Volume 6 Issue 1 (March 2026)
DOI 10.24191/jaeds.v6i1.167
eISSN 2805-5756

Journal of Applied Engineering Design and Simulation

Prediction of Density and Surface Roughness in LPBF-Printed Parts from Recycled SS316L Powder Using Random Forest Regression Model

Yusuf Olanrewaju Busari[1,2], Muhammad Asmadi Abdunkarim Yakoh[1], Mansir Abubakar [3, 4]*
Adel Mohammed Al-Dhahebi[1], Ajibike Joan Farounbi[5]
[1] Faculty of Mechanical Engineering & Smart Manufacturing Research Institute (SMRI), Universiti Teknologi MARA, 40450, Shah Alam, Malaysia
[2] Materials & Metallurgical Engineering Department, University of Ilorin, PMB 1515, Nigeria
[3] Faculty of Computer Science and Mathematics, Universiti Teknologi MARA,40450, Shah Alam, Malaysia
[4] College of Computing and Information Science, Al-Qalam University Katsina, Nigeria
[5] Applied Physics and Material Science Department, Alabama Agricultural and Mechanical Engineering University, Huntsville, 35899, AL, USA
*corresponding author: yusufbusari@uitm.edu.my

ABSTRACT

This paper uses machine learning algorithm to predict the density and surface roughness of 316L stainless steel parts manufactured using recycled powder based on the process parameter and powder characteristics. The advancement in Laser Powder Bed Fusion (LPBF) has enabled the production of complex and high-performance metallic components. However, the high cost of virgin powders and the substantial material waste generated during the AM process present economic and environmental challenges. The developed models with dedicated system interface were built on the understanding of the effect of powder characteristics on the part properties, which included layer thickness, hatch spacing, laser power, and scanning speed. Feature relationships were analysed using a correlation heatmap, highlighting strong interdependencies such as the inverse correlation between density and surface roughness (R = 0.98), and the alignment between laser power and scanning speed (R = 0.74). The RFR model was trained on datasets of varying sizes, and its performance was evaluated using standard error metrics (MAE, MSE, RMSE) and the coefficient of determination (R^2). The model achieved high predictive accuracy, with an R^2 of 0.821 for density and 0.795 for surface roughness from the initial 16 dataset. Error metrics were significantly lower than previous studies: MAE of 0.218 for density and 0.256 for surface roughness. Performance improved with larger datasets, reaching an R^2 of 0.973 for density and 0.942 for roughness at 250 samples, though a slight drop in accuracy was observed beyond this point due to potential data noise. The Random Forest model demonstrated strong capability in predicting quality outcomes in LPBF processes, outperforming earlier works in both accuracy and consistency. The developed system provides a model tool to inform AM optimization effectively, especially when supported by carefully selected features and appropriate dataset sizes.

Keywords: Density; Laser Powder Bed Fusion; Random Forest Regression (RFR); Surface roughness

1.0 INTRODUCTION

Laser powder bed fusion (LPBF) is the most common metal additive manufacturing (AM) technique which offers innovative design freedom and enables the creation of complex geometries. However, it has limitations, including large volume filling, temperature gradients, and potential defects in Ni-based super-alloys [1][2]. Powder recycling is cost-effective and has been studied for stainless steel powder [3]. While the use of recycled powder in LPBF presents challenges, such as changes in powder characteristics and potential reductions properties, it also offers significant economic and environmental benefits. The ability to reuse powder efficiently can reduce material costs and waste, contributing to the sustainability of additive manufacturing processes. However, careful control of powder characteristics and printing parameters is essential to maintain the quality and density of the final components. The density of components produced by LPBF using recycled metal powder is influenced by several key factors, including changes in powder characteristics due to recycling, the particle size distribution, and the printing parameters. These factors collectively impact the powder bed density and the final part density, which are crucial for achieving high-quality LPBF components. Recycled powders often exhibit changes in particle size and morphology, such as increased particle size and irregular shapes due to fragmentation and spatter formation. These changes can affect the flowability and packing density of the powder, which in turn influences the powder bed density and the final part density [4][5][6].

Recycled powders tend to have increased oxygen and other gas contents, which can lead to the formation of non-metallic inclusions and affect the microstructure of the printed parts. This can result in larger voids and a

Received on 11.02.2026
Accepted on 29.03.2026
Published on 30.03.2026

higher inclusion content, potentially reducing the density of the final components [7]. The surface roughness of parts is also affected by the particle size distribution of the powder and the processing parameters such as laser scan velocity and layer thickness. Optimized powder materials and subsequent blasting processes can improve surface roughness [8]. Also, the repeated exposure to high temperatures during the LPBF process can degrade the powder, affecting its rheological properties and leading to a decrease in part density over successive recycling cycles [9]. Recycled powder content of up to 60% can ensure a stable process and satisfactory part quality, while higher proportions may lead to defects and reduced part density [10]. Adjusting LPBF parameters, such as laser power and scanning speed, can mitigate some of the negative effects of using recycled powder. For instance, achieving a relative density of 99.2% is possible by optimizing these parameters, even with recycled powder [11] [12].

Machine learning models have been successfully used to predict surface roughness in various contexts. For instance, models like Random Forest and Support Vector Regression (SVR), Artificial Neural Networks (ANN) have shown robust performance in predicting surface roughness of additively manufactured parts [13][14]. These models can be trained on data from recycled powder to predict and optimize through hyperparameter tuning and cross-validation to enhance predictive accuracy. These models reduce the need for extensive physical testing, leading to cost savings and efficient material research. Machine learning (ML) models have been increasingly employed to model and predict the influence of these variables on manufacturing outcomes [15]. In recent studies, algorithms such as Random Forest Regression (RFR), SVR, ANN, and Gradient Boosting Regression (GBR) have been explored for this purpose. SVR has demonstrated strong correlation in surface roughness prediction ($R^2 \approx 0.85$), though with relatively higher error metrics like RMSE > 1.4. ANN has shown moderate prediction ability ($R^2 \approx 0.60$) but requires larger datasets and tuning to avoid overfitting. K-Nearest Neighbours (KNN) has yielded poor generalization ($R^2 \approx 0.32$) due to its sensitivity to data distribution [16]. In contrast, GBR often achieves high accuracy ($R^2 > 0.90$), but at the cost of complex hyperparameter tuning and higher computational demand. Among these, RFR offers an optimal trade-off between performance, simplicity, and interpretability especially effective when working with small to medium datasets and limited computational resources [17][18].

Although recycled metal powders have great benefits, its also bring important challenges that relate to quality, safety, and performance. Recycled powders are associated lower part density and increased surface roughness, which in turn may lead to challenges with the reliability and mechanical strength of finished printed components [19]. These limitations have slowed the broader industrial adoption of recycled powders in AM workflows. This study is conducted to develop a predictive machine learning framework that optimizes L-PBF processing conditions when using recycled metal powders. The focus is to identify critical process parameters such as laser power, scanning speed and layer thickness that significantly influence density and surface roughness. The evaluation of the accuracy and robustness of the RF model while comparing its performance with other algorithms in various studies will create efficient and intelligent manufacturing strategies by leveraging machine learning for quality prediction and process optimization in metal additive manufacturing.

2.0 METHODOLOGY

The RFR model is used to predict two important quality measurements density (%) and surface roughness (Ra, μm) of metal parts produced by powder bed fusion (PBF) additive manufacturing. The RFR model is built using the random forest regressor algorithm of the scikit-learn library, a popular machine learning toolkit written in python and well-known for its flexibility, robustness, and capability to work with scientific computing workflows as presented in Figure 1. The RFR model is selected because it is effective at working with high-dimensional, non-linear datasets without the problem of overfitting, as ensemble learning is applied**.** Prior to the development of the empirical model, the range of variables to describe the dependent variables of part density and surface roughness are analysed in an exploration plot. This plot is completed to determine the strength of the relationship between the independent variables of the measured powder characteristics and the dependent variable.

