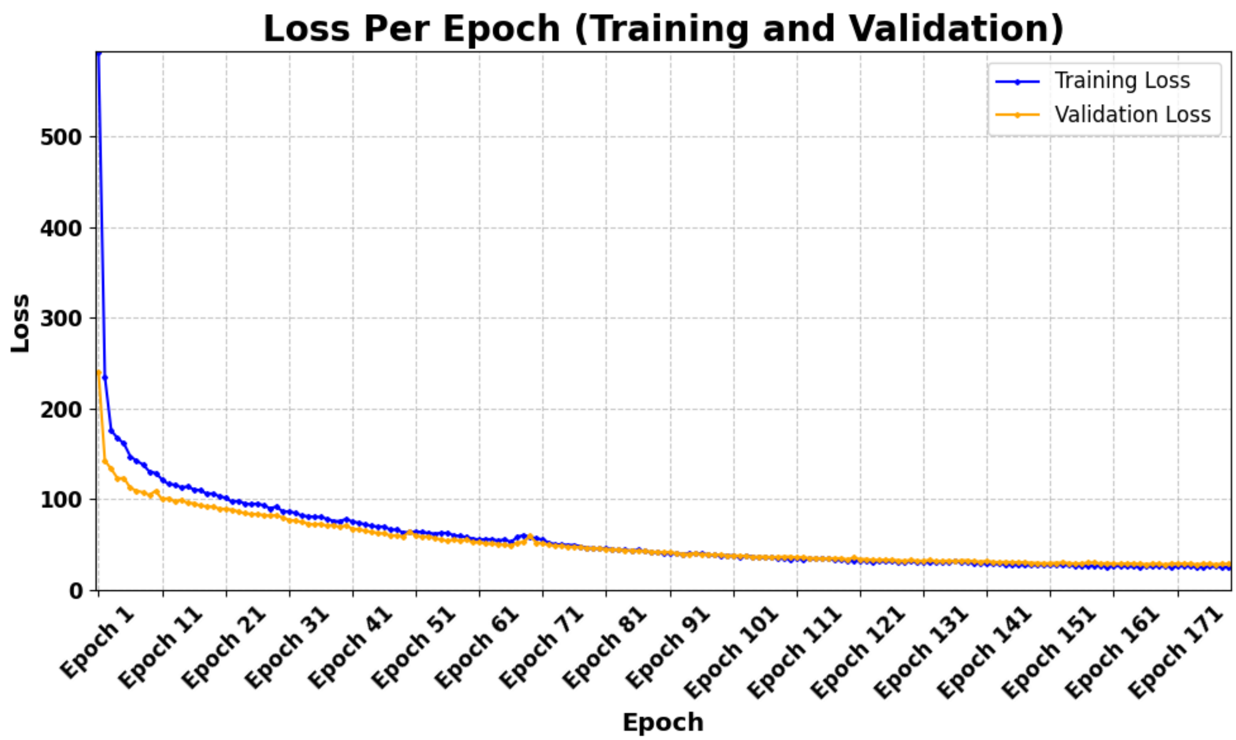


Figure 7.4: Loss Curve: MLP Regression model without SMOTE on all five types of coins

In Fig. 7.5, the figure presents the performance of the MLP regression model applied to predict the Sheldon Coin Grading Scale across five different types of coins. The x-axis and the y-axis are similar to the previous graph. The degree of toning is indicated by the color scale, with lighter colors corresponding to higher degrees of toning. The varying shapes of the dots indicate different types of coins.

A noticeable trend in the graph is that coins with larger differences (both under-predictions and over-predictions) often exhibit a higher degree of toning. These are represented by lighter colors on the graph, particularly in the negative difference region (under-prediction). This suggests that toning might be a significant factor influencing the model's prediction accuracy.

The plot shows that the Washington Quarter and Buffalo Nickel tend to have more significant prediction errors compared to other coin types. The larger errors associated with these coin types might be due to inherent differences in how their features are captured and interpreted by the model. For example, specific characteristics of these two types of coins, such as hair features, might be too complicated to capture and account for by the model. These outliers suggest that in some cases, the model fails to accurately assess the grade, likely due to a combination of high toning and coin type-related challenges.

