

LiDAR-based Characterization of Rock Size Distribution Using Machine Learning Techniques

Santiago Callero¹, Anqi Zhang^{2,3}

¹Department of Petroleum and Geosystems Engineering, University of Texas at Austin

²Center for Perceptual Systems, University of Texas at Austin

³Department of Physics, University of Texas at Austin

Abstract

Drilling for Oil & Gas can be a risky business, and costs are highly dependent on operational surveillance. The feedback that arises from the characterization of the cuttings that come out of the well as drilling progresses can be highly valuable to address downhole dysfunctions and drilling risks. Unfortunately, this is still manually done and evaluated by “specialized personnel”, thus resulting in discrete feedback from the sampled interval and biased based on the expert’s input. Based on this motivation, our project takes on the task of segmenting LiDAR images of rocks into individual regions, and characterizing each of the rock samples’ size class based on extracted geometric features. We applied a complete pipeline of machine learning techniques, including data preparation, model selection, model training, model calibration, model evaluation, and model interpretation. We found that axis minor length, the first Hu moment, and the max height are important features to classify the sieve size class of rocks. Our project serves as a computational foundation of a LiDAR-based automated pipe-sticking anomaly detection algorithm.

Introduction

During the drilling operations of oil or gas, it is common to encounter a “stuck pipe” event, in which a pipe cannot be freed from the well without damaging the involved equipment. Consequences of a stuck pipe can result in one of the most costly problems during drilling (Muqeem et al., 2012), including rig downtime, downhole equipment damage and loss, lost production, as well as the safety risks to personnel. Severe “stuck pipe” events can implicate the complete loss of the drill string and downhole equipment, and even the loss of the already drilled well with the need of a sidetrack well (a new well drilled deviated from the original one). Particle size distribution (PSD) characterization of the removed cuttings from the annulus is considered as a useful and timely indicator (Brankovic et al., 2021) to detect some of the mechanical causes, such as hole caving, sloughing, or collapse.

Thurley (2011) characterized PSD with 3D ranging techniques. Engin & Maerz (2019) furthered the characterization process in a streamlined fashion with Light Detection and Ranging (LiDAR). Using 3D LiDAR data could provide a more accurate, efficient, and user-friendly detection

approach. In our paper, we trained classification models to predict the sieve size class of a rock based on the statistics of its 3D LiDAR data.

Methods

Rocks were placed on the convey belt and passed through LiDAR (**Figure 1**). Location points of rock surfaces were recorded with the Cartesian coordinates (X,Y,Z). Then the points were transformed into the image format. Specifically, the pixel coordinate is inherited from X and Y values, and the pixel value is equal to the Z value.

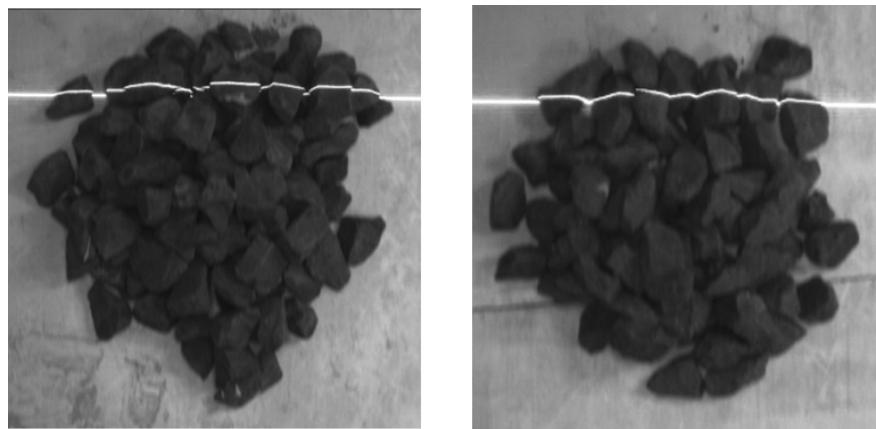


Figure 1. Rock sample in conveyor belt measured by high-frequency laser beam (gray scale image).

Rocks of different sizes were separated by sieving. Each image includes rocks with a certain known size class. Specifically, the largest size class that was left after sieving corresponded to $\frac{3}{4}$ inch-diameter holes. The middle size class went through $\frac{3}{4}$ inch-diameter holes and was left after sieving with $\frac{5}{8}$ inch-diameter holes. The smallest size class went through $\frac{5}{8}$ inch-diameter holes and was left after sieving with $\frac{1}{2}$ inch-diameter holes. From here, we label them as L, M, S classes, respectively.

We then use thresholding to identify the rock region and apply the watershed algorithm to segment the region of each rock. Within each region, we extracted features such as axis major length, axis minor length, and max height. We then built a data table with each rock per row, including the geometric features as independent variables, and the size class of the rock as the dependent variable. No missing and invalid values were present.

The geometric features we extracted include perimeter, area, axis major and minor length of the encapsulating ellipse, mean, median, max, and standard deviation of the rock height, mean,

standard deviation of the gradient of the rock height using the Sobel filter, the gradient of the maximal height location, and the first 4 Hu moments for shape description (Hu, 1962).

The overall data structure was visualized with Principal Component Analysis (PCA), T-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP). They reduced the dimensionality of data for intuitive, visual clustering.

To build an efficient and robust classification model for the natural statistics behind our data set, we dropped the highly correlated features, and split the data into training, validation, and test sets. Then we tuned the hyperparameters with the validation set for several families of classification models: K-nearest Neighbors Classifier (KNN), Support Vector Classifier (SVC), Random Forest Classifier (RF), XGBoost Classifier (XGB), and Multi-Layer Perceptron Classifier (MLP). The default single tree model was used as the baseline model.

KNN assumes data points of the same or similar classes are close to each other. SVC assumes the decision boundary between classes should be as wide as possible. RF assumes the data can be partitioned into regions describable with binary rules. XGB assumes each tree can correct the errors of the previous trees and the data can be described with binary rules. MLP assumes that the data are smooth enough to be well approximated by the specific architecture.

Except for XGB, all other classification models are based on the scikit-learn package in Python (Pedregosa et al., 2011). We looped through different “n_neighbors” for KNN and defined weights as either uniform or distance. We allowed the variation of the regularization parameter for squared L2 penalty, and the kernel type in SVC. For RF, we fitted the number of trees with various categorical criteria (gini, entropy, log loss) and the number of features to consider when looking for the best split (sqrt, log2). The XGB Classifier is optimized from various max depth, number of trees, and L2 regularization weight. For MLP, activation function and layer architecture are modified for the best performance.

Within each family of models, the hyperparameters that induced the highest F1-score were determined with a combination of cross-validated (5-fold) grid search and Bayesian optimization, so that we could compare the performance of two search algorithms and make it more likely to find the global optimum.

For the best model of all families of classification models, we tested its accuracy and F1-score on a separate test set. Then we calibrated it with isotonic and sigmoid functions. We tried two methods: (1) binarize the class label based on the focused class versus the other classes, then fit the model for separate calibration; (2) fit the model with the original 3-class labels and calibrate

it all together. For the multi-class calibration case (1), stratified 5-fold validation is used. Important features were found using Shapley Additive Explanations (SHAP). Partial Dependence Plot (PDP) were plotted for the most important features.

Results

The original LiDAR images of rocks from different sizes of sieves are shown in **Figure 2**. It shows the majority of rocks have largely elliptical shape with one or two pointy ends. Rocks were placed in a single layer, and mildly packed.