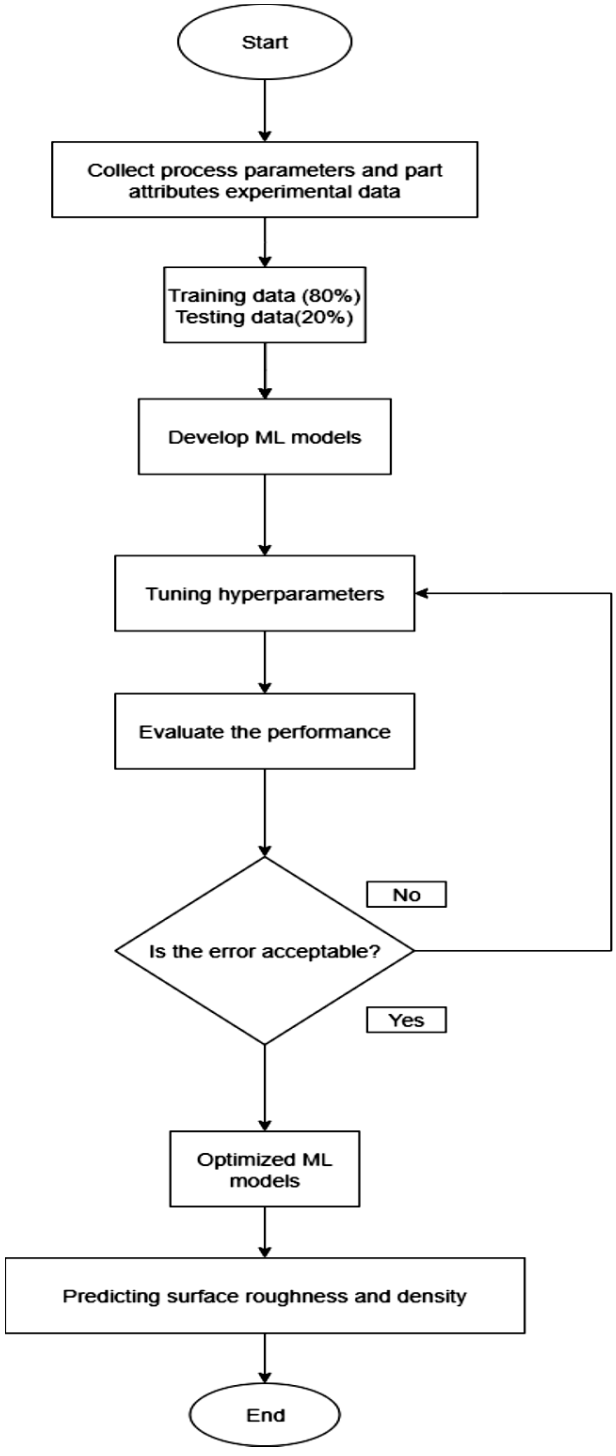


Figure 1. Random Forest Regressor Flowchart

The initial phase involves the assessment of multiple linear regression (MLR) model to guarantee that the accuracy of the model is improved by the inclusion of the specified independent variables. The test statistics are used to calculate the p-value and the inclusion of the variable is statistically significant if the p-value is less than 0.05. The coefficient of determination (R^2) is implemented to evaluate the degree to which the model generated accounts for the dependent variable's variability after all variables are determined to be significant. A higher R^2 value indicates that the model explains a larger proportion of the variation in the dependent variable. It is imperative to validate the model that has been generated to ensure that it accurately and consistently depicts the relationship.

The original measured values are compared to the projected values of the model. The residuals and their interaction with the independent variables employed are the primary focus of the model's testing. The analysis of four interactions is required for these tests, as outlined below. The residuals are plotted against the anticipated values to demonstrate a normal distribution. The residuals exhibit a normal distribution. The residuals do not exhibit any discernible patterns when plotted against the input variables (powder properties in this instance). A graph of the residuals and the prospective input features that are excluded from the model does not indicate any discernible trend. The prerequisites for the validity of an MLR model have been satisfied upon the confirmation of all requirements. The precision of the constructed model can be ascertained by examining the maximal residual generated.

2.1 Data Collection And Preparation

There are significant interaction effects of process parameters regarding surface roughness and relative density [20]. Furthermore, Cr and Si have higher oxygen affinity [21]. The dataset used is presented in Table 1 which initially composed of only 16 observations, each containing 10 relevant features and 2 target variables: density (%) and surface roughness (Ra, μm). Due to the limited size, which is insufficient for training robust machine learning models, data synthesis techniques are applied to expand the dataset to 500 synthetic observations. The dataset generation is conducted using resampling methods, specifically bootstrapping and interpolation methods that generate new synthetic data points based on the statistical patterns and distributions in the original experimental dataset using python to improve model generalization and performance [22].

Table 1: Related Process Parameter dataset

PARAMETER										
Laser Power (W)	Scanning Speed (mm/s)	Chemical Composition (O/N)	Particle Size (μm)	Layer Thickness (μm)	Hatch Spacing (μm)	No of Cycles	Powder Morphology	Density (%)	Surface Roughness (Ra, μm)	Ref
165	950	O+ N+	23	20	70	2	Spherical	98.5	4.2	[23]
180	850	O+ N+	23	20	80	15	Irregular	99.0	3.8	[23]
195	1000	O+ N+	32	20	60	30	Spherical	99.2	3.5	[23]
160	700	O+ N-	30	90	30	25	Spherical	96.5	5.5	[9]
175	800	O+ N-	25	20	80	30	Irregular	98.8	4.0	[24]
165	900	O- N+	35	20	70	10	Spherical	99.1	3.7	[25]
170	750	O- N+	27	25	85	10	Spherical	98.9	4.1	
195	950	O+ N+	29	20	75	30	Spherical	99.3	3.6	
195	1000	O- N-	31	25	80	1	Irregular	97.5	4.8	[26]
160	700	O- N+	28	22	80	2	Irregular	96.8	5.0	
60	500	O- N-	24	18	50	3	Spherical	97.2	4.9	
165	950	O+ N-	30	30	50	15	Irregular	95.9	6.0	[7]
150	800	O+ N+	26	22	80	5	Spherical	99.0	3.2	[27]
150	800	O+ N+	27	30	80	5	Spherical	99.1	3.0	[28]
195	1000	O- N-	32	25	70	20	Irregular	95.5	5.8	[29]
90	400	O+ N+	30	30	70	5	Irregular	96.0	6.2	

Following data generation, initial preprocessing steps are performed. These include the removal of unnamed or index columns and the encoding of categorical features. Specifically, chemical composition and powder morphology are numerically encoded using dictionary-based mapping, ensuring compatibility with machine learning workflows. The missing values in numerical columns is addressed with mean imputation to replace any missing entries with the average value of the respective feature. This approach is particularly important for maintaining data integrity, especially given the mixed nature of real and synthesized data [30]. The complete feature matrix (X) is extracted by dropping the two target columns. The targets are individually defined as y-density for density prediction and y-roughness for surface roughness prediction. The structured and cleaned dataset forms the foundation for model training, validation, and deployment in subsequent steps.

2.2 Library Importation

The project begins by importing essential Python libraries. Pandas and numpy are used for data manipulation due to their powerful and efficient handling of tabular and numerical data. Seaborn and matplotlib.pyplot are selected for data visualization because they offer flexible plotting capabilities, which are crucial for creating correlation heatmaps and scatter plots. For machine learning modeling, the scikit-learn library is utilised, specifically modules for training, model evaluation, and hyperparameter tuning. Finally, tkinter is chosen to develop the graphical user interface (GUI) because it is lightweight, built-in to Python, and well-suited for creating simple yet effective desktop applications without the need for external packages [31] .

