Solid Freeform Fabrication 2021: Proceedings of the 32nd Annual International
Solid Freeform Fabrication Symposium – An Additive Manufacturing Conference
Reviewed Paper

Powder Features Affecting Structural and Mechanical Properties of Additively Manufactured Inconel 718: A Machine Learning Analysis

M. S. Hossain[1,3], D. F. Silva[1,3], A. Vinel[1,3], J. Liu[1,3], and N. Shamsaei[2,3]

[1] Department of Industrial and Systems Engineering, Auburn University, Auburn, AL 36849
[2] Department of Mechanical Engineering, Auburn University, Auburn, AL 36849
[3] National Center for Additive Manufacturing Excellence, Auburn University, Auburn, AL 36849

Abstract

The aim of this paper is to select important Inconel 718 powder properties that can have significant effect on the structural and mechanical properties of Laser-Beam Powder Bed Fusion manufactured specimens. The dataset used was provided by NASA and contains powder rheological, morphological, and chemical composition properties. The output variables considered are melt pool depth, high cycle fatigue life, porosity volume fraction and porosity size. Initially, Pearson correlation coefficient matrix is used to reduce the number of predictor features. Several statistical and machine learning algorithms including stepwise regression, LASSO, and random forest regression are used to identify the powder properties that have the strongest impact on the selected outputs. The variables identified using the different statistical and machine learning techniques are similar, which increases the confidence of the findings.

Keywords: LB-PBF, Correlation coefficient, Variance inflation factor, LASSO, Random forest regression, Stepwise regression

1. Introduction

Laser beam powder bed fusion (LB-PBF) is a metal Additive Manufacturing (AM) process that uses a computer driven, highly focused laser beam which melts a metallic powder layer, resulting in building parts layer-by-layer. The benefits that LB-PBF offers include; the design and customization of products, the fabrication of complex geometries while keeping waste down, the reusability of powder for LB-PBF which reduces production cost, and alleviating of the difficulties of manufacturing superalloys such as Inconel 718, which occur in traditional manufacturing.

The application of machine learning (ML) in additive manufacturing has become widespread. The main application of ML in additive manufacturing (AM) has been optimization of the process parameters. There are comparatively fewer works that discuss the powder properties and no work yet that identifies powder properties for AM that most impact to structural and mechanical properties of produced part. This work tries to fill that gap by identifying the most important powder features that contribute to porosity and fatigue life.

For Inconel 718 AM through LB-PBF, this paper presents selection of relevant powder properties that influence the properties of a produced part. Specifically, variant statistical and machine learning tools, namely correlation coefficient matrix, LASSO, random forest regression, and stepwise regression are applied to meet the objective. The response variables selected for this study are melt pool depth, high cycle fatigue life, porosity volume fraction and porosity size. The results obtained upon the applications of different algorithms agree with each other, adding to the confidence that the correct factors have been identified.

This paper is organized as follows. Section 2 provides a brief literature survey concerning LB-PBF as manufacturing process, Inconel 718 as material, powder properties for metal AM, feature selection algorithms and predictive modeling techniques used in additive manufacturing. Section 3 presents the work methodology along with a brief description on each of the algorithms that were applied. Section 4 focuses on the results obtained. Finally, Section 5 presents concluding remarks and some future directions.

2. Literature Review

AM processes produce physical objects from digital information layer-by-layer, simultaneously defining the object's geometry and determining its material properties. In the world of manufacturing, AM is currently the best solution to fabricate a near-net-shape component with minimal or without post-processing [1, 2]. AM by LB-PBF is a specific subset of metal additive manufacturing technologies in which a metal powder is selectively melted layer-by-layer by a computer-driven highly focused laser beam. The benefits that LB-PBF offers are fabrication of complex shapes without requiring long supply chain networks, production waste that is minimal, and product customization. Another notable advantage of the LPBF technology is that only the locally melted powder transforms into a solid material while the remaining powder can be reused a number of times [3, 4]. These days, LB-PBF is finding its way in advanced manufacturing applications such as manufacturing of shape memory alloy [5], landing gear of aircraft [6] and magnetic materials [7]. The main limitations for LB-PBF are low build rate and the need to clear unmelted powder from internal geometry [8]. All PBF systems can be used to process powders of a relatively large range of metal alloys, such as titanium alloys, stainless steels and Nickel-based super alloys such as Inconel 718 [4].

Inconel 718 is a work hardenable Nickel- Chromium- Iron austenite (γ) based superalloy possessing a wide range of compositions and mechanical properties [9-10]. The alloy can maintain oxidation and corrosion resistance at very high temperature (up to 700^0C) [11]. This engineering material is recognized in the manufacturing world for its good cryogenic and fatigue properties. The superalloy also possess excellent weldabilitty and creep properties as well. Since its initial use in the mid-1960s, Inconel 718 has gained attention in building mechanical components such as gas turbines, rocket engines, turbine blades, and in extrusion dies and containers [12]. Traditionally, Inconel 718 products are manufactured with vacuum induction melting and high precision machining but because of strain hardening of the material, the cutter wears down quickly and machining becomes difficult [1]. As an alternative, the recent emergence of metal AM such as LB-PBF eases the manufacturing of Inconel 718 products with high precision and accuracy.