However, given that Mint State coins (grades 60 to 70) are a primary focus for collectors and the coin grading industry, the model's ability to accurately predict these grades is of paramount importance. The accurate clustering of predictions in this range indicates that the MLP model is reliable in assessing the quality of these high-grade coins. Its strong performance in predicting Mint State coins (grades 60-70) is a significant success.

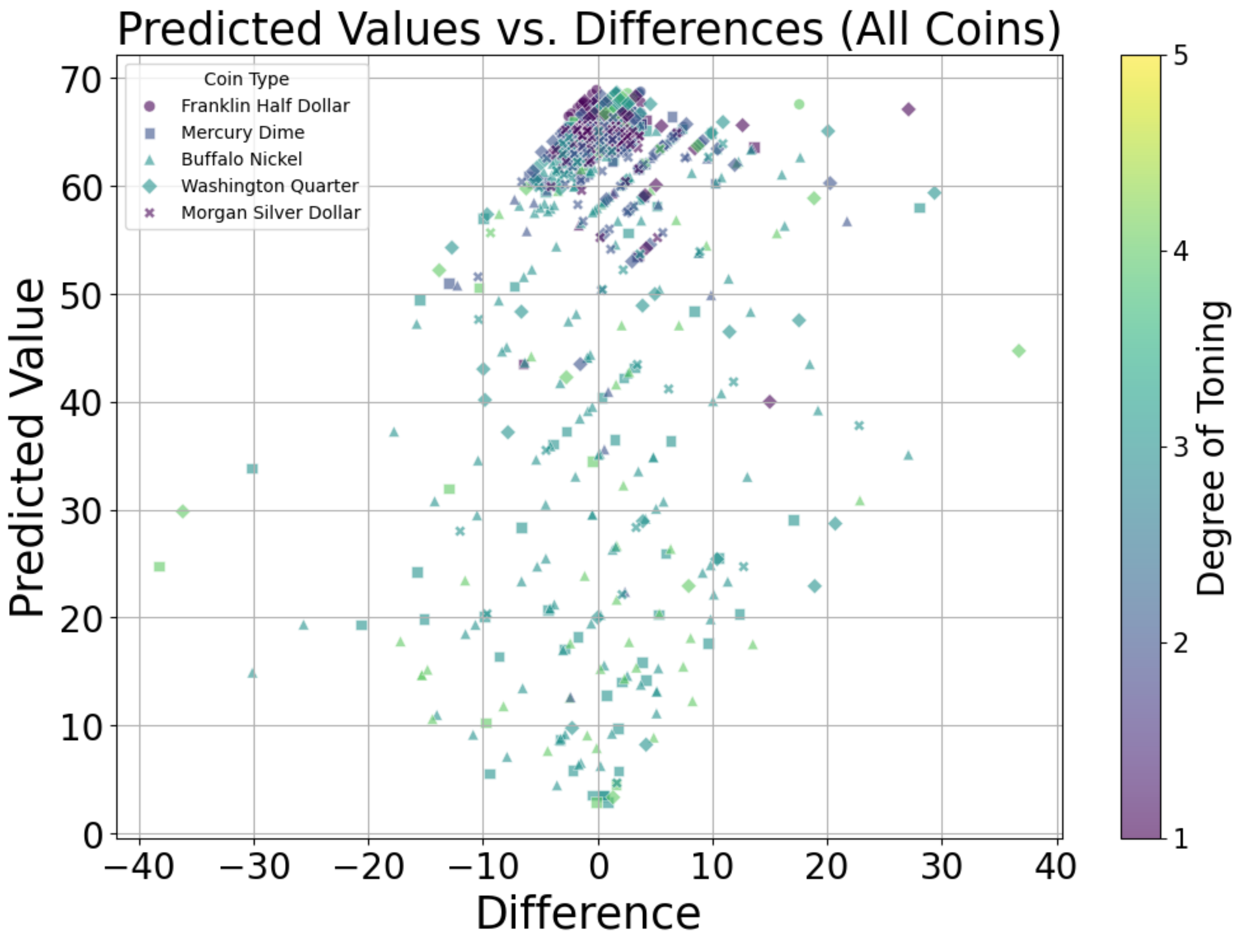


Figure 7.5: MLP in regression mode without SMOTE: Predict Grade vs. Differences (Differences = Predicted Grade - Actual Sheldon Coin Grading Scale); The diagonal clustering in the plot occurs because the predicted values are real numbers (continuous), while the actual values are integers (discrete). When computing the differences (predicted - actual), the results align in such a way that they cluster diagonally based on the integer intervals of the actual values.