2.3 Dataset Loading and Preprocessing

The initial dataset has just 16 samples, which is insufficient for training machine learning models. The data synthesis is utilised to expand the dataset to 500 rows, hence ensuring enhanced representation and learning effectiveness. After the importation the Excel file using pandas, superfluous index columns are removed to avoid redundancy. Categorical data, such as chemical composition and particle shape, are transformed into numerical values by dictionary mapping. This technique is chosen because to its simplicity, maintenance of category uniqueness and ability to facilitate successful interpretation of categorical inputs by the model. The missing values are addressed by mean imputation, a common technique wherein absent entries are replaced with the average value of the corresponding attribute. This ensures data integrity and preserves consistency across the dataset, which is especially crucial when managing a restricted or synthetic dataset as seen in Figure 2 [32] .

```
df = pd.read_excel('500 dataset.xlsx')
df = df.loc[:, ~df.columns.str.contains('^Unnamed')]

# Display the first few rows to check
print(df.head())

# Encode categorical data
chemical_mapping = {val: idx for idx, val in enumerate(df['Chemical Composition'].unique())}
morphology_mapping = {val: idx for idx, val in enumerate(df['Powder Morphology'].unique())}
df['Chemical Composition'] = df['Chemical Composition'].map(chemical_mapping)
df['Powder Morphology'] = df['Powder Morphology'].map(morphology_mapping)

# Handle missing values in the dataset
df.fillna(df.mean(numeric_only=True), inplace=True)
```

Figure 2. Dataset Loading and Preprocessing in Python

2.4 Feature and Target Definition

The dataset is organized into features (X) and targets (y_density and y_roughness) as shown in Figure 3. This is important because supervised learning models must have defined inputs/outputs. The features are important processing parameters and material attributes which are well documented in the additive manufacturing community as they impact the final part quality or attributes. This is why two models are trained to predict density (%) and surface roughness (Ra, µm), because both could be optimized independently since each may respond differently to the same inputs.

```
# Features & Targets
X = df.drop(columns=['Density (%)', 'Surface Roughness (Ra, µm)'])
y_density = df['Density (%)']
y_roughness = df['Surface Roughness (Ra, µm)']
```

Figure 3. Feature and Target Definition in Python

2.5 Train-Test Splitting

An 80% training and 20% testing split is used to evaluate the model's ability to generalise new data, and to ensure that the model has a substantial amount of data to learn from while still having some to test its performance. In an effort to reduce performance leakage from model to model and to evaluate the models independently, this division is conducted for each target variable as depicted in Figure 4.

```
# Train-Test Split
X_train_d, X_test_d, y_train_d, y_test_d = train_test_split(X, y_density, test_size=0.2, random_state=42)
X_train_r, X_test_r, y_train_r, y_test_r = train_test_split(X, y_roughness, test_size=0.2, random_state=42)
```

Figure 4. Train-Test Splitting in Python

2.6 Hyperparameter Tuning with GridSearchCV

A RFR is chosen as the ML algorithm due to its robustness, ability to handle nonlinear relationships, and effectiveness with small to medium-sized datasets. It also reduces overfitting by aggregating multiple decision trees (bagging). To further improve model performance, a GridSearchCV method is applied with 3-fold cross-validation. This approach systematically tests different combinations of hyperparameters and selects the best configuration based on cross-validated R^2 score as shown in Figure 5. Parameters such as n_estimators, max_depth, min_samples_split, and min_samples_leaf are selected because they directly affect tree structure, model complexity, and overfitting control. Cross-validation ensures the model is not overly tailored to a specific data split, promoting better generalization [33].

```
# Hyperparameter grid
param_grid = {
    'n_estimators': [100, 200],
    'max_depth': [10, 20, None],
    'min_samples_split': [2],
    'min_samples_leaf': [1]
}

pipeline_rf = RandomForestRegressor(random_state=42)

# Grid Search for Density
grid_d = GridSearchCV(pipeline_rf, param_grid, cv=3, n_jobs=-1, scoring='r2')
grid_d.fit(X_train_d, y_train_d)

# Grid Search for Roughness
grid_r = GridSearchCV(pipeline_rf, param_grid, cv=3, n_jobs=-1, scoring='r2')
grid_r.fit(X_train_r, y_train_r)
```

Figure 5. Hyperparameter Tuning in Python

2.7 GUI Design Using Tkinter

The prediction model is user-friendly and accessible; thus, a graphical user interface (GUI) is created using Tkinter. Tkinter is selected due to its ease of implementation, is cross-platform, and is built-in, fully integrated with Python. The GUI has many features for users with or without programming experience to input process parameters manually and allows for the use of dropdown menus for selecting categorical variables. New users, without programming experience, can easily access the prediction model in a user-friendly way. The GUI allows the user to predict parameter combinations in real time, which provides flexibility for users in either research laboratories or in industrial settings.

2.8 Prediction Logic Implementation

The predict () function handles input validation, data formatting, and model inference. Helper functions (safe_float, safe_int) are used to ensure that all inputs are numeric and fall within acceptable range. Any missing input values are automatically filled with the mean from the training dataset, ensuring consistency with how missing values were handled during model training. The function then uses the trained Random Forest models to predict both density and surface roughness. The results are displayed with status messages, indicating whether the predictions meet industry-standard thresholds (e.g., density ≥ 99% and Ra between 0.73–4.19 μm) [34] [35]. This output design provides not only the prediction value but also its practical implication.

2.9 Correlation Heatmap Visualization

A correlation heatmap is included to provide deeper insight into the relationships between different process parameters and target properties. This heatmap is generated using Seaborn, which supports well-formatted annotations and color gradients. It is particularly useful for identifying redundant features or key contributors to target variables. This analysis also supports model interpretation and feature selection during development.

2.10 Evaluation and Result Visualization

The scatter diagrams are employed to compare the predicted and actual values of both objective variables in order to verify the predictive capacity of the model. The absolute error for each prediction is denoted by colour-coded error magnitudes, which facilitate the rapid identification of outliers. Figure 6 code represents quantitatively assess accuracy and consistency, the performance metrics R^2, MAE, and MSE are computed. These metrics are chosen because they are the standard in regression analysis and collectively provide a balanced perspective on the precision and reliability of the model [16] .

```
# Plot Actual vs Predicted
y_pred_density = grid_d.predict(X_test_d)
y_pred_roughness = grid_r.predict(X_test_r)
errors_density = np.abs(y_test_d - y_pred_density)
errors_roughness = np.abs(y_test_r - y_pred_roughness)

r2_density = r2_score(y_test_d, y_pred_density)
r2_roughness = r2_score(y_test_r, y_pred_roughness)
mae_density = mean_absolute_error(y_test_d, y_pred_density)
mae_roughness = mean_absolute_error(y_test_r, y_pred_roughness)
mse_density = mean_squared_error(y_test_d, y_pred_density)
mse_roughness = mean_squared_error(y_test_r, y_pred_roughness)
```

Figure 6. Evaluation and Result Visualization in Python

2.11 Application Execution

The last step of the implementation procedure would be launching of the GUI by issuing the command root.mainloop(). This starts up the event listener's main loop, wherein the GUI window remains "alive" or states receptive to user input intervened with several formats, such as when the user may enter input data or select or clicks a button. In a live time environment, registered end-users can simply execute their earlier trained Random Forest models and imagine how variations in the process parameters will affect the outcomes of part quality in a virtual setting without ever having to put these experiments to practice. The essence of the machine learning algorithms, interactive visualization concepts, and user interface side-by-side give the user an entirely working, stand-alone desktop application. This serves the engineers, technicians, and researchers in taking any input values of their choice and instantly getting predictive outputs of parameters important to their process output, like density (%) and surface roughness (Ra, μm). Furthermore, the app furnishes other options of user-friendly visualizations that include heat maps and scatter plots to facilitate deeper interpretation.