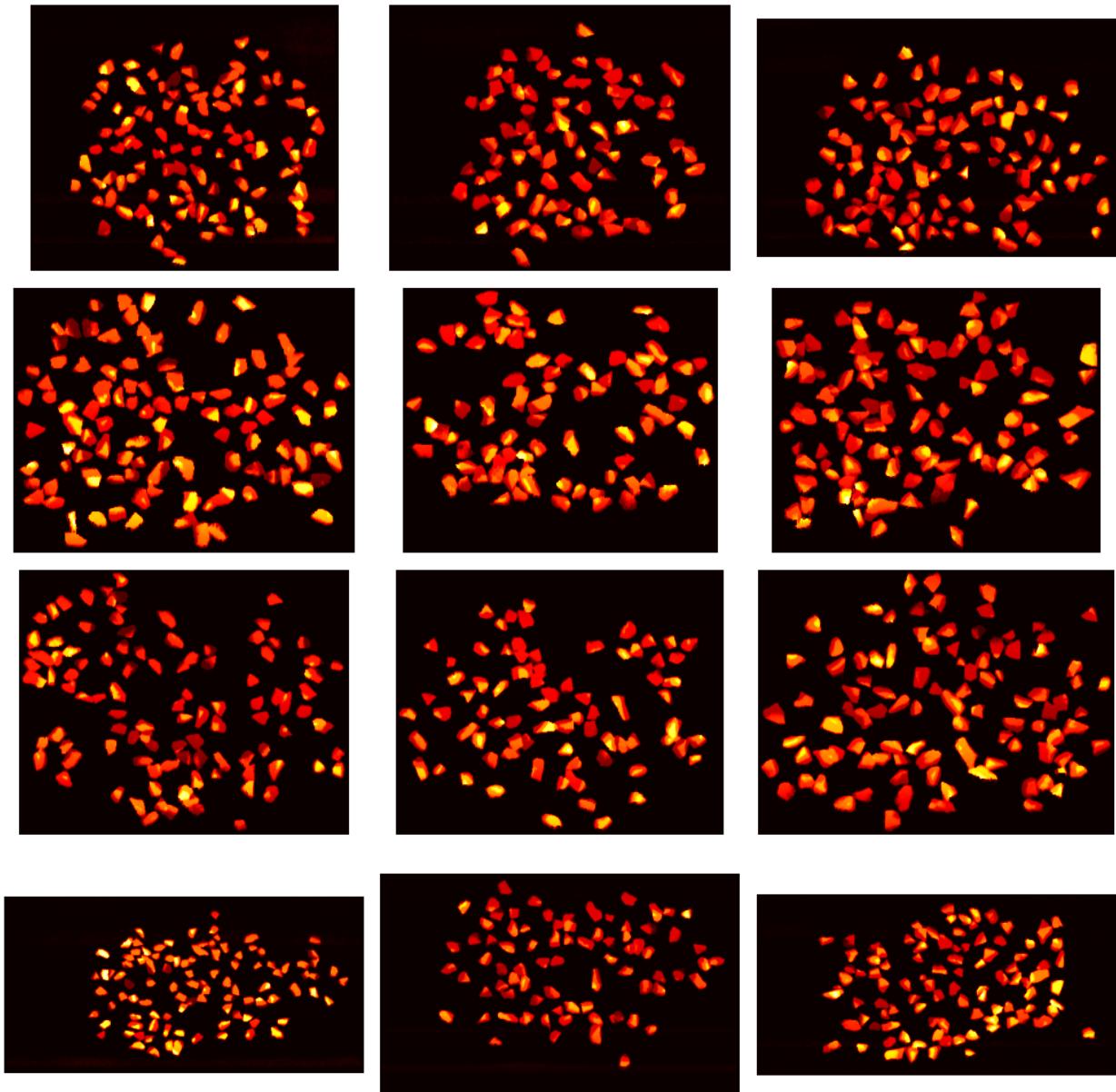


Figure 2. LiDAR images of rocks in the data set. The left, middle, right columns correspond to S, M, L size classes of sieved rocks. Each size class was collected 4 times with a random

arrangement of distinct rocks. The brighter the pixel is, the larger the Z-value at that pixel location is.

The thresholding and watershed algorithms segmented the rock regions by individual rocks, as shown in **Figure 3**. The segmentation performance was visually examined in retrospect, yielding 2.2% of omission (two rocks were segmented as one) and 2.0% percent of duplication (one rock was segmented as two).

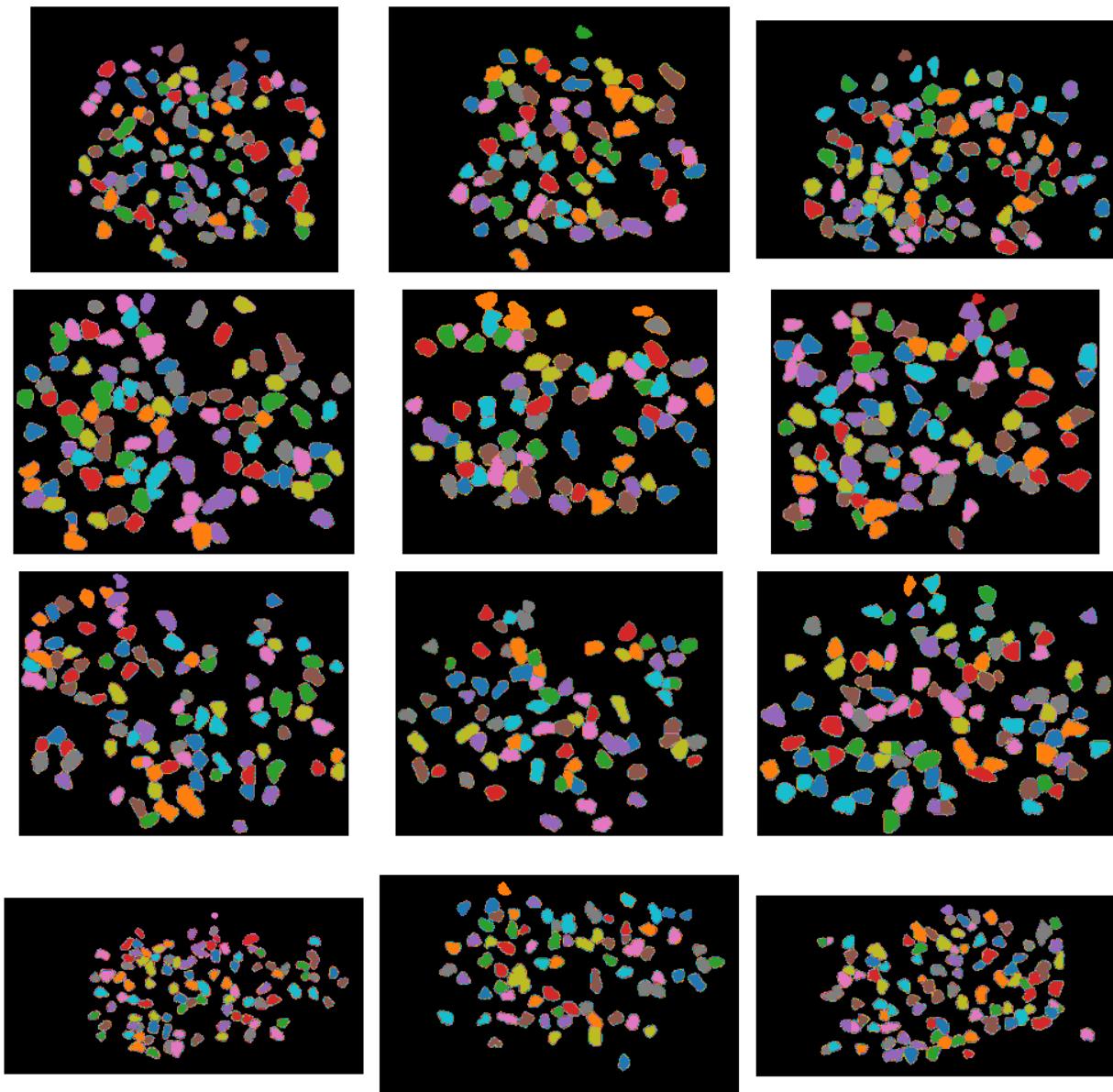


Figure 3. Segmented regions of LiDAR images of rocks. The left, middle, right columns correspond to S, M, L size classes of sieved rocks. Each size class was collected 4 times with a

random arrangement of distinct rocks. Different colors indicate different regions, while neighboring regions with the same color does not necessarily mean they are segmented as the same region. Numbers of rocks by size class are S: 393, M: 304, L: 393.

