

Machine Learning Application on Surface Roughness Forecast and Surface Texture Analysis

Weiting Lin^a

^a*University of California, Davis,
^b*Manufacturing and Sustainable Technologies Research (MASTeR) Laboratory, UC Davis,**

Abstract

This article explores the application of machine learning in predicting various surface roughness parameters, such as arithmetic mean height (S_a), maximum height of the surface (S_z), and texture aspect ratio (Str), from images of aluminum materials. Additionally, it conducts a cluster analysis to distinguish surface irregularities. The discussion extends to data processing and manipulation techniques applied to material image data, starting with the normalization of image data. Subsequently, it utilizes a pre-trained model, such as VGG-16, to extract significant features from the image data. To mitigate extensive computational costs, it implements the Principal Component Analysis (PCA) technique. Through this data processing approach, the study builds Convolutional Neural Network (CNN), XGBoost, and Stack Regression models to fit the image data and forecast the surface roughness parameters. Furthermore, it employs a Random Forest model in cluster analysis to determine the presence of irregularities in the images. Based on each model's performance in forecasting parameters, it's evident that they accurately forecast S_a and S_z , achieving maximum accuracy of 83% and 77% as defined by R-squared. However, they perform poorly on Str due to 57% maximum accuracy. Random Forest models excel in classifying images into four groups with a 95% accuracy and determining material surface uniformity with 86% accuracy.

1. Introduction

The significance of surface roughness in various industries cannot be overstated, as it plays a pivotal role in determining the performance, efficiency, and longevity of products and components. Precise measurement of surface roughness, such as S_a^1 , S_z^2 , and Str^3 , is a challenging task with traditional instruments like surfometers or stylus profilometers, often leading to inaccurate results and inefficiency. In Surface Roughness Prediction with Machine Learning (Wenhe Zhang, 2021)⁴, the paper showcases various machine learning methods for forecasting the R_a^5 , R_z^6 value across different manufacturing processes.

Machine learning, including models like CNN and VGG-16, offers a revolutionary approach to image analysis for faster, more accurate measurements. This project employs three different models: CNN, XGBoost, and Stack Regression, for forecasting the values of S_a , S_z , and Str from images of aluminum materials. It specifically explores four optimization methods for CNN models to determine which offers superior forecasting performance. Ensuring the cleanliness of input images is crucial to prevent data distortion before model fitting.

Furthermore, machine learning extends beyond measurement to classify materials into specific groups, enhancing safety, reliability, and reducing costs by minimizing manual labor and addressing safety concerns with materials that may have sharp edges or toxic chemicals. In Enhanced Material Classification via MobileSEMNet: Leveraging MobileNetV2 for SEM Image Analysis (Cihat Aydin, 2023)⁷ exemplifies MobileSEMNet model the potential of machine learning in material science, achieving an impressive classification accuracy of 10 different materials at 96.87%. This breakthrough demonstrates the transformative impact of machine learning on material classification and surface roughness measurement. In this project, Random Forest models classify images of aluminum materials into four groups and determine the surface uniformity of the images.

2. Data description

In this project, all photographs and data were sourced from VR-6000 Keyence optical profilometer⁸ (Figure 1). The collection process involved placing aluminum blocks on Keyence's observation platform, capturing images of each side of the block, and gathering variable data using Keyence's software. This resulted in a dataset of 1163 entries (Table 3), which were categorized into four groups based on the patterns in the images (Figure 2).

¹ S_a is the average roughness over a measurement area.

² S_z is defined as the sum of the largest peak height value and the largest pit depth value within the defined area.

³ Str is a measure of uniformity of the surface texture

⁴https://www.researchgate.net/publication/350747277_Surface_Roughness_Prediction_with_Machine_Learning

