Variable selection and feature extraction to identify important predictors for roughness in FDM

Avnish Rai, Samarth Karelia, Urjil Pandya, and Palash Gupta

Group No.15

1. Introduction:

Additive manufacturing process, fused deposition modeling uses a thermoplastic or composite material thread in filament form to construct 3D components. The plastic filament is fed via an extruder, melted at the extruding nozzle, and then layer by layer, automatically, is deposited onto the construction platform. The quality of the product is achieved by changing the input variables in the FDM. As seen in figure 1, the procedure includes the following changing variables. To achieve the target product quality, a combination of every changing variable is employed.

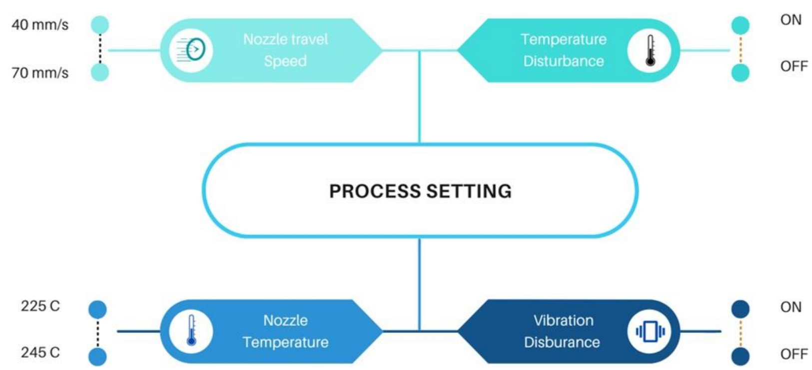

Figure 1: Input setting combination

2. Problem Statement:

This project aims to select the best combination of features/predictors to model the response (i.e., roughness) and to extract the features from the sensors data using functional. regression.

3. Data Structure:

There are 16 potential input combinations to predict response. For more accurate results, replicate these 16 outcomes three times and collect 48 samples to determine the ideal input setting for the desired output. During the process, data is gathered from a variety of sensors to analyze the influence of input variables on process quality. The various sensors used during the process were two tri-axis accelerometers (One located at nozzle and another at bed), Four thermocouples (One located at nozzle and three located at bed), One infrared (IR) sensor.

The surface quality was assessed at ten different locations in 48 samples in order to simulate the responses. The quality was predictor in out functional regression model.

4. Analysis Procedure and Results:

Data from diverse sources was cleaned and fused to perform various analyses, and data pre-processing activities were carried out to ensure improved accuracy. The extraction of features was performed by converting raw data into numerical features that may be handled while retaining the original data sets. It outperforms applying machine learning directly to raw data. It assists us in removing extraneous characteristics that have no discernible effect on the Dependent Variable. A further variable selection process was also carried out in order to identify the best input variable

A) Feature Extraction using Functional Regression (Avnish and Samarth):

A wavelet coefficient is the scalar product of a function (your observation) and a wavelet basis function. These wavelet coefficients are utilized in extracting characteristics from data. We used the wavelet approach to identify features from data of sensors that had the greatest effect on the dependent variable: roughness. The following are the findings of vibration data:

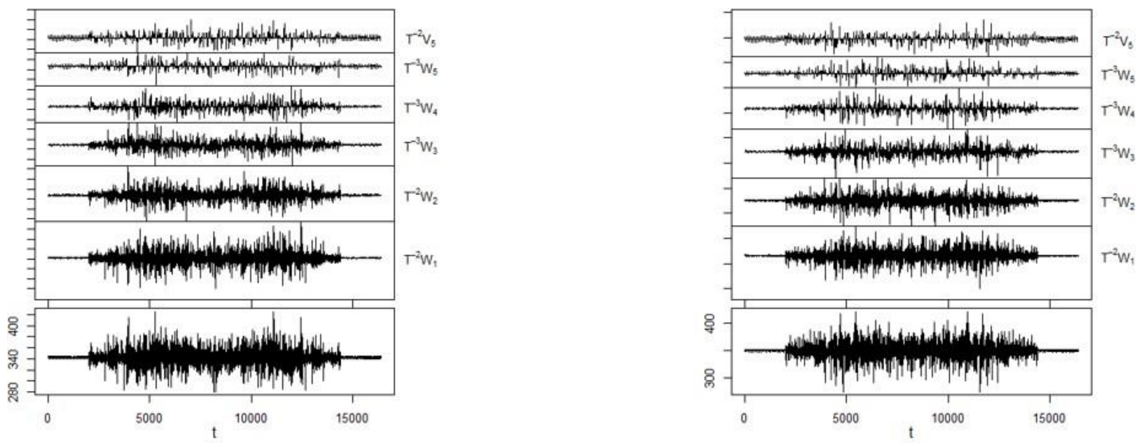

Figure 2: Wavelets for Vibration V1 and V2

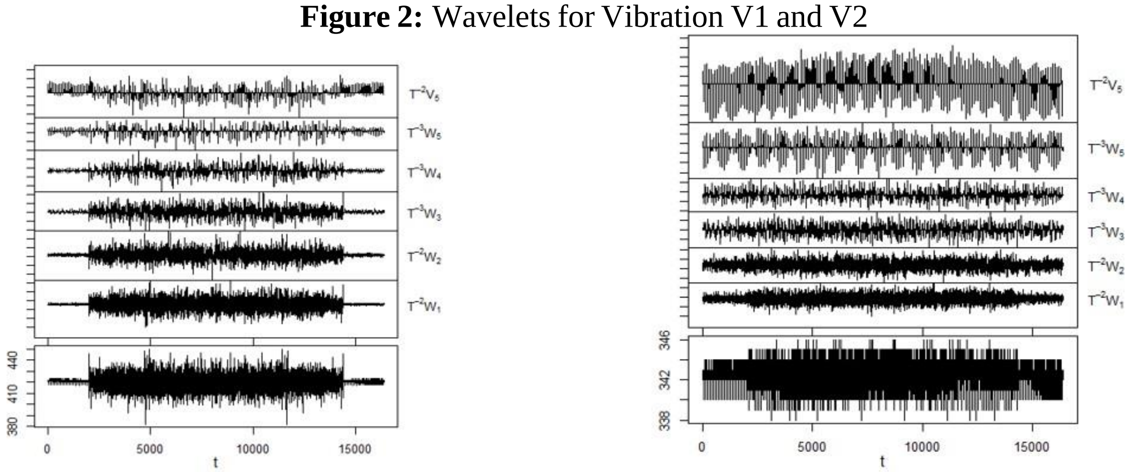

Figure 3: Wavelets for Vibration V3 and V4

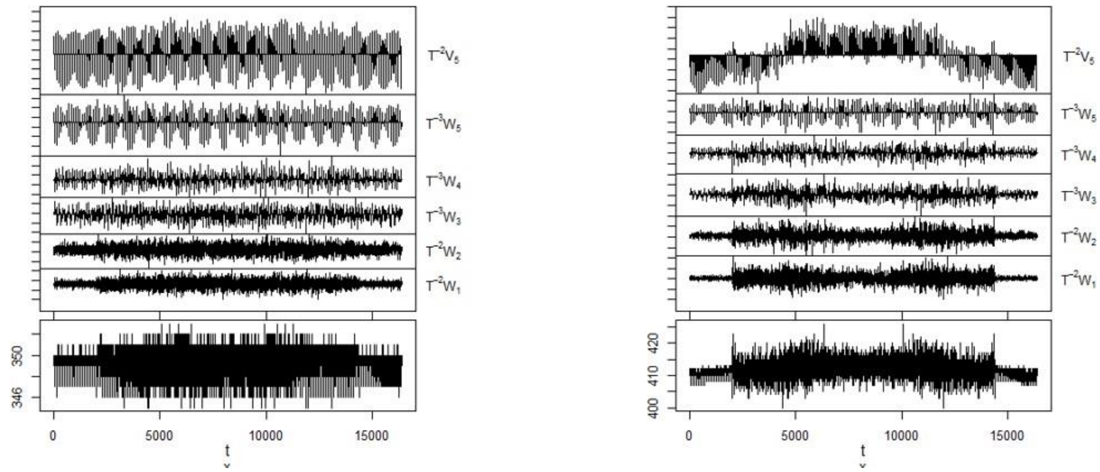

Figure 4: Wavelets for Vibration V5 and V6

B) Feature Extraction using Boruta Algorithm (Urjil and Palash):

Boruta is an algorithm for feature selection. It functions as a wrapper algorithm for Random Forest. The steps involved in the operation of the Boruta algorithm are as follows: To begin, it randomizes the provided data set by making scrambled duplicates of all features (which are called shadow features). Then, using the expanded data set, it trains a random forest classifier and applies a feature importance metric (the default is Mean Decrease Accuracy) to evaluate the value of each feature, with higher being more significant. It analyzes if a genuine feature is more important than the best of its shadow features at each iteration (i.e., whether the feature has a higher Z score than the highest Z score of its shadow features) and continually eliminates features which are deemed highly unimportant. Finally, the algorithm terminates either when all features are validated or rejected, or when it achieves a certain number of random forest runs.