According to **Figure 4**, some geometric features are highly correlated, such as between perimeter and area, between mean height and max height. For model fitting, we decided to drop perimeter, area, mean height, median height, and standard deviation of height, to avoid the wild swings of coefficient estimates due to multicollinearity. **Figure S1** (found in Supplementary Materials Appendix) provides more details on the distribution of geometric features by class and features correlation. Based on the correlation between geometric features and sieve size class, considering the physics of the sieving process, we hypothesize that the axis minor length, max. height, and the weighted 1st Hu moment are the most important features to predict the size class of a rock. We reasoned that if the axis minor length is smaller than the sieve hole diameter, then the rock is more likely to pass through the sieve; that is similar for the max. height and the weighted 1st Hu moment.

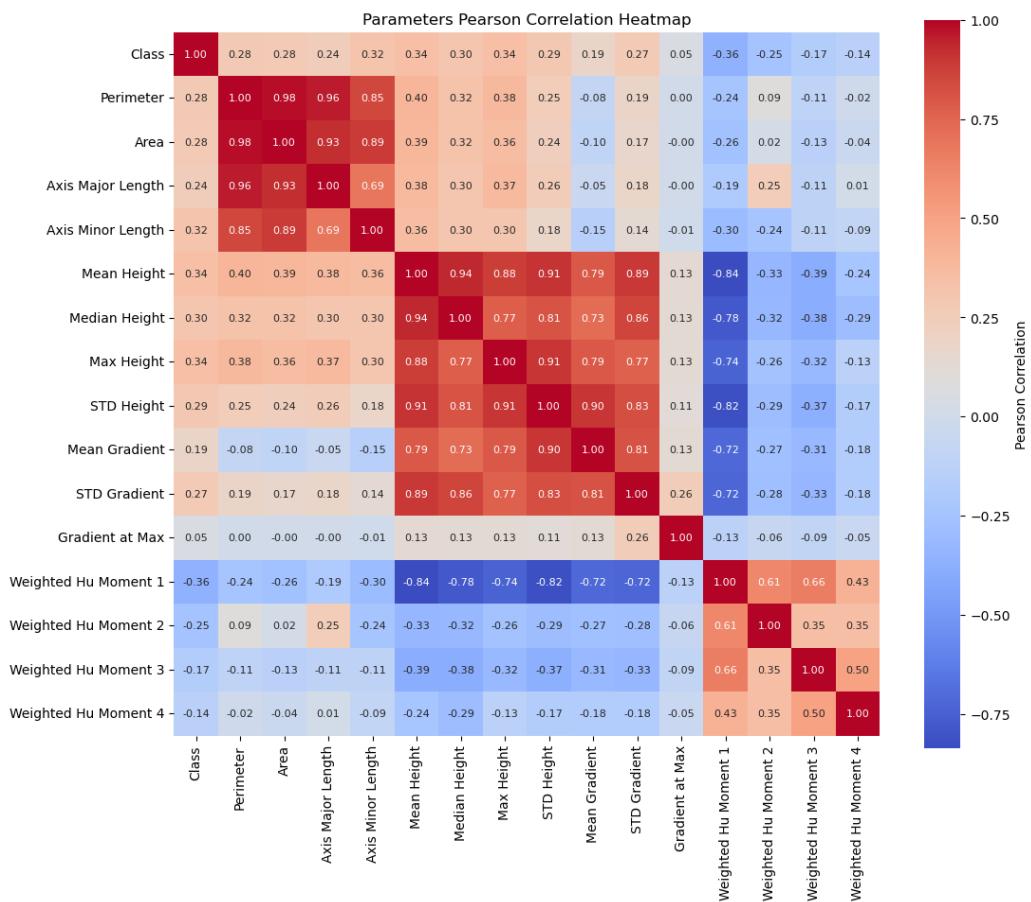


Figure 4. Pearson correlation heatmap between the geometric features of rock samples.

As the next exploratory step, we reduced the dimensionality of selected features into 2 dimensions, so we can understand the data structure visually and expect the classification difficulty of this data set. It seems that distinguishing between the S and L classes is very simple for non-linear models, and the M class is hardest to classify as it overlaps with both other classes.

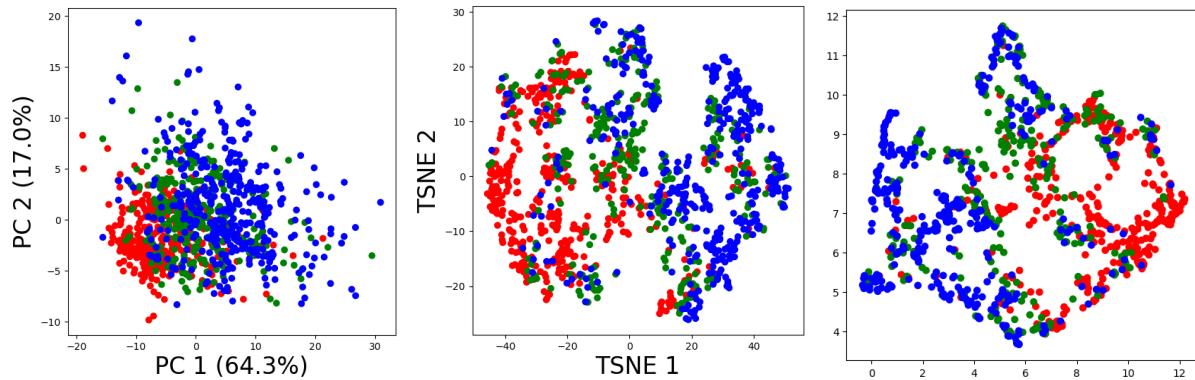


Figure 5. Dimensionality reduction techniques for the rock size data set. From left to right: PCA, t-SNE, and UMAP. The sieve size classes are S (red), M (green), and L (blue).

Table 1 summarizes the hyperparameter tuning for selecting the optimal model among each model family. Based on recall, KNN is the best model. However, based on the precision or F1-score, we chose SVC as the best model for further optimization and analysis. Notice that the single tree baseline model is outperformed by all other classification models here.

Model	Parameters	Dataset	Precision	Recall	F1-score
SVC	'C': 0.1, 'kernel': 'linear'	max_height	0.719	0.706	0.711
KNN	'n_neighbors': 14, 'weights': 'distance'	mean_height	0.698	0.711	0.702
XGB	'max_depth': 4, 'n_estimators': 76, 'reg_lambda': 0.001	max_height	0.705	0.702	0.702
MLP	'activation': 'logistic', 'alpha': 0.001, 'hidden_layer_sizes': (100, 50)	mean_height	0.695	0.702	0.698
RF	'criterion': 'entropy', 'max_features': 'log2', 'n_estimators': 200	max_height	0.7	0.693	0.695
TREE	'criterion': 'gini', 'max_depth': 5	max_height	0.67	0.62	0.64

Table 1. Summary of model selection for each model family and for different metrics. The color continuum from blue to red implies the metric value from small to large.

For more details, the confusion matrices of TREE and SVC are shown below in **Figure 6**. The major improvement lies in the category where the actual rock size is L and the predicted rock size is M or L. SVC is much more reliable to be correct in identifying L-size rocks.