3.0 RESULTS AND DISCUSSION

3.1 Features Importance Analysis Based on Heatmap

The heatmap is a powerful visual aid in the Random Forest Prediction model, helping users to better visualize the relationships between many input parameters and output variables (i.e., density (%) and surface roughness, Ra (μm). Figure 7 is created using the Seaborn package which essentially displays the Pearson correlation coefficients. The Pearson correlation coefficient is a measure of correlation between linear relationships between variables. This feature is very useful to allow users to simply identify which parameters have a significant influence on part quality and how these parameters interact. For example, a heat map can display relationships that have a strong and negative correlation, such as density and surface roughness. As density increases, the surface finishes will be smoother. Alternatively, it could be suggested that laser power and scanning speed have a positive correlation as they are often adjusted together when determining energy input. Overall, the heatmap provides a visual representation of these relationships allowing the user to be guided in basis more diagnostic information and potentially optimize selecting challenges for successful user experience. Again, the heatmap can help reduce training or redundancy in model training and shed light on abnormalities since AM processes generally require specification of multiple interacting variable selection criteria such as temperature, scan strategy speed, and the use of active energy approaches. Thus overall, the heatmap is a distilling process for data driven decisions in allowing complex model behaviour to be distilled into a simple and straightforward format that the user can understand and apply within the context of real-life scenarios.

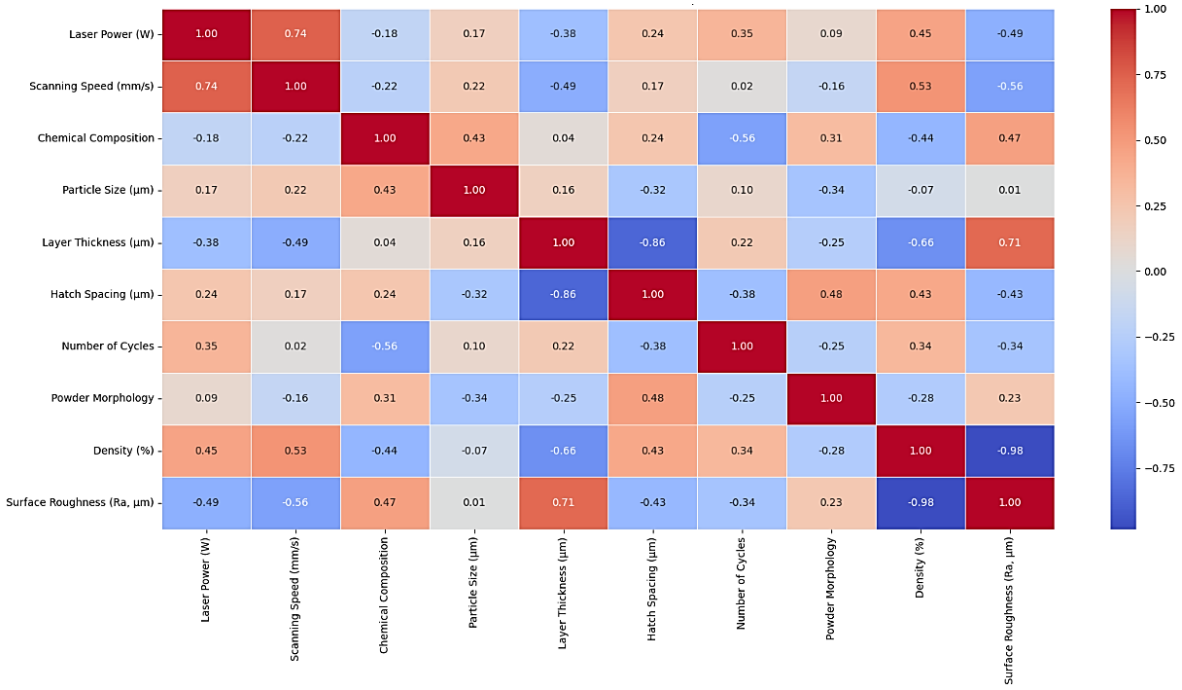


Figure 7. Feature Correlation Heatmap

The feature correlation heatmap offers a comprehensive understanding of the correlations between ten critical process parameters and material properties in LPBF processes. The Pearson correlation coefficient is used to identify strong and moderate linear dependencies of interest in the heatmap. This is particularly important when selecting features to enhance the operationalisation of RFR. One example of a strong negative correlation (R = -0.98) between a parts density and its surface roughness is that as the parts density increases, the surface roughness decreases. The strong negative correlation observed between relative density and surface roughness can be explained by the fundamental physics of the powder bed fusion (PBF) process. In PBF additive manufacturing, process parameters such as laser power, scanning speed, hatch spacing, and layer thickness strongly influence the melt pool stability and powder consolidation. When sufficient energy is delivered to the powder bed, complete melting and proper fusion of particles occur, which promotes higher densification and reduced porosity. This improved consolidation also produces a smoother and more uniform surface morphology, resulting in lower surface roughness values. Conversely, when the energy input is insufficient or process conditions are not optimal, incomplete melting and lack ofbfusion defects may occur. These defects create internal pores and irregular surface features, leading to reduced density and increased surface roughness. Additionally, partially melted particles and balling phenomena may adhere to the surface, further increasing roughness. Therefore, improved densification generally corresponds to smoother surface formation, which explains the observed strong inverse relationship between density and surface roughness.

Furthermore, it suggests that the denser the component, the more laminar behaviour occurs during material fusion. Although RFR is capable of managing redundant features concurrently, the inclusion of both features in larger RFR implementations would impede interpretation. Furthermore, a second robust negative correlation (R = -0.86) exists between layer thickness and hatch spacing, as they are frequently adjusted in inverse proportions as a result of heat input management. The strong positive correlation (R=0.74) between laser power and laser scanning speed indicates that practical operation continuity is possible for overheating and/or underheated components that consistently contribute to heat input. Moderate correlations are also beneficial for expression and ongoing observations. Benefitful insight regarding the interpretation of results is also provided by strong and moderate correlations. For instance, the surface roughness of manufactured parts is improved by higher energy inputs, which results in faster scanning velocities (R = -0.56). Similarly, the surface polish of manufactured parts is improved by higher laser power (R = -0.49). Other moderate correlations, such as the relationship between the number of cycles and the material (i.e., chemical composition) (R=-0.56), may be indicative of thermal degradation or material degradation during continued reuse.

Some characteristics, such as particle size and powder morphology, exhibit minimal direct correlations; however, they may still be relevant through non-linear interactions that RFR can detect. In general, the heatmap advocates for the selection of high-impact features, such as laser power, scanning speed, layer thickness, and either density or surface texture (but not both), in order to enhance the predictive accuracy and model efficiency.

3.2 Features Importance Analysis Based on Heatmap

The image represents a graphical user interface designed to implement the Random Forest Prediction model, which is used for predicting the quality of parts in metal additive manufacturing, specifically during LPBF. It provides users with entry fields to fill in process and material parameters, such as laser power, scanning speed, particle size, layer thickness, hatch spacing, number of cycles, chemical composition, and powder morphology. These percentages are then instantly predicted by the machine learning model fed by historical and synthetic data.

Figure 8 (b) highlights selections, such as 165 W laser power, 1000 mm/s scanning speed, 20 µm particle size, 30 µm layer thickness, and spherical powder morphology. On implementation, this leads to a prediction of the part density of 98.59%, which lies a little below the target of 99% and is hence marked as a "Low Density". Conversley, the surface roughness of 4.13-micron value, as predicted, is within the acceptable range of 0.73 to 4.19 and therefore labelled as "Good Surface Roughness". The tool predicts and is very useful toward the real-time optimization of process conditions.