Researchers have employed techniques to measure rheological properties of Inconel 718 powders [1] to compare with the produced parts' structural or mechanical properties to measure the impact of powder reuse. In [1], the authors concluded that apparent and tapped densities of virgin powder were higher than reused powder because of the deformed shape of reused powder.

However, true density was found to be higher for reused powder because of removal of entrapped gas and remelting. It was also found that smaller particle size resulted in a smoother surface finish because of homogeneous melting. Virgin powder showed better flowability than recycled powders because of virgin powder's more homogenous shape and recycled powder lost sphericity with time and also tended to stick together. Studying 21 samples of Fe and Ni based alloy powders, researchers [14] concluded avalanche surface fractal and avalanche angle are the two factors influence powder flowability. Ellipticity or convexity of powder particle also influence powder flowability by reducing powder density. As per their argument, spherical powders provided higher powder density. In another study [15], six different powders' properties were measured using powder rheometer and rotational shear cell. They concluded that powder flowability was higher for non-cohesive powders and that aeration improved flowability. Powder characterization and flowability were also investigated in [16] using four different powders; they concluded that powders having low values of flow energy, flow rate index, shear strength, compression strength, consolidation index, internal angle of friction, and cohesion had better flowability. Investigating particle morphologies, particle chemistry, and particle microstructure [17], it was concluded that average particle size and particle size distribution influence layer homogeneity; porosity of as built part is also influenced in presence of large particles or agglomerates. They also asserted that maximizing flowability coupled with increasing powder-bed density can be done using narrow particle distribution with less surface deformities. Studying the effects of powder reuse on tensile and fatigue behavior of LB-PBF 17-4 stainless steel parts [18], it was found that fatigue behavior of parts was dependent on location but diminished with powder reuse. Findings from [18] also include that elongated powder showed less flowability and generated higher shear stress, circular equivalent diameter of powder decreased due to powder reuse because of the breaking down of powder particle agglomerates, and with powder reuse bulk and tapped densities of the powder found to be increased. Also, powder reuse did not show any significant impact on microstructure which resulted in no difference in strength of parts produced with either virgin or reused powder. The papers above are in agreement that with powder reuse, spherical powder particles are produced providing better flowability with higher bulk and tapped density.

Much experimental work is being carried out in the AM field every day, the involvement of statistical and machine learning tools is also present in the field. To predict a response variable with high precision and accuracy, it is important to select the important predictor variables that would describe which attributes are more significant in predicting the response variable [19]. Several techniques exist for this purpose. For example, [20] shows how LASSO regression can be used in model selection purpose. Similarly, [19] explained how LASSO regression can be used in selecting important variables from a high dimensional dataset. The article also explains Ridge regression and elastic net regression for selecting important features and discusses their limitations. Stepwise regression is another highly used tool that has been employed in reducing the dimension of the dataset to select the important features [21]. In addition to many other ensemble methods, the application of random forest algorithm is clearly noticeable in selecting important features [22, 23] in different areas. However, in the area of AM; LASSO regression, stepwise regression and random forest regression are yet to be applied for selecting significant features of powder to predict structural and mechanical properties of the finished part. The goal of this article is to fill that gap, by using these techniques to identify the features that best predict porosity, melt-pool depth and fatigue life of Inconel 718 AM parts.

3. Methodology

The purpose of this work is to select important powder morphological and rheological properties which can be used to predict structural and mechanical for LB-PBF manufactured parts. The dataset used for this study was provided by NASA [24] and it includes 16 virgin powders and 3 once recycled powders of Inconel 718 produced by gas or rotary atomization and bought from different suppliers. A set of 31 candidate features and 4 response variables were selected initially out of 191 columns found in the original dataset, but for this work the number of candidate features was reduced down to a level similar to the number of observations, i.e. 19. The response variables chosen for this study are average high cycle fatigue (HCF) life for as fabricated parts, porosity volume fraction, porosity size and melt pool depth. With a view to selecting the important features, a number of statistical and machine learning tools are applied for each of the response variables. In the first step, a correlation matrix is built containing all the 31 candidate features. Next, by applying a cutoff Pearson correlation coefficient value of 0.70, 15 features are selected. Upon finding the 15 features; LASSO, random forest regression, and stepwise regression are applied for each of the response variables to select 10 or less than 10 significant features for each response variable.

3.1 Dataset Description

The original dataset obtained from NASA has 191 columns and 19 rows which was reduced down to 15 features and 4 response variables for 19 observations. This dataset has also been used to study the variation in microstructure and mechanical behavior of LB-PBF processed Inconel 718 product due to powder variability [24]. However, this work tries to discover the impact of morphological and rheological properties of powder on LB-PBF processed Inconel 718 part properties namely porosity volume fraction, porosity size, melt pool depth and high cycle fatigue life. Because of missing values in the original dataset, among the 19 variants of Inconel 718 powders provided by the suppliers, only 13 observations could be used for this study. Each powder either came from 7 different suppliers or from same supplier's different lot. In this study, no reused powder has been used. To reduce the number of candidate features, from 191, several criteria were applied. First, quantile data are avoided, chemical constituents are also rejected since this goes beyond the scope of this study; materials' phase related data are also avoided in the study. Among the morphological data were present in the original dataset; the mean, standard deviation, skewedness and kurtosis of different morphological parameters such as circular equivalent (CE) diameter, circularity, convexity, aspect ratio have been selected in the study. For this study, all powders including fines are considered where particle fines can be defined as powder particles whose diameter is less than or equals to 7 μm. Among the rest of the candidate features from the original dataset bulk density and angle of repose are selected in the first phase of variable selection. Combining all the selected features, in the first phase we have a dataset that consists of 31 candidate features and 4 response variables and 13 observations. Table 1 shows the initially selected 31 candidate features. The 4 response variables used in this work are high cycle fatigue life for as fabricated part, porosity volume fraction and porosity size and melt pool depth. In the table, d5, d10, d50, d90 and d95 are quantiles distributions of CE diameter of powder particle. As an example, d90 would mean the CE diameter of 90% particle are less than d90 value.