Chapter 8

Conclusion & Future Work

8.1 Conclusion

This thesis explores the development of an automated coin grading system designed to reduce the subjectivity and inconsistency of traditional grading methods. By using advanced image processing techniques and machine learning models, this work makes progress in automating the grading process according to the Sheldon Coin Grading Scale. The use of synthetic reference masks and sub-region analysis allowed for accurate extraction of important features from coin images. These methods enabled the system to focus on key aspects of a coin's surface and relief, which are critical for determining its grade. Dividing the coins into 18 distinct regions was particularly effective in capturing the details that affect grading, such as wear patterns and strike clarity.

Additionally, color histograms were employed as part of the feature extraction process to capture the distribution of color tones. This approach provided valuable information about the coin's toning, which is a significant factor in grading. The analysis of color histograms allowed the model to better differentiate between various degrees of toning, serving as another feature for the Sheldon Coin Grading Scale.

The successful application of the MLP model in predicting grades across various coin types demonstrates the potential of automated grading systems to improve and standardize the coin grading process. Moreover, this grading system proves that it could work for any type of coin, with particular effectiveness for Mint State coins. With further refinement and expansion, the proposed system could revolutionize the way coins are graded, offering a standardized, reliable, and efficient alternative to traditional methods.

8.2 Future Work

While the results of this thesis are promising, several areas for future research and development could further enhance the accuracy and applicability of the automated grading system:

1. **Image Artifacts Caused by Coin Holders:** One of the challenges encountered during the image processing phase was the image artifacts introduced by the prongs that hold the coin in place. When attempting to remove the prongs from the images, 5 out of 100 images were affected by incorrect scaling. This misalignment complicated the task of matching the masks accurately and led to erroneous grading results. While this issue occurred infrequently, it still impacted the reliability of the system. For future work, curating the training images to exclude affected samples could help mitigate this problem. Additionally, using a consistent holder design for all coins during imaging could further reduce variability and ensure that the masks align correctly, leading to more accurate grading predictions.

2. **Incorporating Grading Service Bias:** Another aspect to consider in future work is the bias introduced by different grading services, such as PCGS (Professional Coin Grading Service) and NGC (Numismatic Guaranty Corporation). Currently, the majority of coins in the dataset are graded by these two services, and their respective grading standards can influence the perceived value of the coins. However, the dataset

does not currently separate coins based on their grading service. For future work, introducing an additional feature to distinguish between grading services could allow the model to account for these biases, thereby improving predictive accuracy and better aligning its outputs with market expectations.

3. **Exploring the Challenge of Measuring Luster:** Luster is a fundamental component of the Sheldon Coin Grading Scale, contributing significantly to a coin's overall grade. However, capturing luster through 2D images presents a considerable challenge. Graders typically rotate the coin under light to observe how the surface reflects, a dynamic process that is difficult to replicate in a static image. Developing methods to measure luster accurately from 2D images or exploring the use of video or 3D imaging techniques could be crucial steps forward in enhancing the model's ability to assess this critical aspect of coin grading.

4. **Enhancing Toning and Color Analysis:** The degree of toning was identified as a critical factor affecting the accuracy of the grading predictions. Coins with higher degrees of toning—whether positive or negative—tended to result in larger prediction errors. Future work should focus on refining the color histogram analysis and incorporating additional features that better capture the subtleties of toning. Improved differentiation between positive and negative toning effects could lead to more precise grade predictions, particularly for coins where toning significantly influences their perceived quality.

By addressing these challenges, future iterations of the automated grading system could achieve even greater accuracy and reliability, making it a more robust tool for numismatists and collectors alike. The integration of consistent imaging practices, acknowledgment of grading service biases, advanced luster measurement techniques, and refined toning analysis will collectively push the boundaries of what automated coin grading systems can accomplish.

Appendices

Appendix A

Coin Images and Grading Examples

A.1 Representative Images of Each Coin Type

A representative coin from each type (Franklin Half Dollar, Mercury Dime, Buffalo Nickel, Washington Quarter, and Morgan Silver Dollar) is displayed, highlighting both the obverse and reverse sides.