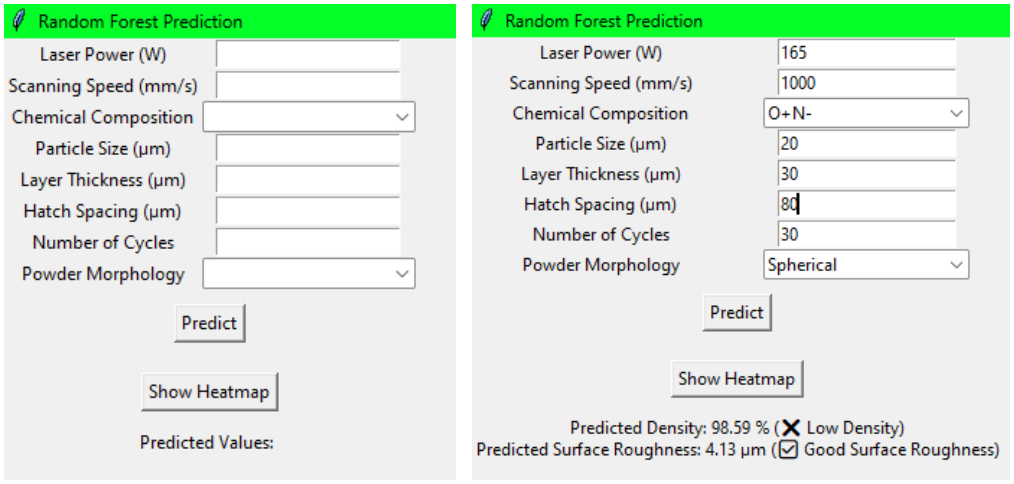


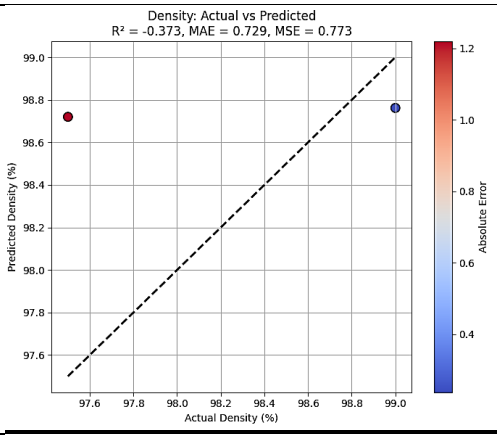
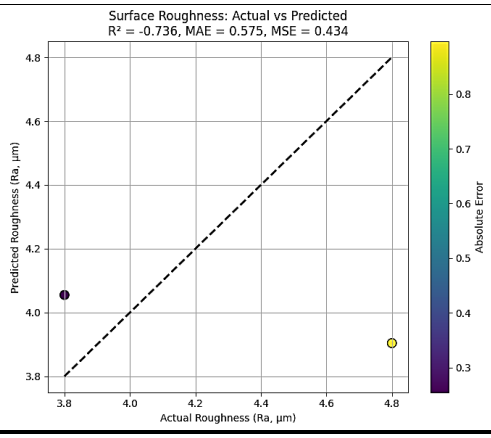
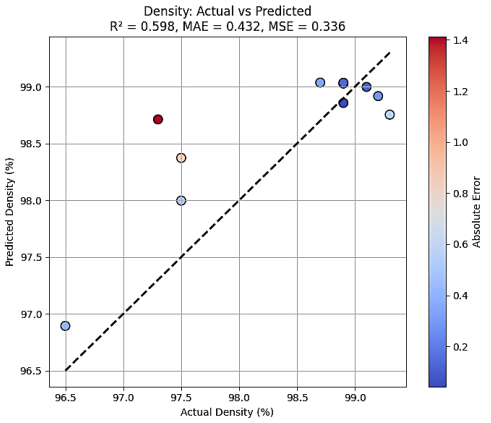
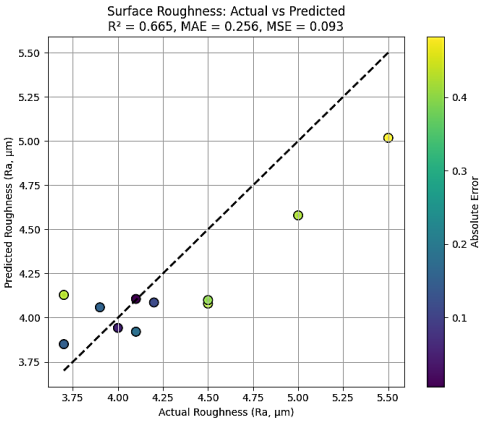
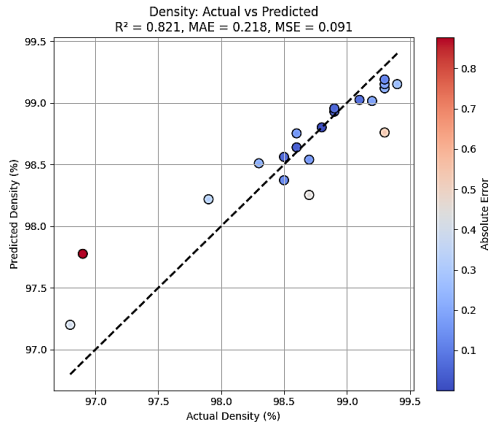
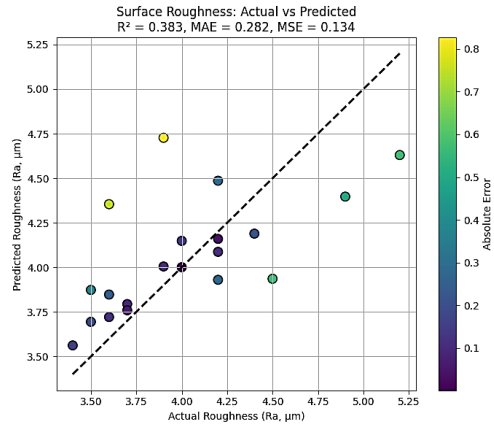
Figure 8. The GUI from the correlation plot