Table 1 Selected candidate features from the Original NASA powder dataset

Apparent density	Percentage of fines	Standard deviation of convexity	Aeration energy
Angle of repose	Mean circular equivalent (CE) diameter	Skewedness of convexity	Normalized aeration sensitivity
Adjusted flow rate	Standard deviation of CE diameter	Kurtosis of convexity	Consolidation energy
d5	Skewedness of CE diameter	Basic flow energy	Permeability
d10	Kurtosis of CE diameter	Stability index	Compressibility
d50	Mean aspect ratio (AR)	Flow rate index	Shear stress upper
d90	Standard deviation of AR	Specific energy	Shear stress lower
d95	Mean convexity	Conditioned bulk density	–

3.2 Pearson Correlation Coefficient Matrix

Since there is a chance of high multicollinearity among the predictor variables, in the first step we tried to get rid of multicollinearity using Pearson correlation coefficient. Correlation coefficient can be defined as a dimensionless number measuring the linear association between two variables, lying in the interval from -1 to $+1$ [25]. A correlation coefficient value of 0 indicates no correlation but does not guarantee independence between the variables. If **x** and **y** are two vectors then the correlation coefficient can be written mathematically as follows where $\bar{x}$ and $\bar{y}$ are the mean of x and y vector respectively.

$$\mathrm{Corr}(x, y) = \frac{\sum_{i=1}^{n}[(y_i-\bar{y})(x_i-\bar{x})]}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2(x_i-\bar{x})^2}}$$

There are 31 candidate features after the initial screening. The goal is to select the predictor features among the 31 candidates in a way such that no two predictor features would have a correlation coefficient of more than 0.70. Upon the application the Pearson correlation coefficient, the desired number of features came down to 15. These are: apparent density, angle of repose, adjusted flow rate, d50, mean CE diameter, standard deviation of CE diameter, skewedness of circularity, kurtosis of convexity, basic flow energy, stability index, conditioned bulk density, aeration energy, consolidation energy, shear stress upper and shear stress lower. Even after this filtering, the number of features is higher than number of observations, as a result, variance inflation factor analysis could not be run on the predictors features. Therefore, we applied several other well recognized machine learning algorithms namely LASSO, random forest regression and stepwise regression to select the most significant features. Below we describe each method.

3.3 Least Absolute Shrinkage and Selection Operator

Unlike ordinary least squares (OLS) which tries to minimize the squared sum of residuals (SSE); in Least Absolute Shrinkage and Selection Operator (LASSO), a little bias is incorporated to reduce the variance. LASSO minimizes a non-linear objective function containing both response and predictor variables [19]. The mathematical formulation for LASSO can be written as following. Given a vector of inputs **X** and outputs **Y**,

$$\hat{\beta} = \text{argmin}_{\beta}\left(\frac{\|\mathbf{Y} - \mathbf{X}\beta\|_2^2}{n} + \lambda\|\beta\|_1\right)$$

$$\text{where } \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 = \sum_{i=0}^{n}(Y_i - (\mathbf{X}\beta)_i)^2 \text{ and}$$

$$\|\beta\|_1 = \sum_{j=1}^{k}|\beta_j| \text{ and } \lambda \geq 0$$

Therefore, LASSO seeks a β vector which minimizes the objective value and that vector can be obtained using desired regularization constant (λ). The number λ measures the severity of penalty. It can range from 0 to $+\infty$. For λ = 0, LASSO line is the OLS line but if $\lambda \rightarrow \infty$, the penalty becomes so large that all the coefficients are forced to be zero. In this work, LASSO is utilized with tuned λ value to capture the important features. The tuning parameter values were generated from a one dimensional array which ranged from 10^{-6} to 10 containing 100,000 regularization constant (λ) values regardless of the response variable. For each response variable, same sized array was generated and it worked as both features and response variables were normalized. Among these values, optimal values of λ were used to perform LASSO operation for each of the response variable. Based on the working principle of LASSO, it picks the λ value which minimizes the squared sum of residuals. The results obtained from different algorithms are listed in the following subsections.