(a) Obverse (b) Reverse

Figure A.1: Franklin Half Dollar

(a) Obverse (b) Reverse

Figure A.2: Mercury Dime

(a) Obverse (b) Reverse

Figure A.3: Buffalo Nickel

(a) Obverse

(b) Reverse

Figure A.4: Washington Quarter

(a) Obverse

(b) Reverse

Figure A.5: Morgan Silver Dollar

A.2 Grading Prediction Examples

A.2.1 Model on Franklin Half Dollar Only

The attached graph illustrates the model's performance for Franklin Half Dollars, with specific examples marked and highlighted with red circles. The obverse and reverse of these example coins are showcased below.

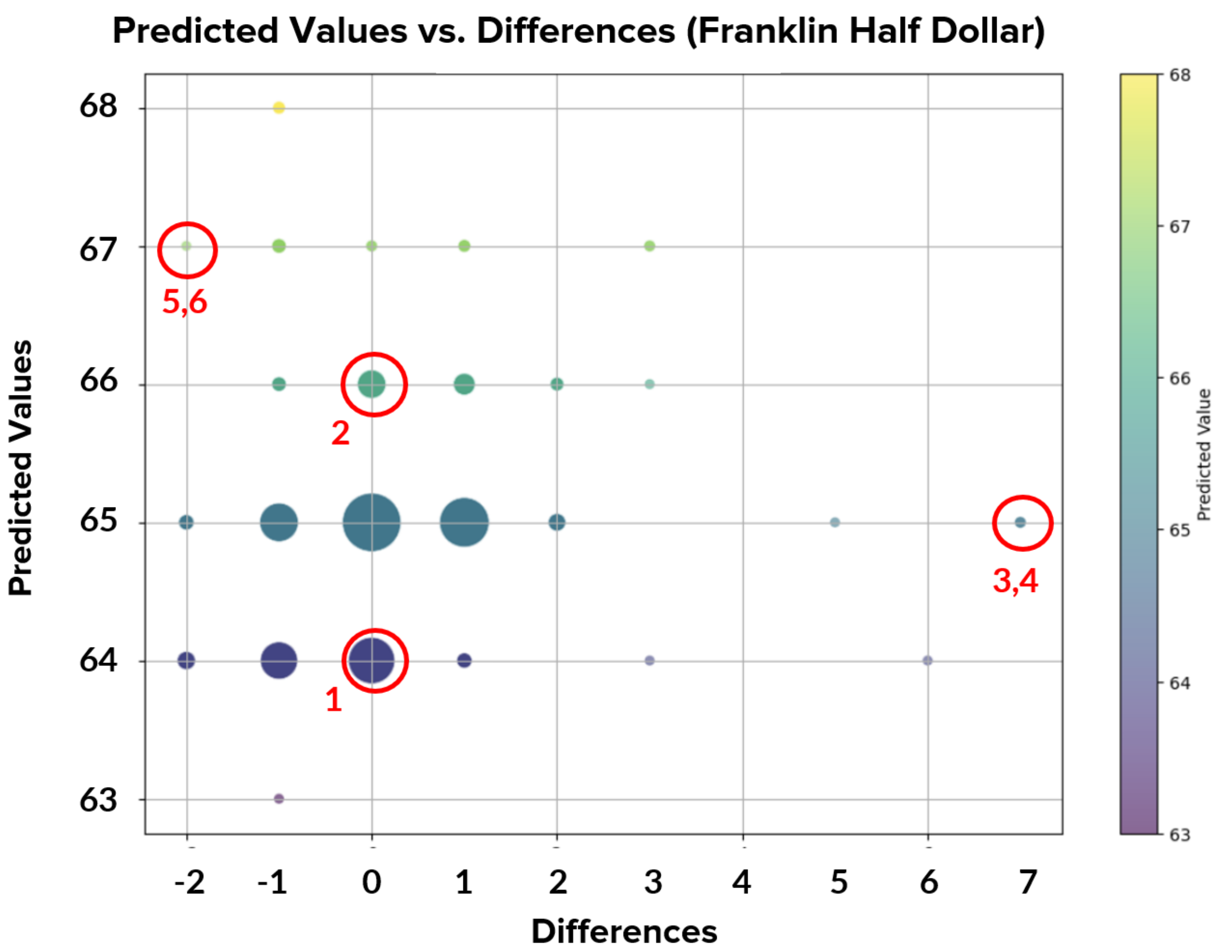


Figure A.6: Example Coins on the graph

Below are two coins (Coin1 & Coin2) where the grading was predicted perfectly.