⁵ R_a is referred to as the arithmetic mean roughness

⁶ R_z is referred to as the maximum roughness

⁷<https://ieta.org/journals/ts/paper/10.18280/ts.400638>

⁸<https://www.keyence.com/products/microscope/macroscope/vr-6000/>

	Types				Total
	1	2	3	4	
Raw data	460	179	365	159	1163

Table 1: The amount of each type across raw data



Figure 1: VR-6000 Keyence optical profilometer with aluminum blocks

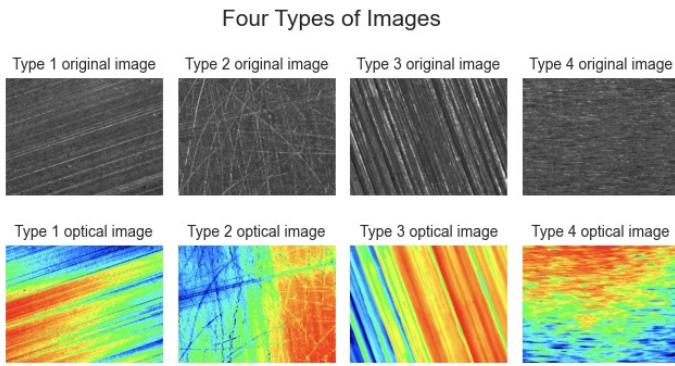


Figure 2: Four sets of original and corresponding optical images.

The distribution plots (Figure 3) illustrate the overall data distribution and the distributions across three parameters. It is evident that skewness is present in all the distribution plots. Therefore, to bring the data closer to a normal distribution—which aids in enhancing the model training process and forecast accuracy—further data transformation is necessary. The following section will provide more detail on data processing.

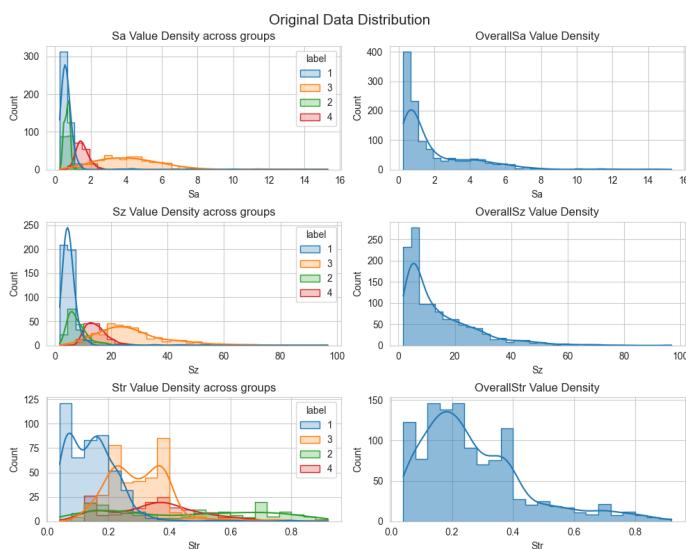


Figure 3: Original data distribution across parameters and overall distribution

The box plots for each parameter across groups reveal distinct outcomes. For the Sa and Sz parameters, type 3 surface images exhibit the highest median values, aligning with their significant surface topography variations. For the Str parameter, half of the type 2 surface images show uniform surfaces with Str value greater than 0.5, consistent with the lack of specific directional patterns. In the Sku⁹ parameter, the majority of images are spiky, as over 50% of the data has Sku values greater than three, indicating a distribution with more pronounced peaks.

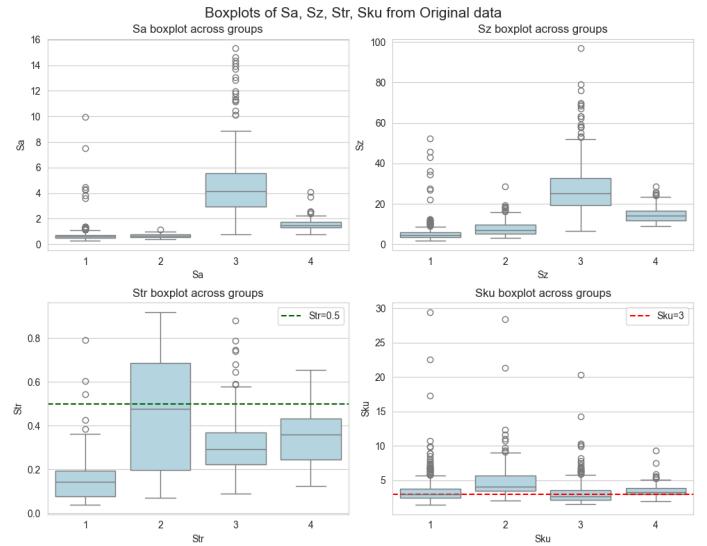


Figure 4: Box plots ach parameter across groups

3. Data analysis

3.1. Data processing

3.1.1. Data transformation

To mitigate data skewness, standard techniques such as log transformation and square root transformation are employed. Table 2 quantifies the proximity of the data distribution to a normal distribution. If the absolute difference between the skewness of the data and 1, the skewness value for a normal distribution, is close to zero, it suggests that the data tend toward a normal distribution.