3.4 Random Forest Regression

Among the feature selection algorithms that work on the method of aggregation, random forest (RF) is one which can be applied for both classification and regression problems [22]. The principle of random forest is to combine large number of binary decision trees built using several bootstrap samples coming from the learning sample **L**. Assuming vector $\mathbf{X} \in \mathbb{R}^P$contains predictors features where, $\mathbf{X} = (X^1, \ldots \ldots, X^n)$ and vector $\mathbf{Y} \in \mathcal{Y}$ label i.e. is a binary $\{0, 1\}$ valued random variable or numerical response having a space, $\mathcal{Y} \in \mathbb{R}^P$then learning set **L** is made of n independent and identically distributed vector $(\mathbf{X}, \mathbf{Y})$. Therefore, learning set can be designated as $\mathbf{L} = \{(X_1, Y_1), \ldots., (X_n, Y_n)\}$. If t is a classifier for a classification problem then t can be defined as a mapping t: $\mathbb{R}^P \rightarrow \mathcal{Y}$ and for regression problem we can write, $Y = s(X) + \epsilon$ where $E[\epsilon|X] = 0$. At each node, a given number of input variables are randomly chosen and the best split is calculated only within the subset. During the process, no pruning is performed so all the trees of the forest are maximal trees. In this work, the number of input variables to be chosen at each node has been kept default which is the rounded down value of $\frac{P}{3}$ where P is the total number of features (i.e. 15) and number of trees are kept fixed at 1000.

3.5 Variance Inflation Factor

The variance inflation factor (VIF) is a popular tool to assess whether there is any multicollinearity among the predictor variables or not. VIF can confirm how much of a regressor's variability is explained by the rest of the regressors in the model due to correlation among those regressors [26]. The value of a VIF has direct impact on the width of the confidence interval for its associated parameter estimate. For independent variables, If r_j^2 represents the coefficient of determination for the j^{th} independent variable, then VIF is mathematically written as following.

$$\mathrm{VIF_j} = \frac{1}{1 - r_j^2} \ \forall \ j = 1, 2, .., p - 1$$

r_j^2 is obtained by fitting a regression model for the j^{th} independent variable on the other independent variables. Although there is no fixed rule for a cutoff value of VIF but a value of 5-10 is often popularly used as cutoff. An r_j^2= 0.90 would signify that 90% of the variability in the j^{th} independent variable is explained by the remainder of the independent variables in the model. In this work, VIF is used to assess the multicollinearity among the features that are chosen by random forest from the 15 selected features.

3.6 Stepwise Regression

Stepwise regression, as the name suggest, works by performing OLS regression step by step by adding variables one at a time or by removing variables one at a time [27]. In forward selection stepwise regression, the model starts with no variables, in the first step one variable is added which is believed to be the most significant. In the following step next best significant variable is chosen out of all the candidate variables. This process continues until all the variables are added or a stopping criterion is met. Similarly, in backward elimination stepwise regression method, the initial model starts with all the candidate variables and in each step model keeps on eliminating variables. The first variable to be removed is the least significant. This process keeps continuing as long as all variables are removed or a stopping condition is fulfilled. In this work, Akaike information criterion (AIC) is used as the model selection criterion for the stepwise regression models. For small cases, where $\frac{n}{K} < 40$, the formula is following.

$$\mathrm{AIC_c} = -2\ln\left[l\left(\hat{\theta}|\mathrm{data}\right)\right] + 2K + \frac{2K(K+1)}{n-K+1}.$$

where, $l(\hat{\theta}|\mathrm{data})$ = the likelihood of the parameter $\hat{\theta}$ in the model, n = number of observations, and K = number of estimable parameters in the model [28]. In this work, $n = 13$ and $K = 3$. The lower the AIC value is, the better fit the model is regardless of forward selection or backward elimination method.

To the best of our knowledge, this is the first research work where the sum of these aforementioned statistical and machine learning techniques have been used in the area of additive manufacturing to predict part properties based on the powder properties.

4. Results and Discussion

After applying the algorithms described above on the dataset for each of the response variables, the obtained results are presented in the following subsections. It is to be noted, before performing each regression analysis, the dataset was normalized using max-min normalization.

4.1 Porosity Volume Fraction

The first algorithm attempted to select the significant features on porosity volume fraction (porosity VF) is LASSO. Since the dataset is high dimensional i.e. it has bigger number of features than the number of observations, it is a suitable dataset where LASSO could be applied. Upon applying LASSO using library 'glmnet' in software R [29], it is found that shear stress lower is the only feature that LASSO finds significant on porosity VF. For LASSO, optimal regularization constant (λ) value found to be 0.1054.

Next, to select the significant features on porosity VF, random forest (RF) regression is performed keeping porosity VF as the response variable and the number of features are kept at 15. Upon performing RF regression using 'randomforest' library [29] in software R, the most significant features are selected. It is also verified using variance inflation factor (VIF) that there is no multicollinearity involved among the selected features. As per random forest regression, shear stress lower is found to be the most significant feature on porosity volume fraction. The ranking of features on porosity volume fraction according to random forest regression is shown in Fig. 1.