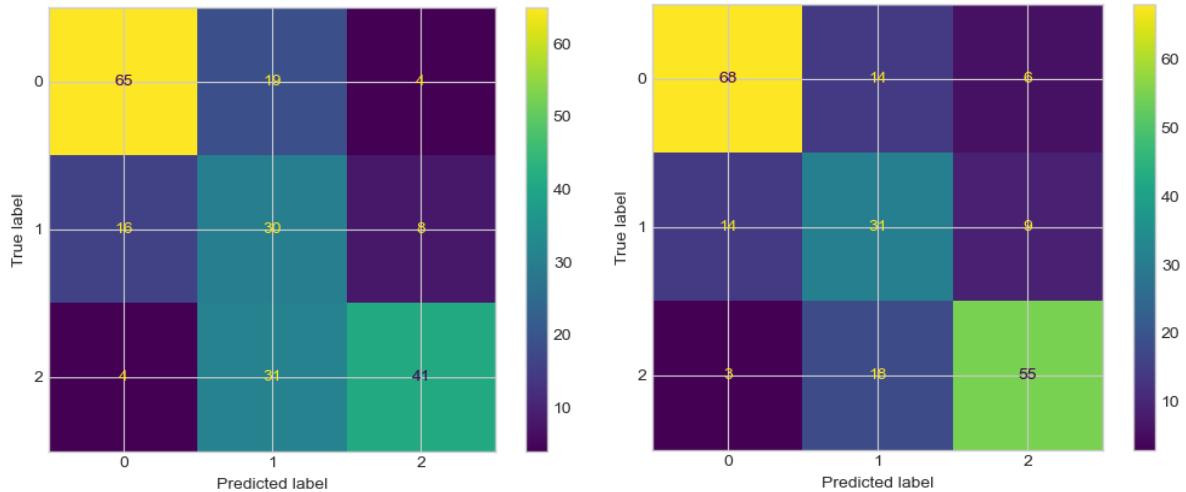


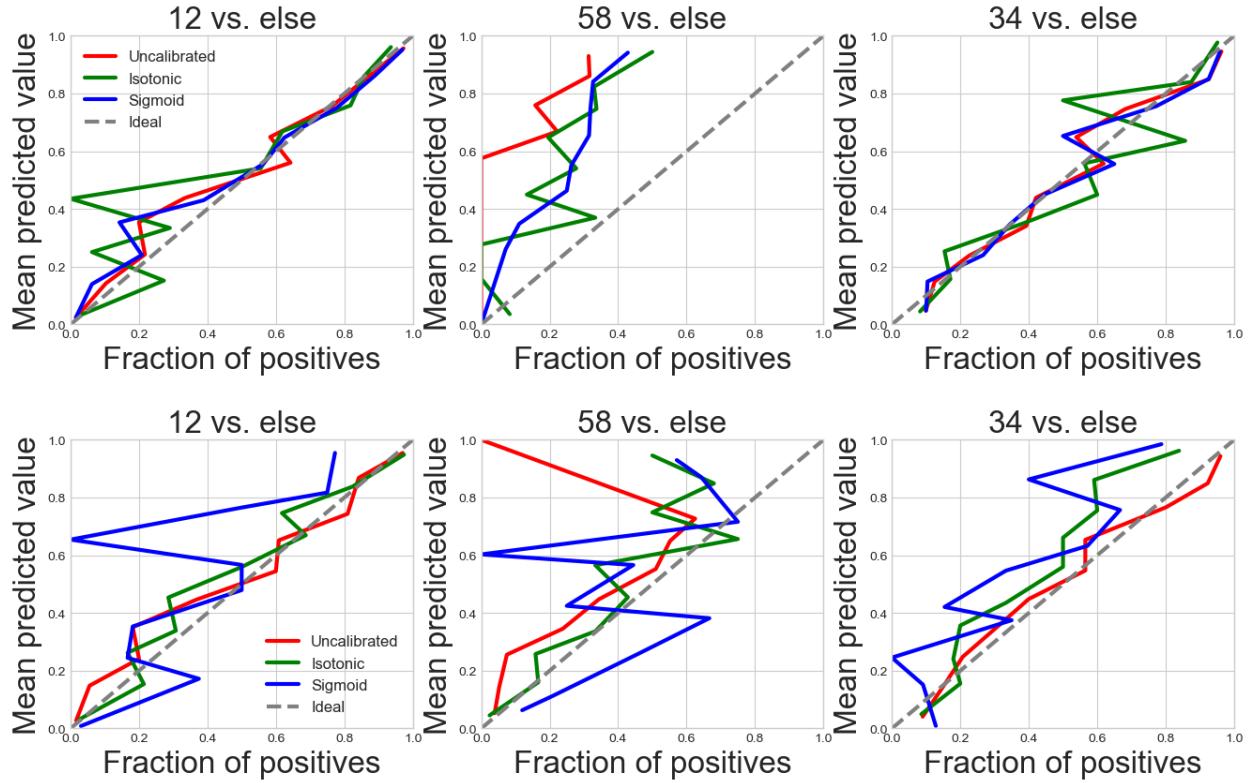
Figure 6. An example of the confusion matrix of the baseline model (TREE) on the left and the best model (SVC) on the right.

We then performed model calibration, testing isotonic and sigmoid functions. As mentioned before, two methods were tested: (1) binarize the class based on the focused class versus the other classes, then fit the model for separate calibration; (2) fit the model with the original 3-class labels and calibrate it all together. For the multi-class calibration case (1), stratified 5-fold validation is used.

Figure 7 below shows the calibration results, where the first and third plot correspond to the first method, second and fourth correspond to the second method.

Although the values for the first and third class look better for the first method, the middle class tends to drift away on the predicted values vs. fraction of positives. A similar effect happens with the second method.

The ROC curves (plots 3rd and 4th on **Figure 7**) show that method 1 yields better results with higher AUCs for the middle class, where the uncalibrated method still works better than the calibrated ones.



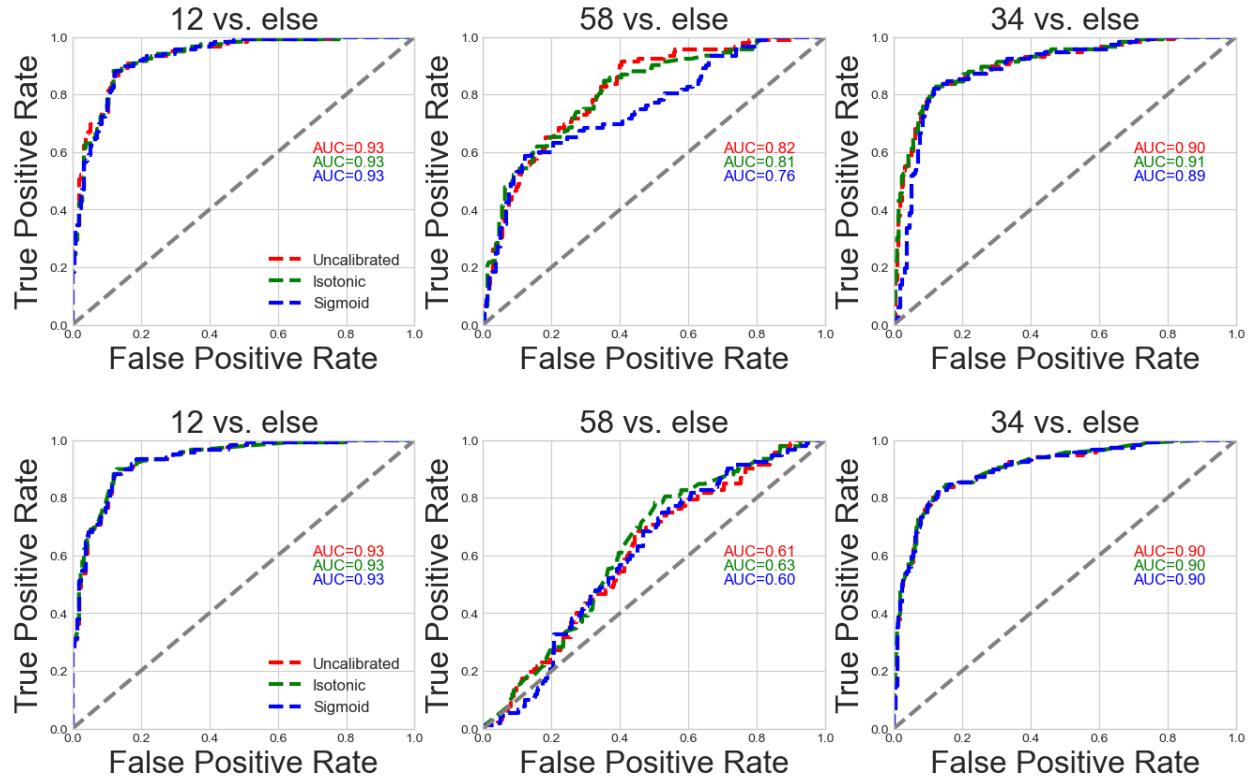
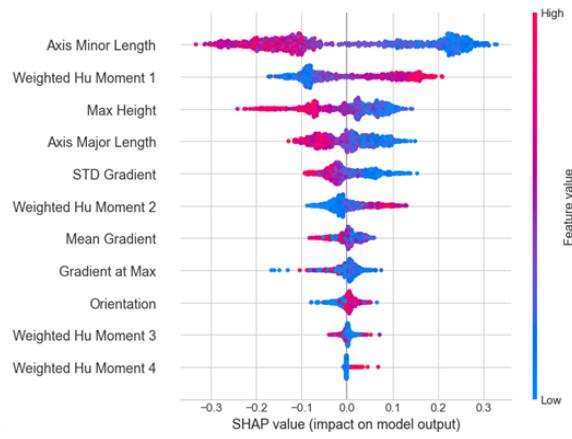
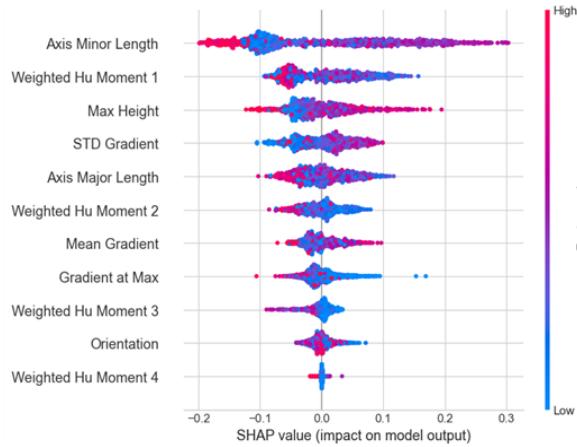