Skewness results

Parameters	Raw data & 2 data transformation methods		
	Raw data	Log	Square root
Sa	1.309	0.433	0.247
Sz	0.852	0.736	0.059
Str	0.231	1.359	0.539

Table 2: The absolute difference between the skewness of the data and 1 across parameters

⁹Sku value is a measure of the sharpness of the roughness profile. Generally, Sku has a normally distributed height when Sku equals 3.

Based on Table 2, data transformation has successfully shifted the distribution of the three parameters closer to a normal distribution. Figures 5 and 6 display the data following log transformation and square root transformation, respectively. Notably, the data after log transformation seems to approach a normal distribution. Consequently, this project employs two distinct response datasets—resulting from these two transformations—in the training models within the modeling phase.

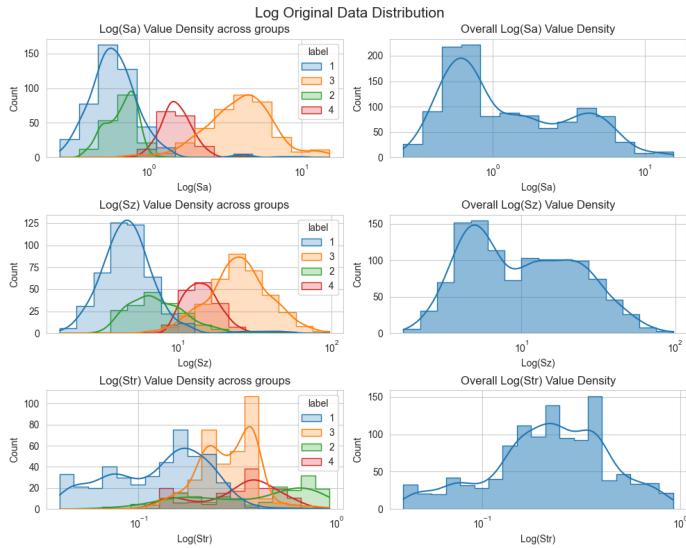


Figure 5: Log transformation data distribution across parameters and overall distribution

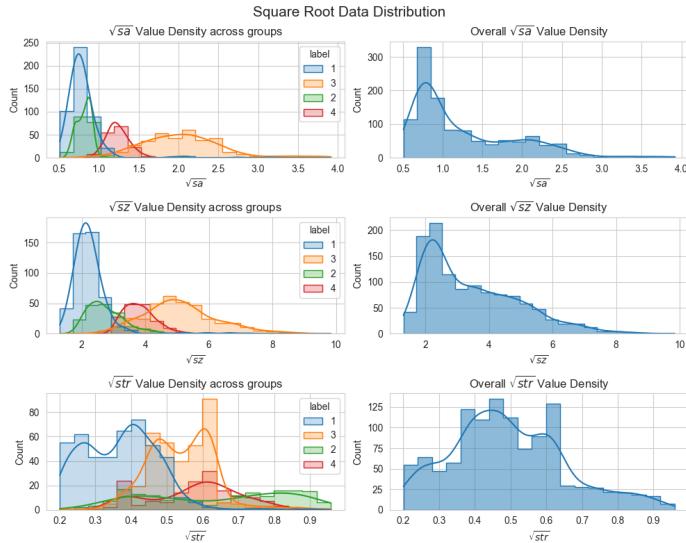


Figure 6: Square root transformation data distribution across parameters and overall distribution

3.1.2. CLAHE method on images

In addition to normalizing the response variable, Sa, Sz, Str, it is equally crucial to address the resolution of the images. In their original form, the images may lack sufficient resolution,

obscuring significant features such as the direction of grooves, valleys, and peaks. To tackle this issue, applying the Contrast Limited AHE (CLAHE) method can balance the contrast in each image, revealing the underlying insights. In Figure 7, the histograms of pixel quantities demonstrate that post-CLAHE images display a wealth of surface texture information, surpassing results from simply adjusting overall contrast and brightness, which is labeled as an enhanced image in Figure 7.