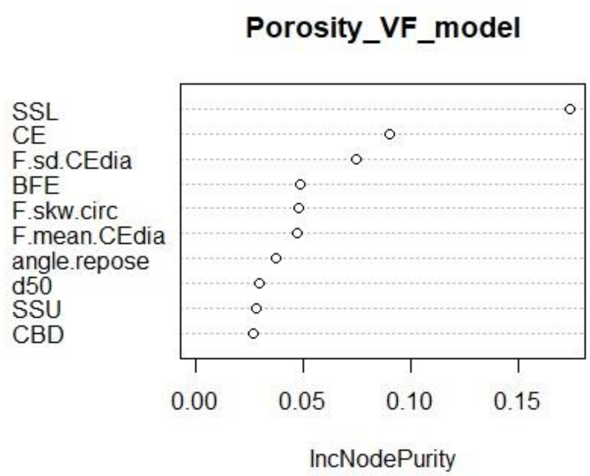

Fig. 1: Ranking of features on porosity volume fraction according to RF regression

Fig. 1 shows that the features are clearly divided into three classes based on importance. The most important feature is shear stress lower. A class of intermediate importance includes Consolidation Energy and Full SD CE diameter. While the remaining features appear to be of little importance. The following table presents the variance inflation factor among the most important features selected by random forest regression. VIF values of the selected features are random forest are shown in Table 2.

Table 2 Variance inflation factor for the selected features by random forest for Porosity VF

Features	VIF value
Shear stress lower	1.319
Consolidation energy	1.140
Full SD CE diameter	1.209

After selecting the significant features using RF regression, both forward selection and backward elimination stepwise regression are performed based on the 10 highest ranked features selected by random forest. It is found that shear stress lower is the most statistically significant features among the selected features. Both forward selection and backward elimination model agree with each other. Table 3 presents different statistics obtained from stepwise regression.

Running the stepwise regression, both forward selection and backward elimination based on features selected by random forest, it is found that shear stress lower is the most statistically significant features among the 10 features attempted. In Table 3, the value of F statistic is shown along with its p-value. Since, the p-value of F statistic less than 0.10, it can be said with 90% confidence that there is a relationship between the features and the response variable.

Table 3 Different statistic for porosity volume fraction for stepwise regression

Regression Model	**Adjusted R^2 value**	**F statistic**	**Residual standard error**	**p-value**
Stepwise regression	0.7455	5.395	0.1395	0.0605

Table 4 lists the features that are significant on porosity volume fraction, for each method. By taking a cut off of 0.05 for node purity score, three most significant features are found according to random forest. In all of the following tables, double * in the parentheses would indicate a confidence level of 99%, single * in the parentheses would indicate a confidence level of 95% and · in the parentheses would indicate a confidence level of 90%.

Table 4 Algorithms and corresponding features significant on porosity volume fraction

Response variable	**LASSO**	**Random forest regression**	**Stepwise regression**
Porosity VF	Shear stress lower	Shear stress lower	Shear stress lower (∗∗)
	−	Consolidation energy	d50 (·)
	−	Full SD of CE diameter	−

Table 4 shows that LASSO, random forest regression, and stepwise regression agree that shear stress lower is the most significant predictor feature for porosity volume fraction.

4.2 Porosity Size

Similar to porosity volume fraction, in the case of porosity size, LASSO is attempted first to select one of more significant features from 15 candidate features. LASSO found shear stress lower, standard deviation of circular equivalent diameter, basic flow energy consolidation energy and median of circular equivalent diameter to be the most significant features on porosity size. The optimal regularization constant value found for lasso is 0.0153.

By taking a node purity value of 0.05 as a cutoff, random forest selected three most important features for porosity size. Fig. 2 shows the ranking of features according to random forest regression. Shear stress lower, consolidation energy, and basic flow energy are found to be the most significant predictor features according to random forest regression.

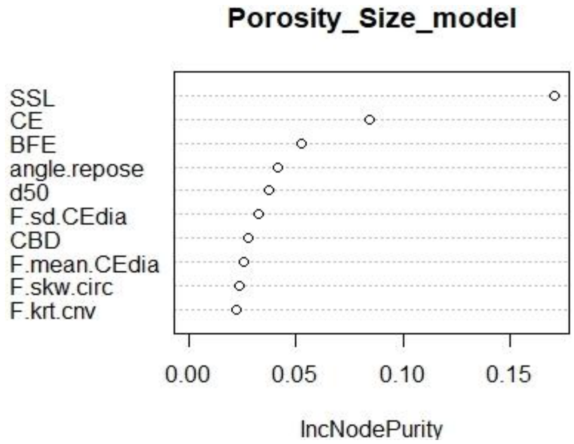

Fig. 2: The ranking of features on porosity size according to RF regression

Table 5 shows the variance inflation factor among the features selected by random forest regression while porosity size is the response variable. From the VIF analysis, it is found that there is no multi-collinearity involved among the features selected by random forest.

Table 5 Variance inflation factor for the selected features by random forest regression for Porosity size

Features	VIF value
Shear stress lower	1.180
Consolidation energy	1.186
Basic flow energy	1.228

Based on the features selected by random forest, stepwise regression is performed as the absence of multicollinearity is already ascertained by VIF analysis. Upon performing stepwise regression using both forward selection method and backward elimination method, significant features on porosity size are determined. Both forward method and backward method agree with each other. Table 6 shows important statistics on stepwise regression.

Table 6 Table showing different statistic for porosity size for stepwise regression

Regression Model	Adjusted R^2 value	F statistic	Residual standard error	p-value
Stepwise regression	0.8817	15.90	0.08677	0.0019

According to Table 6, a very high adjusted R^2 value is noticed in a scale of 0 to 1 signifying a strong fit of the response variable with the features. A p-value less than 0.01 also indicates a confidence level of 99% that response variable is related with the predictor features.