Figure 7. Calibration plots with the “one versus else” strategy. Rows from top to bottom: Mean Predicted Value vs. Fraction of Positives, with calibration procedure (1) in the **Methods** Session; Mean Predicted Value vs. Fraction of Positives, with calibration procedure (2) in the **Methods** Session; ROC curve with calibration procedure (1) in the **Methods** Session; ROC curve with calibration procedure (2) in the **Methods** Session.

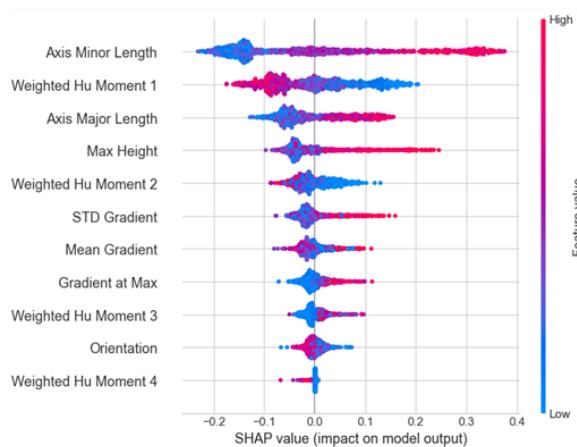
Feature importance analysis for the best performing model was done with SHAPLY (SHAP values), to understand which are the key factors that influence the classification. As observed in **Figure 8**, Axis Minor Length is the feature with the most impact on all 3 classes, followed by the Weighted Hu Moment 1 and Max. Height.



Class 1/2''



Class 5/8''

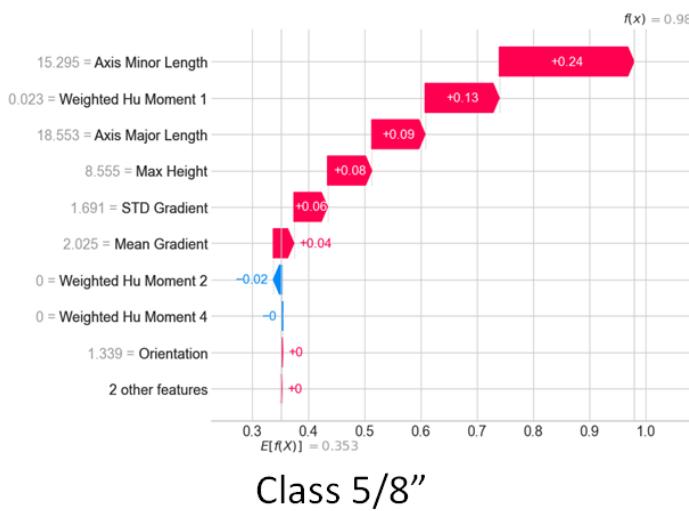
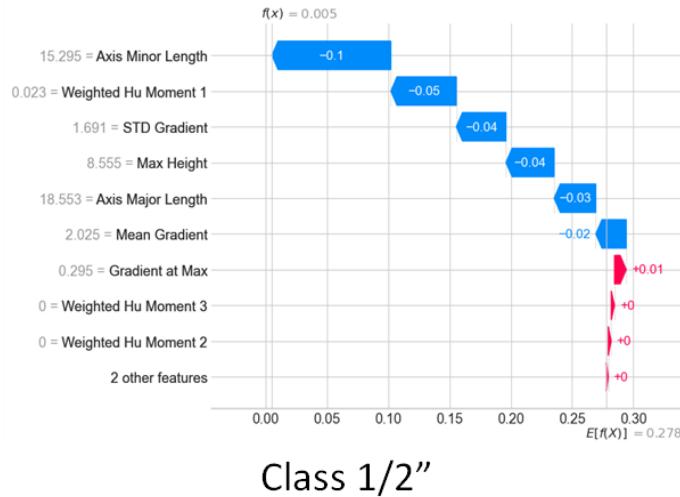


Class 3/4''

Figure 8. SHAP values for best model for classes $\frac{1}{2}''$, $\frac{5}{8}''$ and $\frac{3}{4}''$ (S, M, L)

This is again in line with our initial hypothesis, and in agreement with the physical nature of the classification system (rock sieving). We can also observe that for the smaller class ($\frac{1}{2}$ ") the lower the Axis Minor Length value is, the closer to the class. The same effect happens with the Max. Height, indicating that these values might set the classification threshold for every class. We can observe the inverse effect for the biggest class ($\frac{3}{4}$ ") and a mixture for the middle class ($\frac{5}{8}$ ").

Next, we continue with the SHAP values analysis for a single instance, to understand the impact of features on a single classification instance. As observed in **Figure 9**, the selected instance was a part of class $\frac{5}{8}$ ". In the middle plot it is observed how Axis Minor Length and Weighted Hu Moment 1 are the ones that contribute the most towards the classification. Then the model also takes into account Axis Major Length and the Max. Height. The opposite happens with the minor and major classes, but with the same main features.