(a) Obverse (b) Reverse

Figure A.7: Coin1 - Franklin Half Dollar with PCGS Grade: 64 and Predicted Grade: 64 (Degree 2 Toning)

(a) Obverse (b) Reverse

Figure A.8: Coin2 - Franklin Half Dollar with PCGS Grade: 66 and Predicted Grade: 66 (Degree 3 Toning)

Below are two coins (Coin3 & Coin4) where the model's predicted grade was significantly higher than the actual grade.

(a) Obverse

(b) Reverse

Figure A.9: Coin3 - Franklin Half Dollar with NGC Grade: 58 and Predicted Grade: 65 (Degree 1 Toning)

(a) Obverse

(b) Reverse

Figure A.10: Coin4 - Franklin Half Dollar with PCGS Grade: 58 and Predicted Grade: 65 (Degree 1 Toning)

Below are two coins (Coin5 & Coin6) where the model's predicted grade was lower than the actual grade.

(a) Obverse (b) Reverse

Figure A.11: Coin5 - Franklin Half Dollar with PCGS Grade: 67 and Predicted Grade: 65 (Degree 1 Toning)

(a) Obverse (b) Reverse

Figure A.12: Coin6 - Franklin Half Dollar with PCGS Grade: 67 and Predicted Grade: 65 (Degree 3 Toning)

A.2.2 Model on All Coins

The attached graph illustrates the model's performance for all coins, with specific examples marked and highlighted with red circles. The obverse and reverse of these example coins are showcased below.