From the stepwise regression results, it is found that shear stress lower is the most significant feature and standard deviation of circular equivalent diameter and median of circular equivalent diameter are also very significant. The forward and backward methods agree totally in selecting significant variables. Table 7 lists the important features according to the discussed three algorithms.

Table 7 Algorithms and corresponding features significant on Porosity size

Response variable	LASSO	Random forest regression	Stepwise regression
Porosity size	Shear stress lower	Shear stress lower	Shear stress lower (∗∗∗)
	Full SD of CE diameter	Consolidation energy	Consolidation energy (·)
	Basic flow energy	Basic flow energy	d50 (∗)
	Consolidation energy	–	Full SD of CE diameter (∗)
	d50	–	–

All of the three algorithms agree that shear stress lower is the most significant feature in predicting porosity size. Consolidation energy is also another feature that is selected by all algorithms which can be used in predicting porosity size.

4.3 Melt Pool Depth (MPD)

For melt pool depth, LASSO could not select any important features, as the results were never statistically significant.

To select the important features on melt pool depth, random forest regression is also performed keeping melt pool depth as the response variable and using same parameters as porosity VF and porosity size. The ranking of the features obtained from random forest regression for melt pool depth is shown in Fig. 3.

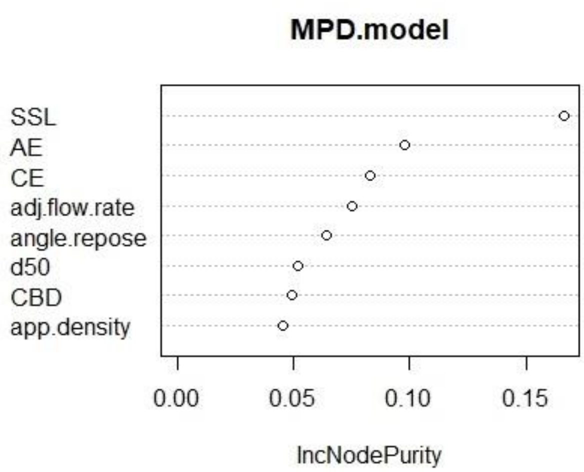

Fig. 3: Ranking of features on melt pool depth according to random forest regression

Upon imposing a cutoff value of 0.05 for node purity score, random forest selected five most significant features. Based on the features selected by random forest regression, VIF analysis is performed. It is found that there is no multicollinearity involved among the features that random forest selected. Table 8 shows the VIF value for each of the significant features selected by random forest when melt pool depth is response variable.

Table 8 Variance inflation factor of selected features by random forest for MPD

Features	VIF value
Shear stress lower	3.066
Aeration energy	2.063
Consolidation energy	2.439
Adjusted flow rate	2.766
Angle of repose	3.858

Upon obtaining the significant features from random forest regression, where there is no multicollinearity involved among the features is ensured, stepwise regression is performed keeping melt pool depth as the response variable. It is found that stepwise regression also selects features that have also been selected by random forest. Table 9 shows the statistics of stepwise regression.

Table 9 Different statistic for melt pool depth for stepwise regression

Regression Model	Adjusted R^2 value	F statistic	Residual standard error	p-value
Stepwise regression	0.7806	9.541	0.1332	0.004971

According to stepwise regression, based on the p-value of the F statistic, it is highly likely that melt pool depth is related with the features. Table 10 lists the variables that are significant on melt pool depth, for each method.

Table 10 Algorithms and corresponding features significant on melt pool depth

Response variable	LASSO	Random forest regression	Stepwise regression
Melt pool depth	–	Shear stress lower	Aeration energy ($**$)
	–	Aeration energy	Angle of repose ($**$)
	–	Consolidation energy	Apparent density ($**$)
	–	Adjusted flow rate	Shear stress lower ($\cdot$)
	–	Angle of repose	d50

According to Table 10, LASSO did not find any significant predictor features on melt pool depth but shear stress lower, aeration energy and angle of repose are found to be predictive on melt pool depth based on both random forest and stepwise regression.

4.4 High Cycle Fatigue life (HCF life)

To select the significant features on high cycle fatigue (HCF) life, LASSO is attempted at first. It is found that LASSO captured 7 significant features for HCF which are angle of repose, d50, mean CE diameter, standard deviation of CE diameter, aeration energy (AE), consolidation energy (CE) and conditioned bulk density. The tuned regularization constant value in LASSO for HCF is found to be 0.03462.

For HCF life, applying random forest regression keeping HCF life as the response variable found some significant features are found. The ranking of the features obtained from random forest regression for high cycle fatigue life is shown in Fig. 4.

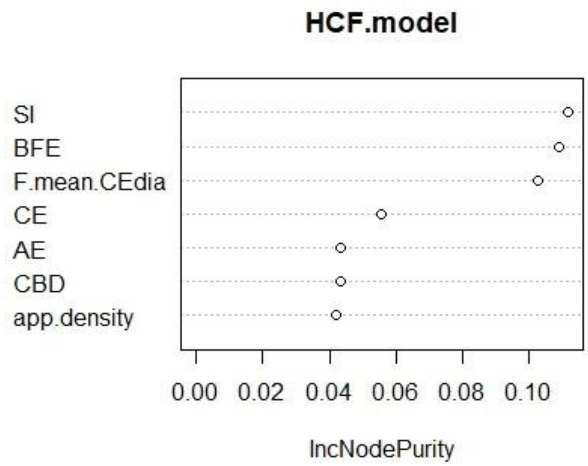

Fig. 4: Ranking of features on HCF life according to random forest regression

Applying a cutoff value of 0.10 for node purity, three significant features are selected by random forest regression. The variance analysis factor of the features selected by random forest regression are performed and the absence of multicollinearity is also ascertained from the VIF values as shown in Table 11.