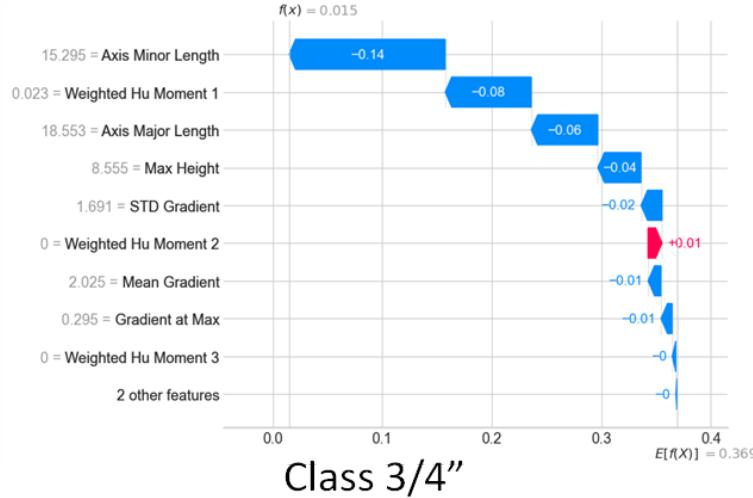
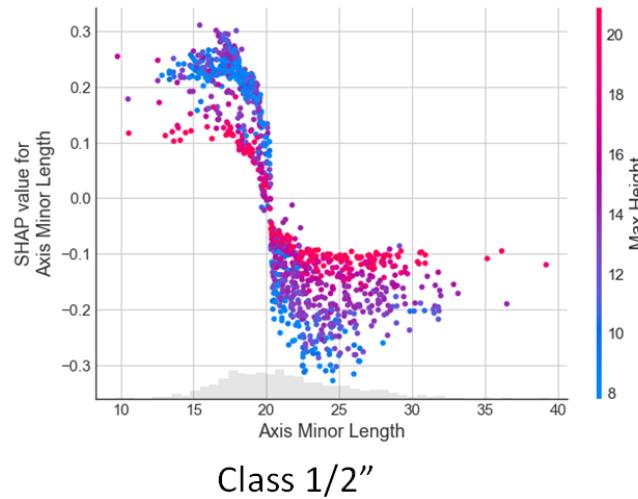
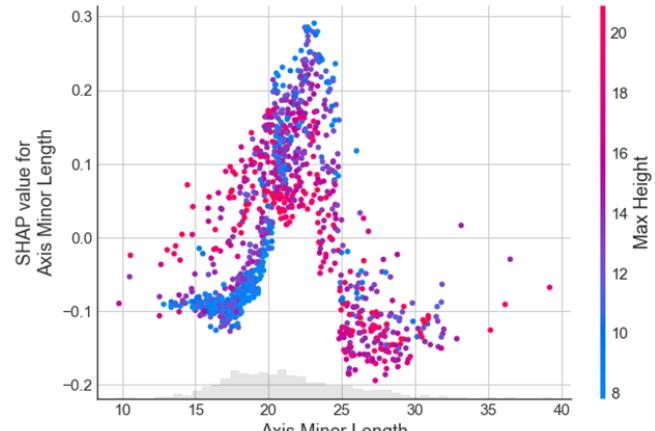


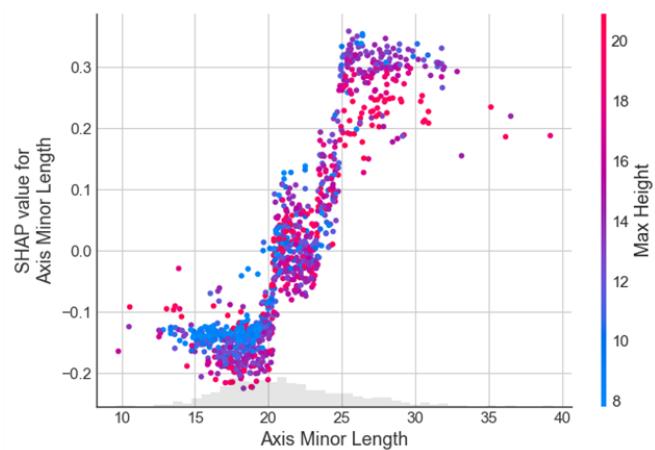
Figure 9. SHAP values for best model for a single instance from class $\frac{5}{8}$ " for classes $\frac{1}{2}$ ", $\frac{5}{8}$ " and $\frac{3}{4}$ " (S, M, L)

We then performed the SHAP Interaction Plots analysis for features Axis Minor Length and Max. Height. As **Figure 10** shows there are clear classification thresholds set for Axis Minor Length on every class. **Figure 10** seems to indicate that values less to 20 mm for Axis Minor Length correspond to class $\frac{1}{2}$ ", from 20 to 25 to class $\frac{5}{8}$ " and from 25 up to class $\frac{3}{4}$ ". Max. Height does not seem to have such a clear effect as the former, with mixed results.





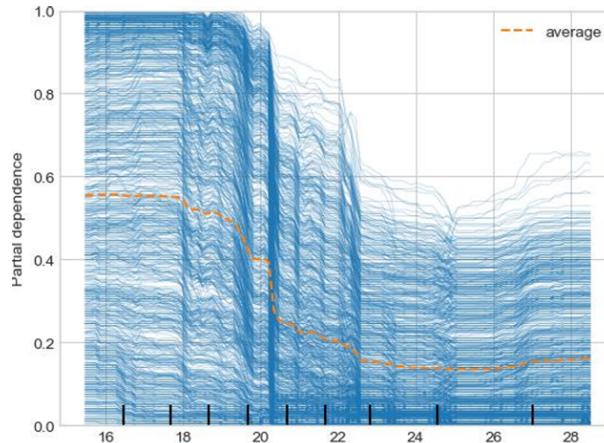
Class 5/8"



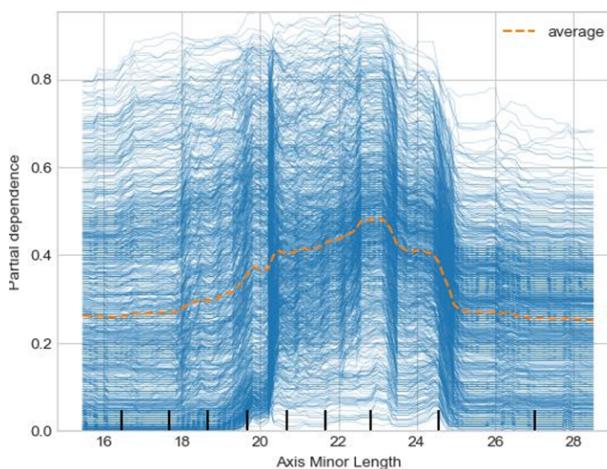
Class 3/4"

Figure 10. SHAP Interaction Plots for features Axis Minor Length and Max. Height

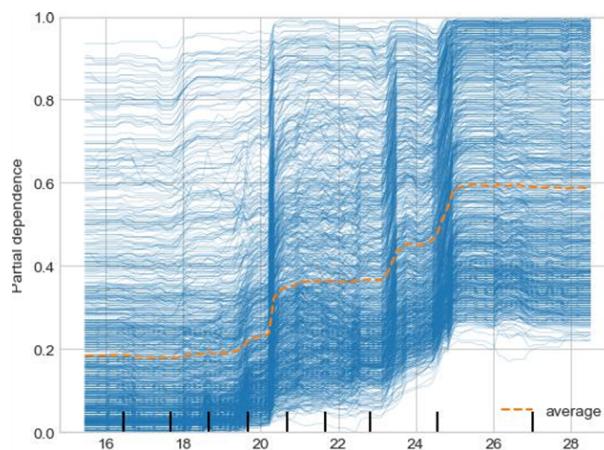
To get further understanding of the impact of the main features, we then ran partial dependence plots both for Axis Minor Length and Max. Height, for all 3 classes.