i) Features Extraction of vibration data:

Boruta performed 199 iterations in 4.903793 secs.
1 attributes confirmed important: V6;
4 attributes confirmed unimportant: V1, V2, V3, V4;
1 tentative attributes left: V5;

Tentative Fix
Boruta performed 199 iterations in 4.903793 secs.
Tentatives rough fixed over the last 199 iterations.
2 attributes confirmed important: V5, V6;
4 attributes confirmed unimportant: V1, V2, V3, V4;

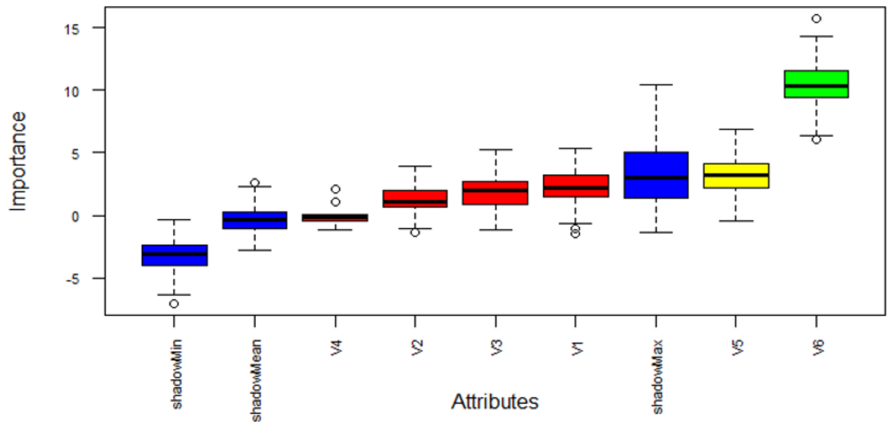

Figure 5: Variable selected Boxplot for vibration

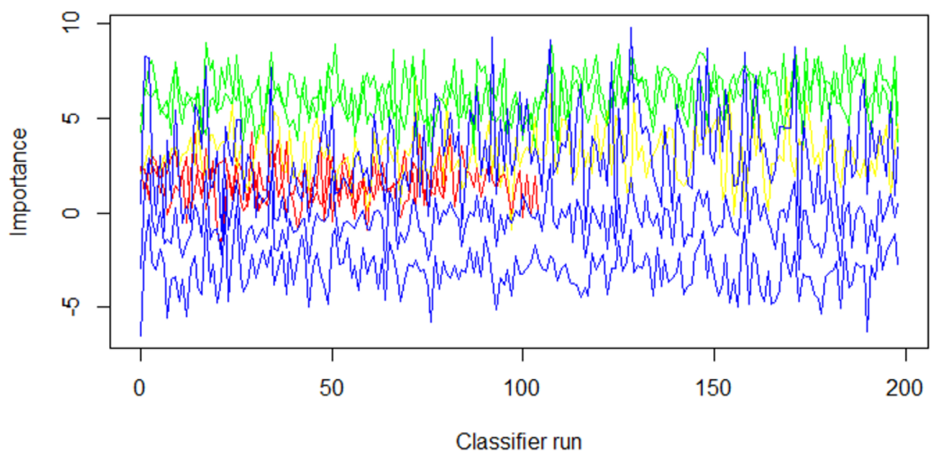

Figure 6: Variable selected for vibration

ii) Feature Extraction of temperature data:

Boruta performed 199 iterations in 4.749274 secs.
2 attributes confirmed important: T1, T3;
2 attributes confirmed unimportant: sr, T4;
1 tentative attributes left: T2;

Tentative Fix
Boruta performed 199 iterations in 4.749274 secs.
Tentatives roughfixed over the last 199 iterations.
3 attributes confirmed important: T1, T2, T3;
2 attributes confirmed unimportant: sr, T4;