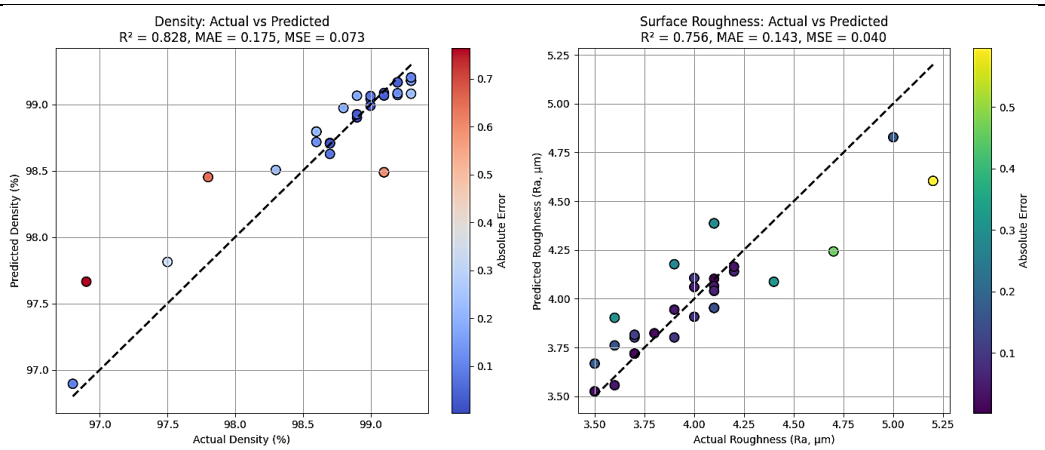
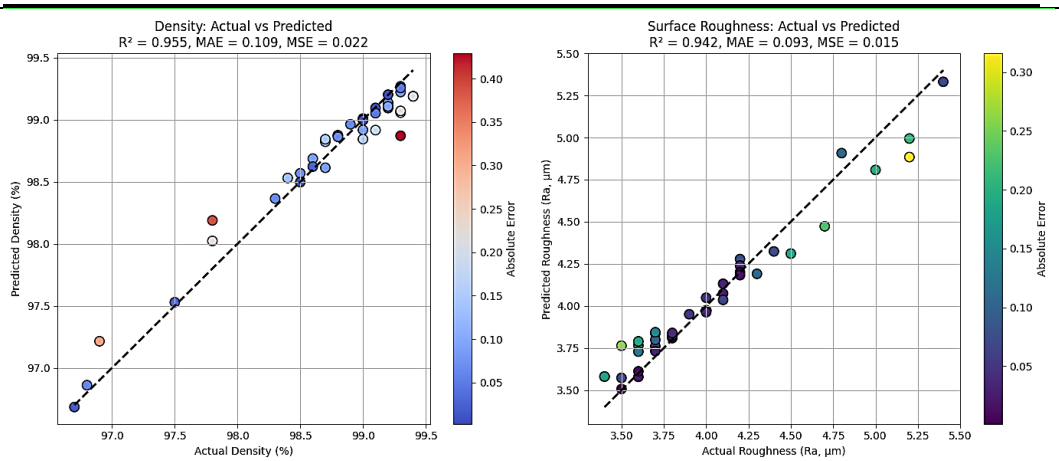
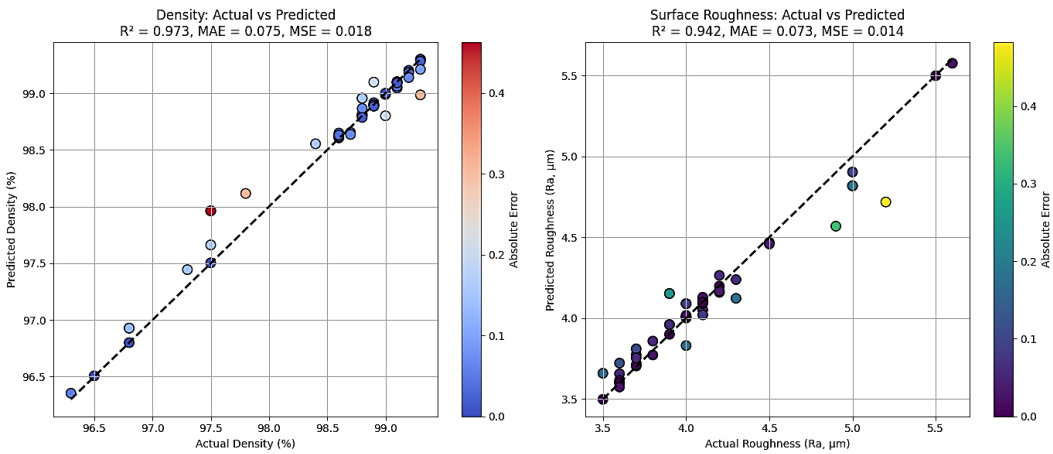
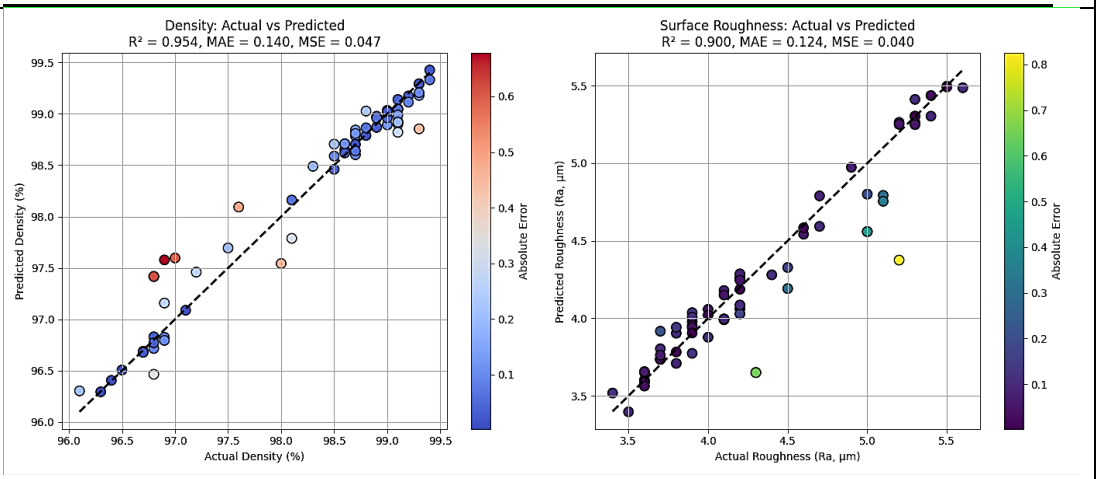
3.3 Result Comparison Between Different Number of Dataset

The RFR model was evaluated on its performance for the density and surface roughness prediction based on the dataset size varying from 10 to 500 in Table 2. At 10 samples, the model lost the most predictive capabilities showing negative R^2 values for both outputs, suggesting extreme underfitting. The model began to show improved performance at 50 samples with R^2 outputs of = 0.598 for density and 0.665 for surface roughness with MAE and MSE decreasing as well. When there were 100 samples, the density prediction began to develop more predictive capabilities (R^2 = 0.821); however, surface roughness dropped to R^2 = 0.383, suggesting that the model could not make sense of everything due to too many dimensions for the possible outputs or images.

The model performance stabilized at 150 samples where it reached R^2 of 0.828 and 0.756 for density and surface roughness, respectively. The development of lower MAE and MSE showed that the model reduced the errors visually as well. The performance peaked and was stable at 250 samples when the R^2 dropped to 0.973 for density and 0.942 for surface roughness respectively. Both error metrics also appeared as their lowest measures. It can be deduced that the model performed at its best predictive capabilities and generalization capabilities with 250 samples. The model performance decline with a dataset size of 500 samples is attributed to noise and inconsistencies in the synthetic data rather than overfitting. The optimal dataset size for accurate predictions of density and surface roughness was found to be 250 data points. The slight performance drop is linked to variability from bootstrapping and interpolation techniques used in data generation. The Random Forest Regressor mitigates overfitting risk through ensemble learning, and hyperparameter tuning with GridSearchCV further aids in robust model performance. Thus, the changes in performance reflect the variability from synthetic data rather than model overfitting.

Table 2: Result Comparison Between Different Number of Dataset

No. Data	Density	Surface Roughness
10		
50		
100		

150	
200	
250	
500	

3.4 Comparison with Existing Study for Density

The RFR model was evaluated on its performance predicting the density of parts produced by Powder Bed Fusion Additive Manufacturing. To check the robustness and accuracy of the constructed model, the performances were compared against the machine-learning models noted in Barrionuevo et al.'s work which used various machine learning algorithms, such as ANN, SVM, K-Nearest Neighbours KNN and Random Forest for similar outputs. The article used a dataset of approximately 100 sample parts in which both studies were trained and evaluated [29]. A detailed comparison of performance metrics such as R^2, MAE and RMSE is presented in Table 3.

Table 3: Performance Comparison Between Present Study and Existing Literature

Model	Study	R^2	MAE	RMSE	Remarks
Random Forest	This Study	0.821	0.218	0.3067	High accuracy with low error metrics
Random Forest	Barrionuevo et al. (2023)	0.5198	0.443	0.4443	High R^2 but significantly higher errors compared to this study
KNN	Barrionuevo et al. (2023)	0.3225	0.5216	0.3225	Moderate accuracy and error
SVM	Barrionuevo et al. (2023)	0.2529	0.7566	0.4319	High R^2, but less stable in error metrics

The Random Forest model created in this study reported an R^2 value of 0.821, which is a relatively high R^2 value, demonstrating a strong association between the predicted and actual density values. This value indicates that the model explains approximately 82.1% of the variability in the data set, demonstrating the model's capability of capturing complex relationships between the input features. If we compare the stated R^2 value of 0.5198 noted by Barrionuevo et al. (2023), it is clear that this model has much better overall explanatory power and generalization performance.

The Random Forest model in this study had a MAE of 0.218, which is much better than the 0.443 MAE obtained for the comparison study. A lower MAE represents a more accurate model because it resulted in smaller average errors in part density predictions, which is paramount when aiming to provide consistent and reliable predictions. In additive manufacturing, even small reductions in quality parameters can mean big differences in the part performance.

The RMSE, which provides an interpretable error measure in the same units as the target variable, was calculated to be 0.3067 in this study. Although slightly higher than initially reported, this RMSE is still substantially lower than the 0.4443 observed in the comparative study. A lower RMSE indicates fewer large prediction errors and a generally well-performing model. This result confirms that the Random Forest model developed in this project remains more reliable and accurate in predicting density outcomes, especially when applied to recycled powder-based L-PBF processes.