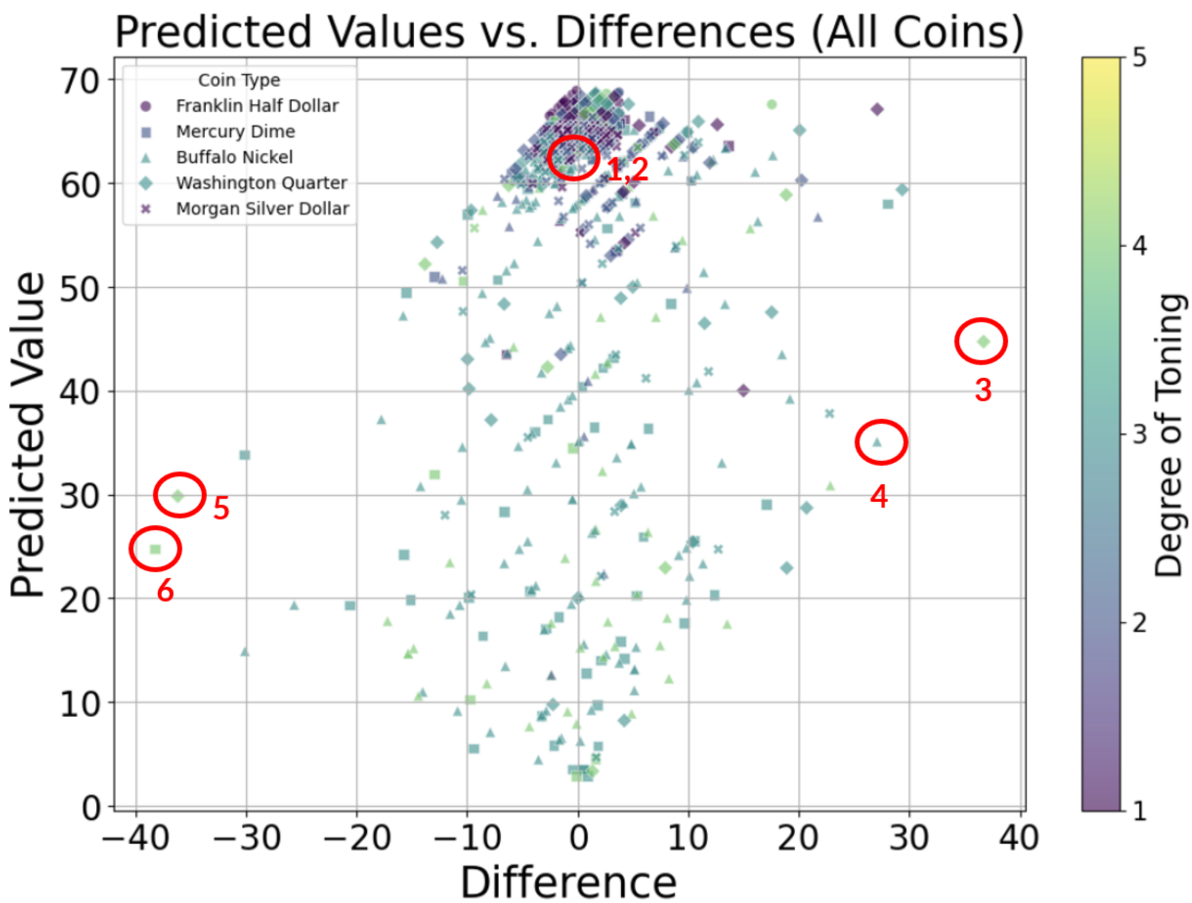


Figure A.13: Example Coins on the graph

Below are two coins (Coin1 & Coin2) where the grading was predicted perfectly.

(a) Obverse

(b) Reverse

Figure A.14: Coin1 - Buffalo Nickel with PCGS Grade: 62 and Predicted Grade: 62 (Degree 3 Toning)

(a) Obverse

(b) Reverse

Figure A.15: Coin2 - Morgan Silver Dollar with PCGS Grade: 64 and Predicted Grade: 64 (Degree 3 Toning)

Below are two coins (Coin3 & Coin4) where the model's predicted grade was significantly higher than the actual grade.

(a) Obverse (b) Reverse

Figure A.16: Coin3 - Washington Quarter with PCGS Grade: 8 and Predicted Grade: 45 (Degree 4 Toning)

(a) Obverse (b) Reverse

Figure A.17: Coin4 - Buffalo Nickel with PCGS Grade: 8 and Predicted Grade: 35 (Degree 3 Toning)

Below are two coins (Coin5 & Coin6) where the model's predicted grade was significantly lower than the actual grade.

(a) Obverse (b) Reverse

Figure A.18: Coin5 - Washington Quarter with NGC Grade: 66 and Predicted Grade: 30 (Degree 4 Toning)

(a) Obverse (b) Reverse

Figure A.19: Coin6 - Mercury Dime with PCGS Grade: 63 and Predicted Grade: 25 (Degree 4 Toning)

Bibliography

[1] U.S. Mint, "Anatomy of a coin." https://www.usmint.gov/learn/collecting-basics/anatomy-of-a-coin, 2024. Accessed: 2024-05-29.

[2] U.S. Asset Advisors, "Sheldon grading scale." https://usassetadvisors.com/sheldon-grading-scale/, 2024. Accessed: 2024-05-29.

[3] M. R. Atighehchian, "Coin wear estimation and automatic coin grading." https://spectrum.library.concordia.ca/982454/1/Atighehchian_McompSc_S2017.pdf, 2017. Published Master's Thesis at Concordia University.

[4] R. Bassett, P. Gallivan, X. Gao, E. Heinen, and A. Sakalaspur, "Development of an automated coin grader: A progress report," in *Proceedings of the 8th Annual Mid-Atlantic Student Workshop on Programming Languages and Systems*, 2002.

[5] I. Schlag and O. Arandjelović, "Ancient roman coin recognition in the wild using deep learning-based recognition of artistically depicted face profiles," in *Proceedings of the ICCV Workshop*, pp. 2898–2906, 2017.

[6] X. Pan and L. Tougne, "Image analysis and deep learning for aiding professional coin grading," in *2018 International Conference on Image and Video Processing, and Artificial Intelligence* (R. Su, ed.), vol. 10836, p. 1083605, International Society for Optics and Photonics, SPIE, 2018.

[7] S. Gakhar, "Local image patterns for counterfeit coin detection and automatic coin grading," master's thesis, Concordia University, 2020.

[8] J. McMorrow-Hernandez, "4 fundamental components of mint state grading." https://www.pcgs.com/news/4-fundamental-components-of-mint-state-grading, 2022. Accessed: 2024-05-29.

[9] D. L. R. Coins, "Inventory of coins." https://www.davidlawrence.com/inventory/755761, 2023. Accessed: 2024-05-26.

[10] R. A. Bassett, *Computer-based objective interactive numismatic system*. Phd dissertation, Pace University, 2003. AAI3106821.