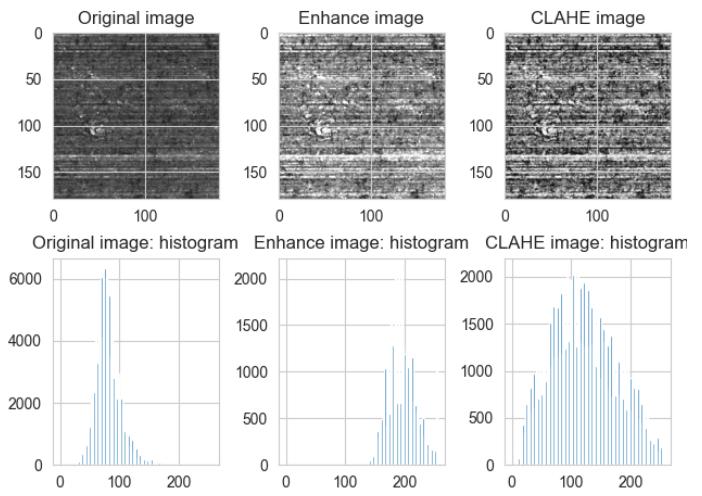


Figure 7: Distribution of Log transformation data without outlier across parameters and overall distribution

The Figure 8 below represents Histogram Equalization. The essence of the CLAHE method involves applying Histogram Equalization to each block, which is a small area of the image; collectively, these blocks compose the entire image. The method employs contrast limiting techniques to prevent excessive noise. In this technique, the histogram is truncated at a predetermined value. Any excess in the histogram bins beyond this limit is evenly redistributed across all bins, thereby moderating the enhancement of contrast and controlling noise.

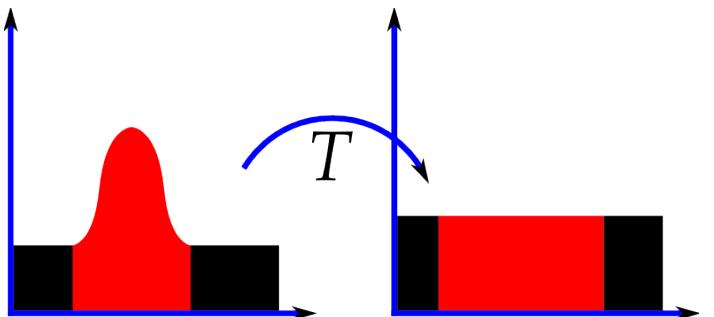


Figure 8: Histogram Equalization (source: https://en.wikipedia.org/wiki/Histogram_equalization)

3.1.3. Uniformity classification

Based on information from the Keyence website¹⁰, the Str parameter is used to determine the uniformity of a material's surface. However, it does not offer a robust standard for uniformity. The provided information suggests that as the Str value approaches 1, the surface is more uniform, whereas a value closer to 0 indicates a surface with a clear direction. Therefore, this project adopts 0.5 as the critical value for the Str parameter: if the Str value is greater than or equal to 0.5, the surface of the image is classified as uniform; if the value is less than 0.5, it is classified as non-uniform.

The Figure 9 indicates the uniform and non-uniform surfaces among four types of material surfaces.

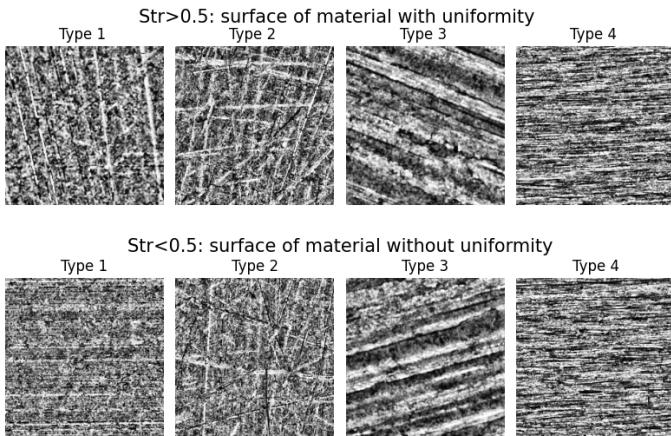


Figure 9: The number of images with uniform surfaces among the four types of surface images.