The Random Forest model produced from this project provides more excellent performance in predicting the relative density of components made from Powder Bed Fusion Additive Manufacturing than the Random Forest model by Barrionuevo et al. (2023). The model calculated better values, as shown by the lower values for all the metrics (R^2, MAE, RMSE, and MSE), which means that this model predictions are better made and more robust. Specifically, the considerably lower error metrics for the model (MAE, RMSE, MSE) mean that the model is a much more reliable option for density and relative density predictions in an industrial context.

The improved performance can largely be attributed to the improved methods and optimizations utilized to model the data, such as feature engineering, hyperparameters, tuning and ensemble methods. The previous study of Barrionuevo et al. (2023) showed that the Random Forest modelling techniques from this study represent a more reliable and effective option for industrial applications involving optimizing the Powder Bed Fusion process and predicting the quality of the part.

3.5 Comparison with Existing Study for Surface Roughness

The RFR model was employed to predict surface roughness (Ra) in LPBF printed parts. The model was trained and evaluated using performance metrics including the R^2, MAE, MSE and RMSE. A dataset comprising of 50 samples was used in training and evaluation [14]. The effectiveness of the model was assessed using a performance comparison of Dejene et al. (2024) to highlight the accuracy and robustness of the model relative to an established benchmark in the literature as presented in Table 4.

Table 4: Performance Comparison

Model	Study	R^2	MAE	RME/ MSE	Remarks
Random Forest (RFR)	This Study	0.665	0.26	0.305 (RMSE)	Good accuracy with low error metrics; dataset size: 50
Random Forest (RFR)	Dejene et al. (2024)	0.97	0.51	0.66 (RMSE)	High R^2 but relatively higher error values
Support Vector Regression	Dejene et al. (2024)	0.85	0.99	1.41 (RMSE)	Moderate prediction performance for Ra

With the Random Forest model in the context, a 0.665 R^2 value was obtained with respect to surface roughness; hence, it had a moderate variance explanation within the data. Meanwhile, Dejene et al. recorded an R^2 of approximately 0.97 within their RFR model; the lower R^2 recorded in this study is mainly due to a much smaller data sample taken (50 samples only). Therefore, the limited observations restrict the scope of the model to capture and subsequently generalize to more complex. However, 0.665 R^2 is enough to denote a working model that can apply well in a scarce data situation.

The model achieved an MAE of 0.256, substantially smaller than the 0.51 MAE recorded by Dejene et al. This means that the average is extensively smaller between predicted and actual surface roughness values, thus indicating better overall accuracy. Hence, even with fewer data points, the model produces precise predictions for Ra and hence can be used in any application that demands uniformity in surface finish quality.

In RMSE terms, the model had an error of 0.305, while Dejene et al.'s model returned an error of 0.66. RMSE heavily weighs large deviations in prediction, which means the smaller the value below, the better the model is at controlling large deviations. This supports the findings that the model predicted more stable and accurate values, especially in controlling big outliers in the estimation of surface roughness, even though the overall R^2 value might be lower due to data limitation.

Although the R^2 in this study is lower, both MAE and RMSE show improved performance in error minimization. This suggests the model may be slightly underfitted due to fewer data samples but still offers high precision in the predictions it makes. The Random Forest Regression model developed in this study shows promising results for predicting surface roughness (Ra) in L-PBF printed parts, even when trained on a smaller dataset (50 samples). Compared to the model by Dejene et al. (2024), the model achieved lower MAE and RMSE, indicating higher prediction accuracy and fewer large errors, despite having a lower R^2 due to dataset size limitations. These findings highlight the effectiveness of proper preprocessing, feature engineering, and model tuning. While expanding the dataset in future work could improve R^2, the current model is already a valuable tool for quality control and surface optimization in additive manufacturing applications.

4.0 CONCLUSION

This work presented the use of the RFR model to predict quality indicators (Qis) such as density and surface roughness for L-PBF parts made from recycled metal powders. The structure of the model was meant to balance practicality for real-world use; it is easy-to-use, includes a rationally simple graphic user interface, and has effective data preprocessing and feature engineering techniques. With 250 samples, we found the model highly accurate across all dataset size combinations by achieving the lowest average MAE and MSE values and R^2 values of 0.973 for density and 0.942 for surface roughness, indicating a strong predictive constant and degree of model generalizability. Applying a correlation heatmap also helped to explain feature selection and model explainability by illustrating the relationships between some of the key input features, including a strong inverse correlation density with surface roughness (R = -0.98) and a positive correlation laser power and scanning speed (R = 0.74). While the model was accurate, the research was limited by the small size of the dataset, particularly in the first step of model evaluation where we observed underfitting when sample sizes were only as small as 10. The use of recycled powder data also raises the question of how generalizable the model will be across material types and machine settings since these data may not truly capture the full variability that would be seen in a true industrial reuse situation. However, there is a lot of promise for the advancement of sustainable additive manufacturing with the developed system. When using recycled powders, its real-time predictive capability integrated through a graphical user interface (GUI) offers industry professionals a useful tool to optimize process parameters, lower experimental costs, and improve part quality. Future work should expand the data set to include recycled powder condition, and powder composition, which allows the model to be relevant for a wider variety of conditions stemming from recycled powders. Further evaluating different methods, such as Gradient Boosted Trees or other Physics-Informed Neural Networks would also help to enhance the predictive competency of the data set. In addition, real-time monitoring and feedback for adaptive learning with sensors may produce a subsequently more dynamic value creation system applicable for real-life scenarios in manufacturing. Overall, this research provides the structural base for a data driven, sustainable process optimization model to best assist continuously rapid improvements of operational choices in additive manufacturing processes of metals.

ACKNOWLEDGEMENT
The authors wish to extend their sincere gratitude to the dedicated staff of Smart Manufacturing Research Institute (SMRI) Universiti Teknologi MARA (UiTM) Shah Alam for their invaluable support and the Department of Applied Physics and Material Science, Alabama Agricultural and Mechanical Engineering University, Huntsville, 35899, AL, USA.

AUTHORS CONTRIBUTION
Supervision; Y.O. Busari, Conceptualization; Y. O. Busari, , Data curation; M.A.A Yakoh, M. Abubakar, Formal analysis; A.M. Al-Dhahebi, Data collection; M.A.A Yakoh, M. Abubakar, , A .J. Farounbi, Validation, Writing - Original Draft, Writing - Review & Editing; Y.O. Busari , M. Abubakar.