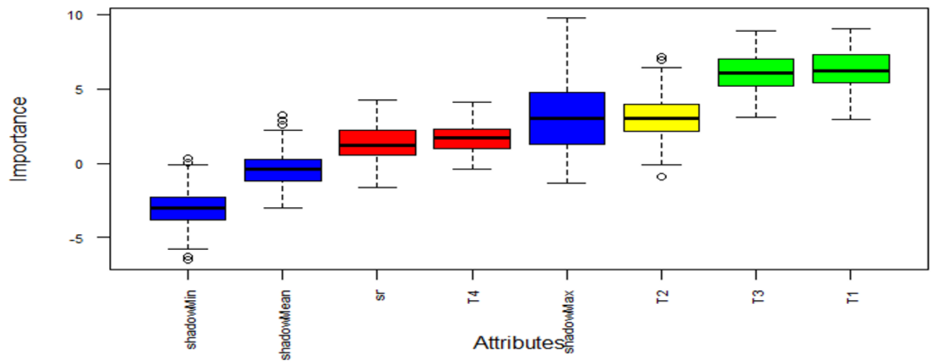

Figure 7: Variable selected Boxplot for Temperature

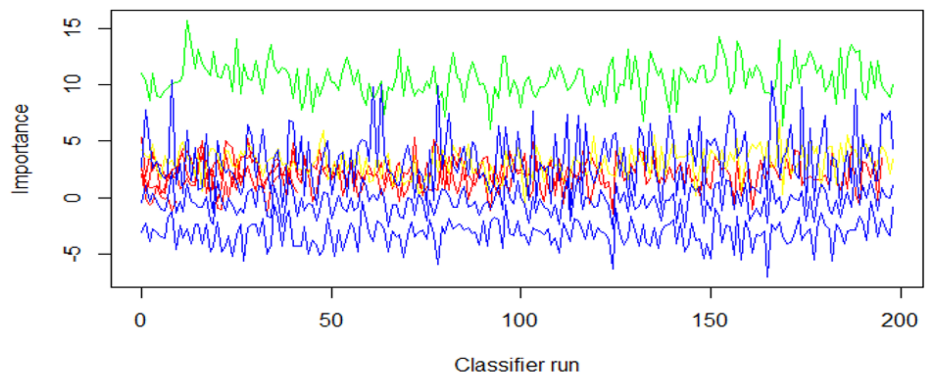

Figure 8: Variable selected for Temperature

iii) **Variable Selection by Lasso:** Lasso will eliminate many features and reduce overfitting in your linear model. RMSE comes out to be : 1.257071.

We can observe that as per Boruta and Lasso the thermocouple features T1, T2, T3 and Vibration features V1 andare the most influential on the dependent variable i.e 'roughness'. Adding to it the RMSE values are showcased as well.

We can observe that as per RIDGE the thermocouple features T1, T2, T3 and Vibration features V1 and V6 are themost influential on the dependent variable i.e 'roughness'. Adding to it the RMSE values are showcased as well.

iv) **Variable Selection by Ridge:** Ridge will reduce the impact of features that are not important in predicting your y values. RMSE comes out to be1.35976

:

v) **Variable Selection by Elastic**: Elastic Net combines feature elimination from Lasso and feature coefficient reduction from the Ridge model to prove your model's predictions. RMSE comes out to be : 1.32852. We can observe that as per Boruta and Lasso the thermocouple features T1, T2, T3 and Vibration features V1,V4 V6 are the most influential on the dependent variable i.e 'roughness'. Adding to it the RMSE values are showcased well. Depicted is the plot showing cross – validated Mean square error for lasso with is the most optimum regression our case. The value of Lambda is here chosen by cross validation. As lambda increases, shrinkage occurs so that variables that are at zero can be thrown away. So, a major advantage of lasso is that it is a combination of both shrinkage and selection of variables. In cases with very large number of features, lasso allow us to efficiently find the sparse model that involve a small subset of the features

vi) **Selection of best lamdba**:

0.257575 : Elastic net.

1.229339 : Ridge regression.

0.06118449 : Lasso regression.

We can conclude that by performing the LASSO regression we get the better value of tuning parameter

5. Conclusion:

In a nutshell, T1, T2, T3 and V1, V4, V6 are the variables that have the most influence on the product's manufacture. These factors influence the product's roughness, which is a dependent variable. The Boruta Algorithm and Feature Extraction Using Functional Regression were the techniques utilized to determine the variables.

6. Future Direction:

We can take manufacturing time series analysis into consideration for further work. For instance, how the roughness responds to the passage of time. Aside from this, we can concentrate on determining the ideal temperature and vibrations of each sensor where the quality of the roughness is quite high.

References:

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3092303/pdf/nihms248823.pdf
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7219245/
- https://www.sciencedirect.com/science/article/pii/S1566253519303902