However, there is a significant discrepancy between the number of uniform and non-uniform images; the number of non-uniform images is approximately 8.6 times that of the uniform images.

Uniform	Non-uniform
120	1043

Table 3: The number of images within the uniform and non-uniform classifications

Upon closer examination of each type in Figure 10, it is clear that type 2 images constitute the majority of the uniform surface images and it complies with the Figure 2.

To increase the number of uniform images, this project rotates the uniform images by 90 degrees, 180 degrees, and 270 degrees, and also flips the images horizontally and vertically. As a result, we now have 720 uniform images in Table 4 and a total of 1763 images as the dataset for classification models, which will be discussed in the data modeling section. The dis-

¹⁰<https://www.keyence.com/ss/products/microscope/roughness/surface/str-texture-aspect-ratio.jsp>

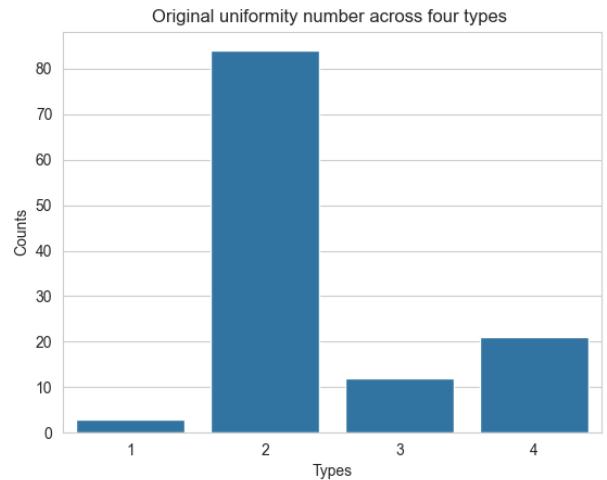


Figure 10: The original number of images with uniform surfaces among the four types of surface images.

tribution of uniform and non-uniform images across four types after adjustment is shown in Figures 11.

Uniform	Non-uniform
720	1043

Table 4: The number of images within the uniform and non-uniform classifications after increasing uniform images

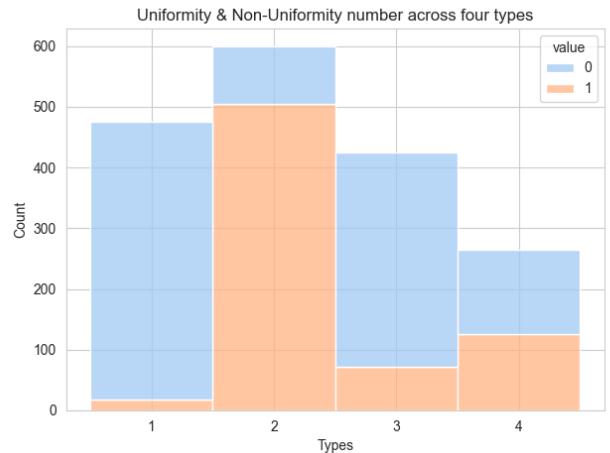


Figure 11: The final number of images with uniform and non-uniform surfaces among the four types of surface images (0:non-uniform, 1:uniform).

3.1.4. Number of cleaned samples

As previously mentioned in the abstract, this project aims to achieve two objectives: constructing a model for predicting surface roughness parameters and for image classification, which includes classifying images into four types, and determining the uniformity of the image surface.

The datasets for the two objectives vary. Utilizing flipped and rotated images as datasets for predicting surface roughness parameters may result in a lack of variation, leading to a poor model fit. Therefore, the dataset size for predicting surface roughness parameters will remain the same as the original dataset, with 1163 entries. For classification tasks, augmenting the dataset with flipped and rotated images is feasible because the classification outcomes, being categorical predictors, are less sensitive to variations than continuous predictors. Consequently, the total dataset size for the classification model is 1763.

Parameter prediction model	Classification model
1163	1763

Table 5: The number of images cross different objectives

3.2. Data modeling

3.2.1. Feature extraction and feature reduction

Since the quantity of each image type in this project is not substantial and the distribution is uneven, it is preferable to use one model to fit all data types rather than creating four separate models for each dataset. A single model with a consolidated training sample will likely yield better forecasting results.