Class 1/2"



Class 5/8"

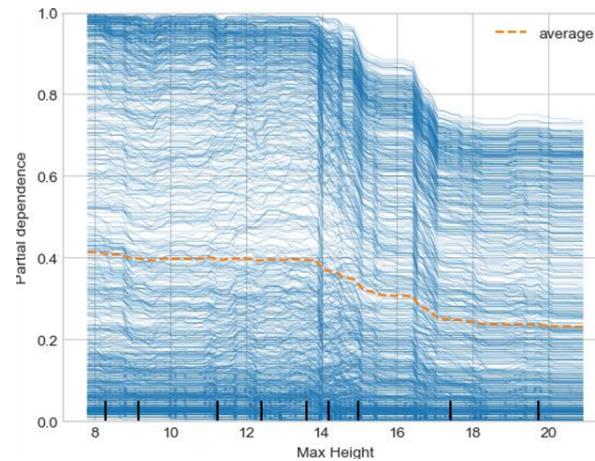


Class 3/4"

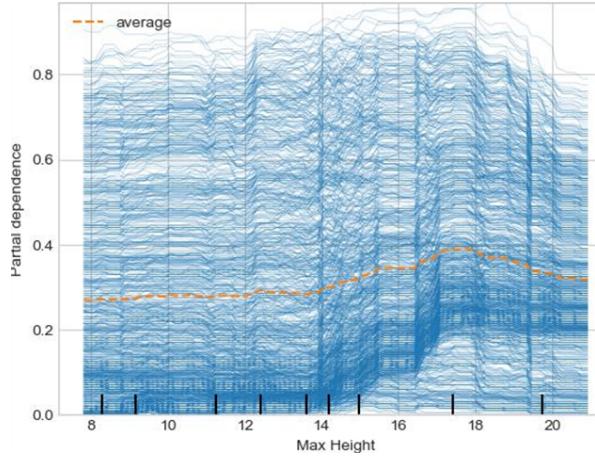
Figure 11. Partial Dependence Plot for feature Axis Minor Length, for all 3 classes

As observed in **Figure 11**, feature Axis Minor Length has different thresholds for classification among classes (as observed previously in Figure X for SHAP values). Thus, it has an important impact on the classification results.

In **Figure 12** it can be observed that although Max. Height has an impact on the final classification, it won't be as high as Axis Minor Length. For this feature there's also different ranges that seem to set classification thresholds.



Class 1/2"



Class 5/8"

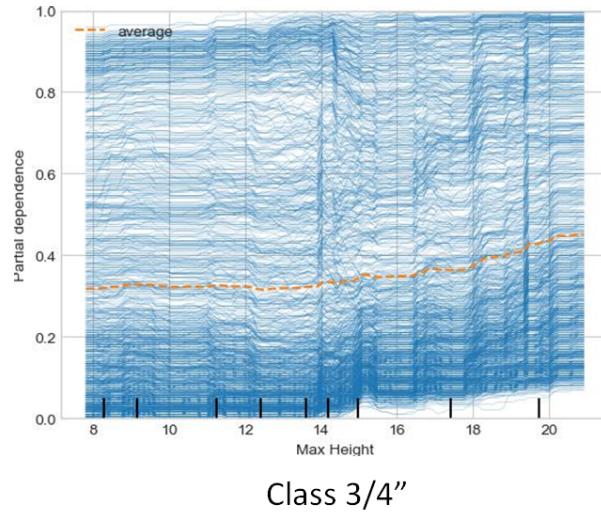


Figure 12. Partial Dependence Plot for feature Max. Height, for all 3 classes

As an additional test, we used Z-score Normalization for the dataset and compared resulting model performance versus the ones trained with the original dataset.

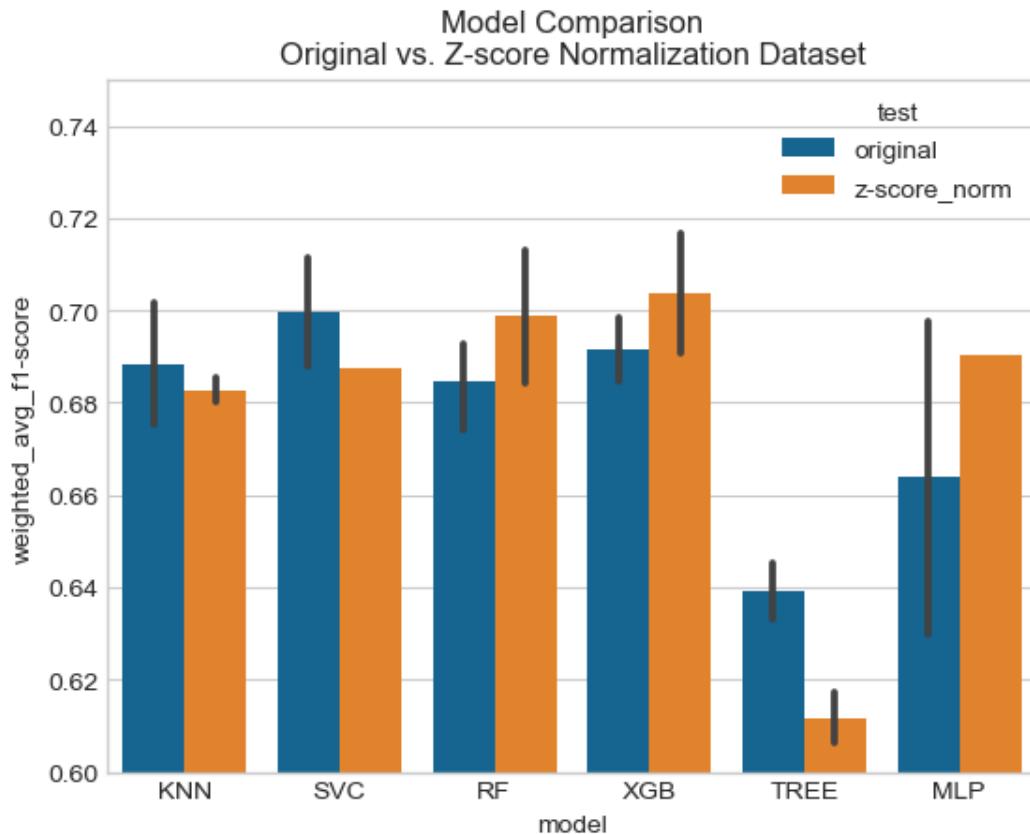


Figure 13. Weighted Avg. for F-1 Score for different models over original vs. normalized (scaled) datasets

As **Figure 13** shows the effect will not be the same for every model, and overall the difference in performance measured with the F-1 score is small. Thus, it was concluded that the feature normalization is not required.

Discussion

There are several factors resulting from the lab generated data that are worth taking into account for future development and deployment for the real-world application.

First, there's inherent classification error generated during the data generation phase, both by the sieving error and the segmentation error. These will then impact the performance of the classification models. Although the segmentation error was quantified, the sieving error was not. Future work should focus on quantifying and minimizing this error for input data.

Then there are several limitations in the model that come from the way that data was generated in laboratory conditions.

- Single-layer & Spatially separated samples: Rock samples were laid in single and sparse layers in the belt (as shown in **Figure 14**), making the segmentation task a bit easier.

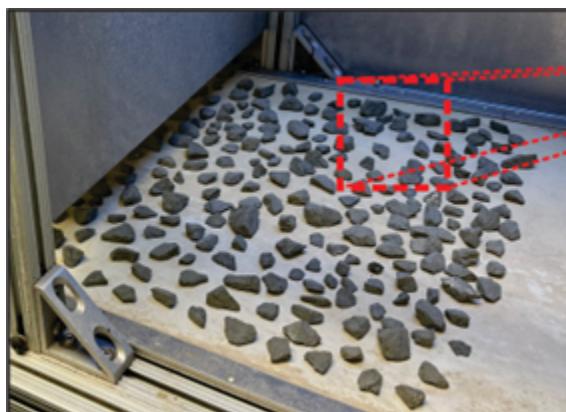


Figure 14. Single-layered sparse sample on sensor belt

In the real application the material is not necessarily spaced out, and can be presented as grouped or layered material (**Figure 15**).



Figure 15. Real application samples

- Clean samples: Cuttings coming from the well will generally be wet in drilling mud, and won't be as dry as the samples ran in the lab. This will also represent additional segmentation challenges.

Thus, training the models on laboratory data will impact its performance on the real-world application. Future work needs to consider data generated while drilling a well, and samples ran at the lab can also consider these types of conditions (wet cuttings and grouped samples).