Table 11 VIF for the selected features from random forest regression for HCF life

Features	VIF value
Stability index	1.372
Basic flow energy	1.387
Full mean CE diameter	1.013

After obtaining the important features from LASSO and random forest, stepwise regression is performed and the value of different statistic from stepwise regression is listed in Table 12. The significant features selected by three algorithms are listed in Table 13.

Table 12 Different statistic for high cycle fatigue life for stepwise regression

Regression Model	Adjusted R^2 value	F statistic	Residual standard error	p-value
Stepwise regression	0.7735	7.835	0.1148	0.01216

Based on the p-value of the F-statistic, it can be asserted with a confidence of 90% that average HCF life is related with the predictor features.

Table 13 presents the different algorithms and corresponding significant variables on high cycle fatigue life.

Table 13 Algorithms and corresponding features significant on average HCF life

Response variable	LASSO	Random forest regression	Stepwise regression
High cycle fatigue life	Angle of repose	Stability index	Full mean CE diameter (∗∗)
	d50	Basic flow energy	Full SD CE diameter (∗∗)
	Full mean CE diameter	Full mean CE diameter	Aeration energy (∗∗)
	Full SD CE diameter	–	Angle of repose (∗)
	Conditioned bulk density	–	–
	Consolidation energy	–	–
	Aeration energy	–	–

Based on the table we find, LASSO and stepwise regression results agree with each other and the significant predictor variables are angle of repose, aeration energy, median of circular equivalent diameter, standard deviation of circular equivalent diameter.

5. Conclusions

From this study of effect of powder's morphological and rheological properties on the LB-PBF manufactured Inconel 718 part properties the following conclusions can be drawn.

1. In this research work, an attempt is made to select the powder features of Inconel 718 manufactured in LB-PBF AM that have the greatest impact on porosity size, porosity volume fraction, melt pool depth and HCF life of the finished part. To meet the objective, several statistical and machine learning tools namely Pearson correlation coefficient, LASSO regression, random forest regression and stepwise regression were employed. In the first step, 31 predictor features and 4 response variables were selected from a dataset consisted of 191 columns. In the next step, a Pearson correlation coefficient cut off value of 0.70 was used to reduce the number of features to 15. Upon obtaining 15 features, different statistical and feature selection techniques were applied to select the most significant predictor features.

2. Applying LASSO, random forest regression and stepwise regression based on the features selected by random forest, it is found that shear stress lower is the most significant feature for both porosity size and porosity volume fraction. Shear stress lower is also found to be highest ranked feature according to random forest. All the algorithms are in agreement about shear stress lower's significance on both porosity volume fraction and porosity size.

3. The most significant features impacting melt pool depth according to random forest regression are shear stress lower, aeration energy and angle of repose. This finding from random

forest regression agrees with forward selection and backward elimination stepwise regression. LASSO regression, however, did not find any features to be significant.

4. For high cycle fatigue, the seven most important features are selected using LASSO. Running a multiple linear regression taking seven predictor features, it is found that mean of circular equivalent diameter is the most significant feature for predicting average high cycle fatigue life. This finding also agrees with LASSO and random forest regression results.

The findings from this research are in agreement with existing literature [1]. This work only identifies the important predictive feature from the perspective of selecting predictor variables that influence the part quality of Inconel 718 manufactured in LB-PBF. Future work is required to create predictive models based on the selected features. However, this work enables such possibilities, by identifying the important ones to use.

Acknowledgement

The authors of this article would like to thank NASA and Brian West, Additive Manufacturing Engineer at NASA for supporting this research work by providing the Inconel 718 powder dataset for LB-PBF manufacturing. The authors would also like to thank everyone involved in the research [24] for setting the initial baseline of the current research work.

References

[1] Nguyen, Q.B., Nai, M.L.S., Zhu, Z., Sun, C.N., Wei, J. and Zhou, W., 2017. Characteristics of inconel powders for powder-bed additive manufacturing. *Engineering*, *3*(5), pp.695-700.

[2] Thompson, M.K., Moroni, G., Vaneker, T., Fadel, G., Campbell, R.I., Gibson, I., Bernard, A., Schulz, J., Graf, P., Ahuja, B. and Martina, F., 2016. Design for Additive Manufacturing: Trends, opportunities, considerations, and constraints. *CIRP annals*, *65*(2), pp.737-760.

[3] Agrawal, P., Thapliyal, S., Nene, S.S., Mishra, R.S., McWilliams, B.A. and Cho, K.C., 2020. Excellent strength-ductility synergy in metastable high entropy alloy by laser powder bed additive manufacturing. *Additive Manufacturing*, *32*, p.101098.

[4] Nicoletto, G., 2017. Anisotropic high cycle fatigue behavior of Ti–6Al–4V obtained by powder bed laser fusion. *International Journal of Fatigue*, *94*, pp.255-262.