To ensure enhanced model performance in forecasting Sa, Sz, and Str values, extracting potential features from images using a pre-trained model is essential. The pre-trained model used in this project is the VGG-16 (Figure 12 is its structure) a type of CNN model. Its layers progressively learn finer details, making it well-suited for a broad range of image recognition tasks. Since the objective is to extract features using the VGG16 model, only the convolutional layers are utilized, omitting the fully connected layers that are typically used for classification tasks in the final stages of the network.

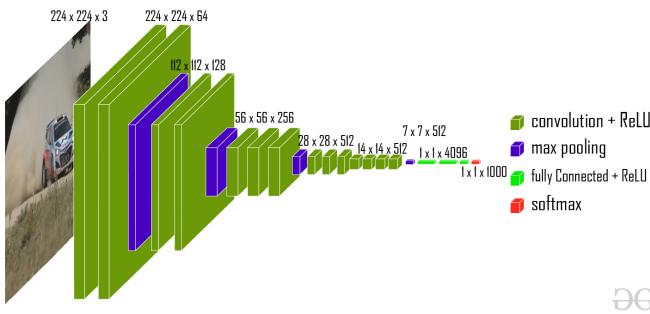


Figure 12: VGG16 construction (source: <https://www.geeksforgeeks.org/vgg-16-cnn-model/>)

However, the number of features obtained after processing with VGG16 stands at 25,088, a substantial figure that may result in increased computation time for further model training. Another advantage is that employing a feature reduction method can also enhance the stability of feature quality. Thus, implementing Principal Component Analysis (PCA) subsequent to the VGG16 model—to trim down the features to a quantity that accounts for 90% of the image variability—could

mitigate this computational challenge. Table 6 illustrates the count of features throughout the feature extraction process of each image.

Feature extraction process	Input(raw data)	VGG16	PCA
Feature number	150,528	25088	309

Table 6: The number of features in the feature extraction process

3.2.2. Models - Predicting parameter values

The objective of this section is to predict the Sa, Sz, Str of the aluminum material surface. This project utilizes three types of models, each enhanced through regularization methods to improve performance:

1. Convolutional Neural Network (CNN)

CNN, a type of deep neural network, is extensively used for analyzing 2D imagery. In this phase and Figure 13, the CNN model comprises connected layers, which include two dense layers and two dropout layers. Because the VGG-16 model utilized in this project consists exclusively of convolutional and max pooling layers. Therefore, it is essential to construct dense and dropout layers to complete the architecture of a comprehensive CNN model.

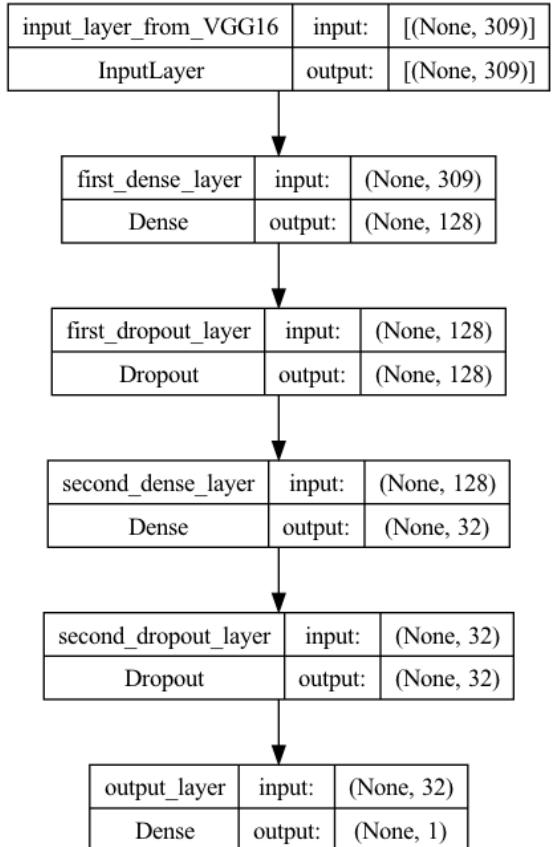


Figure 13: CNN model structure

Optimization methods:

- Stochastic gradient descent (SGD) 5

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x^{(i)}, y^{(i)}) \quad (1)$$

where θ represents the parameters, η is the learning rate, and $\nabla_{\theta} J(\theta; x^{(i)}, y^{(i)})$ is the gradient of the objective function with respect to the parameters for a single sample $(x^{(i)}, y^{(i)})$.