[11] P. Bhatia and various authors, "An improved hough transform for circle detection," in *Proceedings of the conference on image processing*, 2015.

[12] Council for Six Sigma Certification, *Lean Six Sigma Black Belt Certification Training Manual*, 2018. https://www.sixsigmacouncil.org/wp-content/uploads/2018/09/Lean-Six-Sigma-Black-Belt-Certification-Training-Manual-CSSC-2018-06b.pdf.

[13] M. Ansari, D. Kurchaniya, and M. Dixit, "A comprehensive analysis of image edge detection techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 12, pp. 1–12, 11 2017.

[14] S. L. Bangare and A. S. Dubal, "Reviewing otsu's method for image thresholding," *International Journal of Applied Engineering Research*, vol. 10, no. 9, pp. 21515–21526, 2015.

[15] J. Dannreuther, D. Hall, P. C. G. Service, and Q. Bowers, *The Official Guide to Coin Grading and Counterfeit Detection: Professional Coin Grading Service*. Official Guide to Coin Grading & Counterfeit Detection, House of Collectibles, 1997.

[16] Bullion Exchanges, "A guide to coin grading." https://bullionexchanges.com/blog/a-guide-to-coin-grading, 2023. Accessed: 2024-05-26.

[17] B. Conn and O. Arandjelović, "Towards computer vision based ancient coin recognition in the wild — automatic reliable image preprocessing and normalization," in *2017 International Joint Conference on Neural Networks (IJCNN)*, (Anchorage, AK, USA), pp. 1457–1464, IEEE, 2017.

[18] Professional Coin Grading Service, "Toning: Does it help or hurt a coin's value?." https://www.pcgs.com/news/toning-does-it-help#:~:text=While%20attractive%20toning%20cannot%20raise,%2FPR%2D67%20and%20up., 2015. Accessed: 2024-05-29.

[19] M. F. Fauzi, S. W. Hardianto, and A. Gunawan, "Determination of hue saturation value (hsv) color feature in kidney," *Journal of Physics: Conference Series*, vol. 2157, no. 1, p. 012020, 2022.

[20] M. A. Babyak, "What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models," *Psychosomatic Medicine*, vol. 66, no. 3, pp. 411–421, 2004.

[21] Professional Coin Grading Service, "Eye appeal: What it means and why it matters." https://www.pcgs.com/news/eye-appeal#:~:text=Eye%20Appeal%20is%20defined%20as,coloration%20to%20the%20coin's%20surface, 2019. Accessed: 2024-05-29.

[22] D. L. R. Coins, "dlrc color rating explained." https://blog.davidlawrence.com/dlrc-color-rating-explained/#:~:text=Our%20%22color%20rating%22%20or%20degree,%2C%20gold%2C%20and%20copper%20example., 2024. Accessed: 2024-08-22.

[23] American Numismatic Association, *The Official ANA Grading Standards for United States Coins*. Atlanta, GA: Whitman Publishing, 7th ed., 2013.

[24] "Multilayer perceptron - an overview | sciencedirect topics." https://www.sciencedirect.com/topics/computer-science/multilayer-perceptron. Accessed: 2024-05-29.

[25] L. Pinto-Coelho, "How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications," *Bioengineering (Basel)*, vol. 10, p. 1435, Dec 2023.

[26] S. J. Berggren, "Automated assessment of norwegian l2 essays." https://www.duo.uio.no/bitstream/handle/10852/69420/thesis.pdf?sequence=1&isAllowed=y. Accessed: 2024-05-29.

[27] S. Zheng, H. Xu, A. Mukhtar, N. Khalilpoor, and A. Yasir, "Estimating residential buildings' energy usage utilising a combination of teaching-learning-based optimization (tlbo) method with conventional prediction techniques," *Engineering Applications of Computational Fluid Mechanics*, vol. 17, p. 2276347, 2023.

[28] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[29] M. A. Muslim, T. L. Nikmah, D. A. A. Pertiwi, Subhan, Jumanto, Y. Dasril, and Iswanto, "New model combination meta-learner to improve accuracy prediction p2p lending with stacking ensemble learning," *Intelligent Systems with Applications*, vol. 18, p. 200204, 2023.

[30] D. Shanthi and N. Chethan, "Genetic algorithm based hyper-parameter tuning to improve the performance of machine learning models," *SN Computer Science*, vol. 3, no. 5, pp. 1–13, 2022.