Conclusion

Modeling results and analysis show that the model classification logic follows a similar logic to the sieving classification. And thus, Max. Height and Axis Minor Length work as predictive features.

As observed in **Figure 16**, the main features distribution has an important level of overlap between different classes (the same was observed for PCA analysis). This will increase the complexity of the classification problem, mainly for the middle class, as observed from the model results.

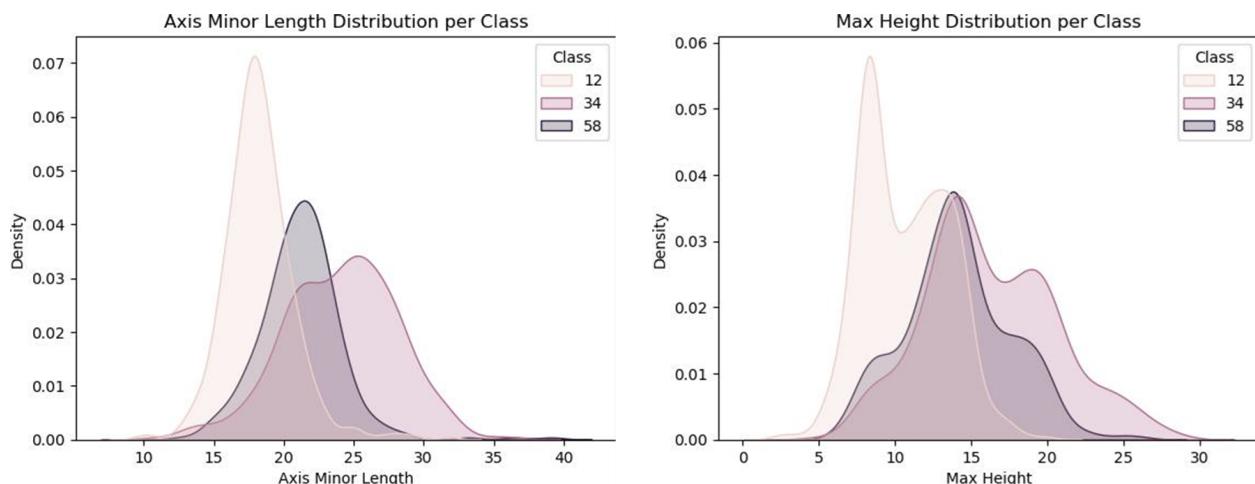


Figure 16. Axis Minor Length and Max. Height distribution per class

Future work needs to focus on:

- Minimizing the input data error coming from the sieve classification error and segmentation error
- Samples measured need to accurately reproduce real-world conditions (cuttings wet in drilling mud, and grouped samples), to consider the model deployment in real conditions
- A broader range of size classes need to be considered to encompass the entire range of sizes of cuttings generated while drilling

Acknowledgment

Santiago Callero and Anqi Zhang contributed to this project equally.

References

- Brankovic, A., Matteucci, M., Restelli, M., Ferrarini, L., Piroddi, L., Spelta, A., & Zausa, F. (2021). Data-driven indicators for the detection and prediction of stuck-pipe events in oil&gas drilling operations. *Upstream Oil and Gas Technology*, 7, 100043.
- Engin, I. C., & Maerz, N. H. (2019). Size distribution analysis of aggregates using LiDAR scan data and an alternate algorithm. *Measurement*, 143, 136-143.
- Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2), 179-187.
- Muqeem, M. A., Weekse, A. E., & Al-Hajji, A. A. (2012, April). Stuck Pipe Best Practices—A Challenging Approach to Reducing Stuck Pipe Costs. In SPE Saudi Arabia Section Technical Symposium and Exhibition. OnePetro.
- Thurley, M. J. (2011). Automated online measurement of limestone particle size distributions using 3D range data. *Journal of Process Control*, 21(2), 254-262.
- Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

Supplementary Materials

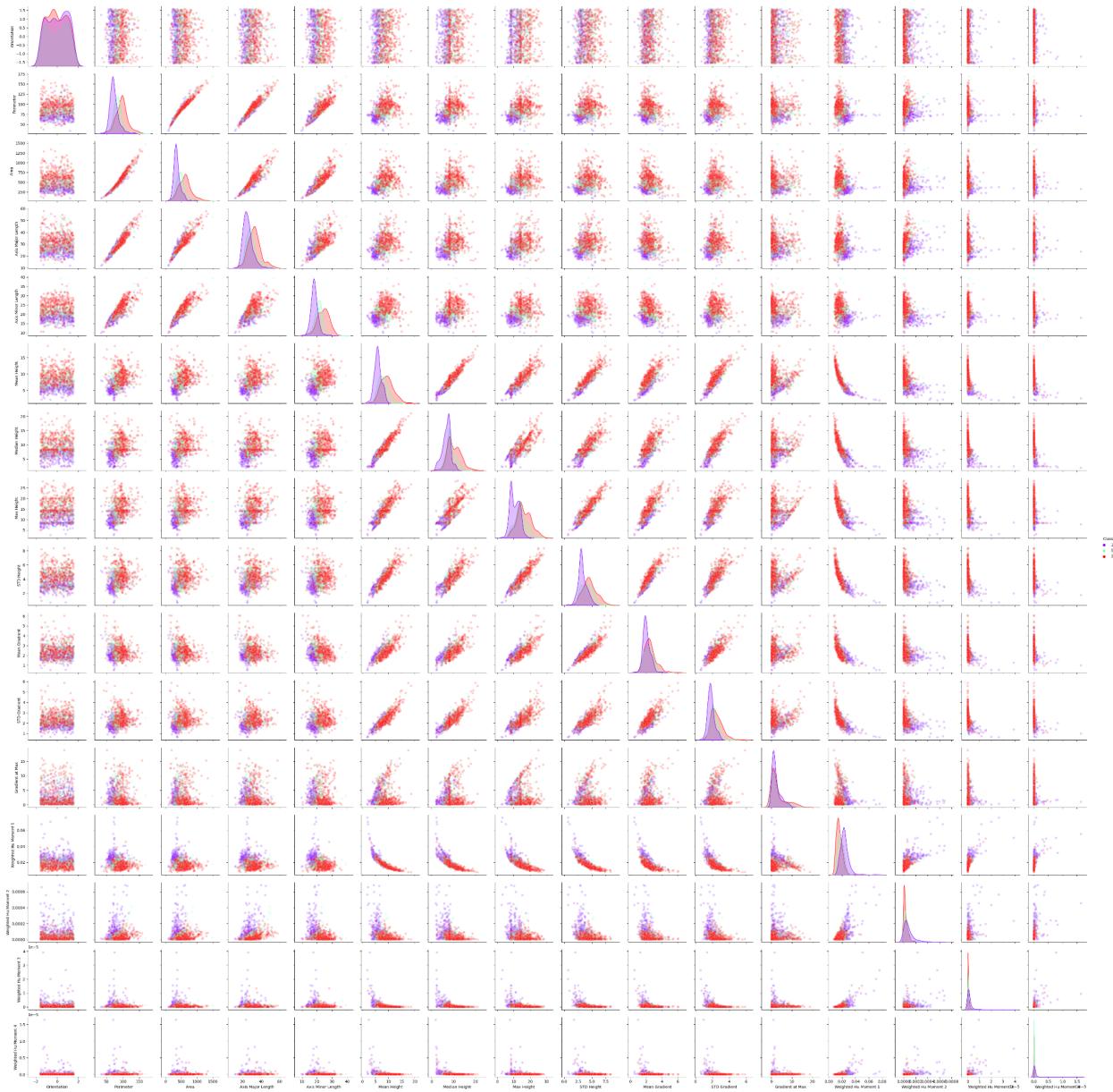


Figure S1. Pair plot between the geometric features of rock samples. The diagonal line shows the histogram of each feature by the size class.