- Root Mean Square Propagation (RMSprop) 5

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t \quad (2)$$

where $E[g^2]_t$ is the moving average of the squared gradients, and η is the learning rate.

- Adam 5

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (3)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (4)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (5)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (6)$$

$$\theta_{t+1} = \theta_t - \frac{\eta \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (7)$$

where m_t and v_t are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients, respectively. β_1 and β_2 are the exponential decay rates for these moment estimates, and ϵ is a small scalar used to prevent division by zero.

- Adagrad 5

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{G_t + \epsilon}} \odot g_t \quad (8)$$

where G_t is a diagonal matrix where each diagonal element i, i is the sum of the squares of the gradients with respect to θ_i up to time step t , and \odot denotes element-wise multiplication.

2. XGBoost

Represents “Extreme Gradient Boosting” and is typically used for supervised learning problems. Optimizing the objective function, which encompasses both the loss function and the regularization function, is crucial for determining the best model, aiming to minimize the value of the objective function. XGBoost employs decision tree ensembles as its model, comprising several sets of Classification and Regression Trees (CART).

3. Stacked Regression

Stacked regression involves combining base models including regression or classification (CART models) with a Meta-learner, which is another model used to find the optimal integration of base models. This approach enhances predictions based on the results from the base models. In this case, the base model of Stacked Regression includes linear regression and decision tree regression and ridge regression serves as the Meta-learner, synthesizing the weights of the base models to improve prediction performance.

3.2.3. Models - Classification

The objective of this section is to classify aluminum material surfaces into four groups and to determine their uniformity. In this section, the model employs Random Forest, which consists of numerous decision trees and combines them to provide precise and accurate predictions.

3.3. Forecast

3.3.1. Model performance-Sa, ,Sz, Str value prediction

In the data transformation section, this project adopts two methods—log transformation and square root transformation—on surface roughness parameter values. To quantify the performance of models, using the R-square (R^2)¹¹ value appears to be a reasonable method. Instead of running each model once, we run each model 50 times to estimate the R-square statistics, which include the maximum, minimum, median, and mean values. Besides, serving 30% of data as testing data and remaining 70% of data as training data.

1. CNN models

The tables from 7 to 12 showcase the performance of CNN models for three parameters across four optimizers and two data transformation methods. To determine which transformation method more effectively enhances model training and parameter prediction, the median R-square (R^2) is utilized for comparison due to its reduced sensitivity to outliers, offering valuable insights.

Firstly, within Table 7 and Table 8, it is evident that CNN models with log-transformed Sa input outperform those with square root-transformed Sa input in terms of median R-square (R^2) across four optimizers. Notably, within the Adagrad optimizer, the CNN model with log-transformed Sa input achieves the highest max R-square (R^2) at 0.83 compared to the other three optimizers. This suggests that, to some extent, this model accurately predicts the Sa parameters with approximately 83% accuracy.

¹¹stands for the proportion of the variance that could be explained by the model, a negative value may indicate that the model is not suitable

Best R^2 : XGBoost

Parameters	Data transformation methods	
	Log	Square root
Sa	0.75	0.67
Sz	0.68	0.69
Str	0.57	0.22

Table 13: Performance of XGBoost model across 2 data transformation methods for Sa, Sz, Str prediction

ividually. In both Table 14 and Table 15, both data transformation methods implemented on Str as input of the stacked regression model provide a similar Str parameter median prediction accuracy of around 26%.

Log: Stacked regression

parameters	R square (R^2) statistics			
	Max	Min	Median	Mean
Sa	0.76	0.72	0.75	0.75
Sz	0.72	0.70	0.71	0.71
Str	0.27	0.24	0.26	0.26

Table 14: Performance of stacked regression model with log-transformed input for Sa, Sz, Str prediction

\sqrt{x} : Stacked regression

Parameters	R square (R^2) statistics			
	Max	Min	Median	Mean
Sa	0.70	0.66	0.68	0.68
Sz	0.66	0.63	0.65	0.65
Str	0.29	0.25	0.26	0.26

Table 15: Performance of stacked regression model with square root-transformed input for Sa, Sz, Str prediction

To sum up, CNN model with Adagrad optimizer could bring the best Sa and Sz prediction accuracy at 83% and 77% respectively with log-transformed data as model input. As for Str parameter prediction, using XGBoost model with log-transformed data as model input could provide the best Str parameter prediction with nearly 57% accuracy. Based on the results above, using log-transformed data as the model input for predicting three parameters is more effective than using square root-transformed data as the model input.