DECLARATION OF COMPETING OF INTEREST
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES
[1] C. Guo *et al.*, 'Additive manufacturing of Ni-based superalloys: Residual stress, mechanisms of crack formation and strategies for crack inhibition', *Nano Materials Science*, no. June, 2022.
[2] L. A. Tochiro *et al.*, 'Powder bed fusion of high-Mn-N Ni-free austenitic stainless steel: achieving low porosity and high mechanical strength through process parameter selection', *International Journal of Advanced Manufacturing Technology*, vol. 131, no. 3–4, pp. 1377–1396, 2024.
[3] Y. O. Busari *et al.*, 'Evaluation of recycling strategies for SS316L powder in laser powder bed fusion: impacts on mechanical properties and fatigue performance', *International Journal of Advanced Manufacturing Technology*, vol. 139, no. 3, pp. 1123–1144, 2025.
[4] A. Lanzutti *et al.*, 'Effect of powder recycling on inclusion content and distribution in AISI 316L produced by L-PBF technique', *Journal of Materials Research and Technology*, vol. 23, pp. 3638–3650, 2023.
[5] S. Li, B. Chen, C. Tan, and X. Song, 'Study on Recyclability of 316L Stainless Steel Powder by Using Laser Directed Energy Deposition', *Journal of Materials Engineering and Performance*, vol. 31, no. 1, pp. 400–409, 2022.
[6] K. Riener, S. Oswald, M. Winkler, and G. J. Leichtfried, 'Influence of storage conditions and reconditioning of AlSi10Mg powder on the quality of parts produced by laser powder bed fusion (LPBF)', *Additive Manufacturing*, vol. 39, no. February, p. 101896, 2021.
[7] T. Delacroix, F. Lomello, F. Schuster, H. Maskrot, and J. P. Garandet, 'Influence of powder recycling on 316L stainless steel feedstocks and printed parts in laser powder bed fusion', *Additive Manufacturing*, vol. 50, no. September 2021, 2022.
[8] A. B. Spierings, N. Herres, and G. Levy, 'Influence of the particle size distribution on surface quality and mechanical properties in AM steel parts', *Rapid Prototyping Journal*, vol. 17, no. 3, pp. 195–202, 2011.
[9] R. Douglas, R. Lancaster, T. Jones, N. Barnard, and J. Adams, 'The Influence of Powder Reuse on the Properties of Laser Powder Bed-Fused Stainless Steel 316L: A Review', *Advanced Engineering Materials*, vol. 2200596, 2022.
[10] Q. Wu *et al.*, 'Numerical investigation on the reuse of recycled powders in powder bed fusion additive manufacturing', *Additive Manufacturing*, vol. 77, no. June, p. 103821, 2023.
[11] I. A. Pelevin *et al.*, 'New Scanning Strategy Approach for Laser Powder Bed Fusion of Nd-Fe-B Hard Magnetic Material', *Metals*, vol. 13, no. 6, 2023.
[12] I. A. Pelevin *et al.*, 'Laser Powder Bed Fusion of Chromium Bronze Using Recycled Powder', *materials*, vol. 14, no. 3644, pp. 1–11, 2021.
[13] F. Jáñez-Martino, E. Fidalgo, S. Martínez-Pellitero, S. Giganto, and E. Alegre, 'Estimation of arithmetical mean roughness using a regression approach and computer vision in stainless steel specimens manufactured by additive manufacturing', *International Journal of Computer Integrated Manufacturing*, vol. 0, no. 0, pp. 1–12, 2025.
[14] B. V. S. Reddy, A. M. Shaik, C. C. Sastry, J. Krishnaiah, S. Patil, and C. P. Nikhare, 'Performance evaluation of machine learning techniques in surface roughness prediction for 3D printed micro-lattice structures', *Journal of Manufacturing Processes*, vol. 137, no. November 2024, pp. 320–341, 2025.
[15] M. A. Equbal, A. Equbal, Z. A. Khan, and I. A. Badruddin, 'Machine learning in additive manufacturing: A comprehensive insight', *International Journal of Lightweight Materials and*

Manufacture, vol. 8, no. 2, pp. 264–284, 2025.
[16] F. Ahmed, U. Ali, D. Sarker, E. Marzbanrad, K. Choi, and Y. Mahmoodkhani, 'Study of powder recycling and its effect on printed parts during laser powder-bed fusion of 17-4 PH stainless steel', *Journal of Materials Processing Tech*, vol. 278, p. 116522, 2020.
[17] D. N. Aqilah, A. K. M. Sayuti, Y. Farazila, D. Y. Suleiman, M. A. N. Amirah, and W. B. W. N. Izzati, 'Effects of Process Parameters on the Surface Roughness of Stainless Steel 316L Parts Produced by Selective Laser Melting', *Journal of Testing and Evaluation*, vol. 46, no. 4, pp. 1673–1683, 2018.
[18] E. Toyserkani, D. Sarker, O. O. Ibhadode, F. Liravi, P. Russo, and K. Taherkhani, 'Monitoring and Quality Assurance for Metal Additive Manufacturing', in *Metal Additive Manufacturing*, Wiley & Sons Ltd, 2021.
[19] W. Abd-Elaziem, S. Elkatatny, T. A. Sebaey, M. A. Darwish, M. A. Abd El-Baky, and A. hamada, 'Machine learning for advancing laser powder bed fusion of stainless steel', *Journal of Materials Research and Technology*, vol. 30, no. April, pp. 4986–5016, 2024.
[20] N. Emminghaus, C. Hoff, J. Hermsdorf, and S. Kaierle, 'Residual oxygen content and powder recycling: Effects on surface roughness and porosity of additively manufactured Ti-6Al-4V', *Additive Manufacturing*, vol. 46, 2021.
[21] N. E. Gorji *et al.*, 'A new method for assessing the recyclability of powders within Powder Bed Fusion process', *Materials Characterization*, vol. 161, no. October 2019, p. 110167, 2020.
[22] G. Assunção, R. Izbicki, and M. Prates, 'Is Augmentation Effective in Improving Prediction in Imbalanced Datasets?', *Journal of Data Science*, pp. 1–16, Oct. 2024.
[23] R. Douglas, N. Barnard, N. Lavery, J. Sullivan, T. Jones, and R. Lancaster, 'The effect of powder recycling on the mechanical performance of laser powder bed fused stainless steel 316L', *Additive Manufacturing*, vol. 88, no. June, p. 104245, 2024.
[24] S. M. Yusuf, E. Choo, and N. Gao, 'Comparison between virgin and recycled 316l ss and alsi10mg powders used for laser powder bed fusion additive manufacturing', *Metals*, vol. 10, no. 12, pp. 1–18, 2020.
[25] C. Lu *et al.*, 'A comprehensive characterization of virgin and recycled 316L powders during laser powder bed fusion', *Journal of Materials Research and Technology*, vol. 18, pp. 2292--2309, 2022.
[26] G. Jacob, S. Watson, C. Brown, and S. Watson, 'Effects of powder recycling on stainless steel powder and built material properties in metal powder bed fusion processes NIST Advanced Manufacturing Series 100-6 Effects of powder recycling on stainless steel powder and built material properties in metal', *NIST Advanced Manufacturing Series 100-6*, p. 59, 2017.
[27] A. Gopaluni, C. Nayak, A. Piironen, T. Kantonen, and A. Salminen, 'Effects of powder recycling on laser-based powder bed fusion produced SS316L parts', in *NOLAMP- Nordic Laser Materials Processing Conference (19TH-NOLAMP-2023)*, 2023, vol. 1296, no. 1, p. 012021.
[28] N. D. Dejene, H. G. Lemu, and E. M. Gutema, 'Effects of process parameters on the surface characteristics of laser powder bed fusion printed parts: machine learning predictions with random forest and support vector regression', *International Journal of Advanced Manufacturing Technology*, vol. 133, no. 11, pp. 5611–5625, 2024.
[29] G. O. Barrionuevo, J. A. Ramos-Grez, M. Walczak, and C. A. Betancourt, 'Comparative evaluation of supervised machine learning algorithms in the prediction of the relative density of 316L stainless steel fabricated by selective laser melting', *International Journal of Advanced Manufacturing Technology*, vol. 113, no. 1–2, pp. 419–433, 2021.
[30] J. Gehrke, R. Ramakrishnan, and V. Ganti, 'RainForest - A framework for fast decision tree construction of large datasets', *Data Mining and Knowledge Discovery*, vol. 4, no. 2–3, pp. 127–162, 2000.
[31] H. A. Salman, A. Kalakech, and A. Steiti, 'Random Forest Algorithm Overview', *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, 2024.
[32] A. K. Kausik, A. Bin Rashid, R. F. Baki, and M. M. Jannat Maktum, 'Machine learning algorithms for manufacturing quality assurance: A systematic review of performance metrics and applications', *Array*, vol. 26, no. March, p. 100393, 2025.
[33] E. Scornet, 'Tuning parameters in random forests', *ESAIM: Proceedings and Surveys*, vol. 60, pp. 144–162, Dec. 2017.
[34] M. Gor *et al.*, 'Density Prediction in Powder Bed Fusion Additive Manufacturing: Machine Learning-Based Techniques', *Applied Sciences (Switzerland)*, vol. 12, no. 14, 2022.
[35] D. Wu, Y. Wei, and J. Terpenny, 'Surface Roughness Prediction in Additive Manufacturing Using Machine Learning', in *Proceedings of the ASME 2018 13th International Manufacturing Science and Engineering Conference. Volume 3: Manufacturing Equipment and Systems*, 2018.