[5] Chen, X., Liu, K., Guo, W., Gangil, N., Siddiquee, A.N. and Konovalov, S., 2019. The fabrication of NiTi shape memory alloy by selective laser melting: a review. *Rapid Prototyping Journal*.

[6] LaTour, A., Charles, M., Raghunathan, N. and Ahlfors, M., 2020. Towards production of landing gear materials with laser powder bed fusion. *International Journal of Refractory Metals and Hard Materials*, *91*, p.105272.

[7] Bittner, F., Thielsch, J. and Drossel, W.G., 2020. Laser powder bed fusion of Nd–Fe–B permanent magnets. *Progress in Additive Manufacturing*, *5*(1), pp.3-9.

[8] Fayazfar, H., Salarian, M., Rogalsky, A., Sarker, D., Russo, P., Paserin, V. and Toyserkani, E., 2018. A critical review of powder-based additive manufacturing of ferrous alloys:

Process parameters, microstructure and mechanical properties. *Materials & Design*, *144*, pp.98-128.

[9] Thomas, A., El-Wahabi, M., Cabrera, J.M. and Prado, J.M., 2006. High temperature deformation of Inconel 718. *Journal of materials processing technology*, *177*(1-3), pp.469-472.

[10] Hosseini, E. and Popovich, V.A., 2019. A review of mechanical properties of additively manufactured Inconel 718. *Additive Manufacturing*, *30*, p.100877.

[11] Jia, Q. and Gu, D., 2014. Selective laser melting additive manufacturing of Inconel 718 superalloy parts: Densification, microstructure and properties. *Journal of Alloys and Compounds*, *585*, pp.713-721.

[12] Chlebus, E., Gruber, K., Kuźnicka, B., Kurzac, J. and Kurzynowski, T., 2015. Effect of heat treatment on the microstructure and mechanical properties of Inconel 718 processed by selective laser melting. *Materials Science and Engineering: A*, *639*, pp.647-655.

[13] Strößner, J., Terock, M. and Glatzel, U., 2015. Mechanical and microstructural investigation of nickel-based superalloy IN718 manufactured by selective laser melting (SLM). *Advanced Engineering Materials*, *17*(8), pp.1099-1105.

[14] Spierings, A.B., Voegtlin, M., Bauer, T.U. and Wegener, K., 2016. Powder flowability characterisation methodology for powder-bed-based metal additive manufacturing. *Progress in Additive Manufacturing*, *1*(1), pp.9-20.

[15] Freeman, R., 2007. Measuring the flow properties of consolidated, conditioned and aerated powders—a comparative study using a powder rheometer and a rotational shear cell. *Powder Technology*, *174*(1-2), pp.25-33.

[16] Freeman, R. and Fu, X., 2008. Characterisation of powder bulk, dynamic flow and shear properties in relation to die filling. *Powder Metallurgy*, *51*(3), pp.196-201.

[17] Sutton, A.T., Kriewall, C.S., Leu, M.C. and Newkirk, J.W., 2017. Powder characterisation techniques and effects of powder characteristics on part properties in powder-bed fusion processes. *Virtual and physical prototyping*, *12*(1), pp.3-29.

[18] Soltani-Tehrani, A., Pegues, J. and Shamsaei, N., 2020. Fatigue behavior of additively manufactured 17-4 PH stainless steel: The effects of part location and powder re-use. *Additive Manufacturing*, *36*, p.101398.

[19] Fonti, V. and Belitser, E., 2017. Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, *30*, pp.1-25.

[20] Zhao, P. and Yu, B., 2006. On model selection consistency of Lasso. *The Journal of Machine Learning Research*, *7*, pp.2541-2563.

[21] Shanableh, T. and Assaleh, K., 2010. Feature modeling using polynomial classifiers and stepwise regression. *Neurocomputing*, *73*(10-12), pp.1752-1759.

[22] Genuer, R., Poggi, J.M. and Tuleau-Malot, C., 2010. Variable selection using random forests. *Pattern recognition letters*, *31*(14), pp.2225-2236.

[23] Gregorutti, B., Michel, B. and Saint-Pierre, P., 2017. Correlation and variable importance in random forests. *Statistics and Computing*, *27*(3), pp.659-678.

[24] Sudbrack, C.K., Lerch, B.A., Smith, T.M., Locci, I.E., Ellis, D.L., Thompson, A.C. and Richards, B., 2018. Impact of powder variability on the microstructure and mechanical behavior of selective laser melted alloy 718. In *Proceedings of the 9th International Symposium on Superalloy 718 & Derivatives: Energy, Aerospace, and Industrial Applications* (pp. 89-113). Springer, Cham.

[25] Montgomery, D.C. and Runger, G.C., 2010. *Applied statistics and probability for engineers*. John Wiley & Sons.

[26] Craney, T.A. and Surles, J.G., 2002. Model-dependent variance inflation factor cutoff values. *Quality Engineering*, *14*(3), pp.391-403.

[27] Derksen, S. and Keselman, H.J., 1992. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, *45*(2), pp.265-282.

[28] Burham, K.P. and Anderson, D.R., 2002. Model selection and multimodel inference: A practical information-theoretic approach. Springer. *New York*.

[29] Hothorn, T., 2021. CRAN task view: Machine learning & statistical learning.