Parameters	Sa	Sz	Str
Best model	CNN with Adagrad	CNN with Adagrad	XGBoost
R^2 value	0.83	0.77	0.57

Table 16: Best models for three parameters prediction

3.3.2. Model performance-Group classification, Uniformity classification

In this section, using the Random Forest model is sufficient, due to its strong classification ability. Both Random Forest

models contain 100 decision trees, which are helpful to classify the surface image into four groups or to determine the uniformity of the surface image.

Moreover, using 30% of data as testing data to determine the performance of classification models. In the confusion matrix, calculating sensitivity is crucial because it can decrease the type II error rate, which represents the proportion of undetected positive results. Sensitivity provides insights into the proportion of correctly identified positives among those with the condition.

Precision is another vital indicator of model classification performance. It provides straightforward insight into the accuracy across each group.

However, to solve the condition regarding imbalanced samples across groups, F1-score is beneficial to balancing the precision and sensitivity. In general, high precision, sensitivity, and F1-score indicate strong model performance.

$$\text{Precision} = \frac{\text{True Positive}}{\text{Number of Samples}} \quad (9)$$

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (10)$$

1. Group classification

According to Table 17, the number of samples across four groups is imbalanced. Hence, F1-score plays an important role in this case. It is obvious that F1-score in each type of image group exceeds 0.9 and the overall accuracy is around 95%. Hence, the Random Forest model showcases its strong ability to classify images into four groups.

Type	Precision	Sensitivity (Recall)	F1-score	Number of samples
1	0.93	0.92	0.92	143
2	0.95	1	0.97	180
3	0.98	0.92	0.95	127
4	1	1	1	79

Table 17: Result for group classification

2. Uniformity classification

In this case, 0 represents the non-uniformity surface image, and 1 represents the uniformity surface image. In Table 18, the number of samples across two groups is imbalanced, so considering F1-score as a model performance indicator is reasonable. Both groups share the F1-score surpassing 0.8, and the model has better classification performance in determining the non-uniformity surface images with higher sensitivity at 0.9. The overall accuracy is around 86%

In conclusion, Random Forest models demonstrate exceptional performance on classification tasks, with an overall accuracy exceeding 85% in the two cases mentioned above.

Groups	Precision	Sensitivity (Recall)	F1-score	Number of samples
0	0.88	0.90	0.89	313
1	0.85	0.81	0.83	216

Table 18: Result for uniformity classification (0:non-uniformity, 1:uniformity)

4. Discussion

Given the results from applying machine learning methods to surface roughness prediction, image classification, and determination regarding uniformity of the surface, it is demonstrated that the application of machine learning to surface texture analysis is feasible. This approach could reduce costs associated with workload and the purchase of precision measurement equipment. However, the success of the models heavily depends on data collection; insufficient data and unbalanced samples across groups can significantly impact model performance. Moreover, not all surface roughness parameters are suitable for the same machine learning models, as seen with the Str parameter in this project. This is because each surface parameter has different measurements, and the factors affecting these parameters vary. Therefore, understanding the background of surface parameters and employing data processing methods to ensure data quality suitable for the proper model is crucial.

5. Reference

1. CLAHE

- Image Contrast Enhancement Using CLAHE
- Histogram Equalization

2. Optimizer equation

- SGD

Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press

- RMSprop

Understanding RMSProp: An Adaptive Learning Rate Method

- Adam

Kingma, D. P., Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

- Adagrad

Duchi, J., Hazan, E., Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research, 12, 2121-2159. Available

3. Convolutional Neural Network (CNN)

- Basic Introduction to Convolutional Neural Network
- 4. XGBoost
- XGBoost Tutorials
- 5. Stacked Regression
- How to Use Stacking to Choose the Best Possible Algorithm?