Machine Learning-Driven Optimization of Laser Powder Bed Fusion: From Powder Characterization to Process Parameters Tailoring for Ti6Al4V Powders

by

Farima Liravi

A thesis

presented to the University Of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2024

© Farima Liravi 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Ali Dolatabadi, Ph.D. Professor Department of Mechanical and Industrial Engineering University of Toronto
Supervisor:	Ehsan Toyserkani, Ph.D. Professor Department of Mechanical and Mechatronics Engineering University of Waterloo
Internal Member:	William Melek, Ph.D. Professor Department of Mechanical and Mechatronics Engineering University of Waterloo
Internal Member:	Peng Peng, Ph.D. Assistant Professor Department of Mechanical and Mechatronics Engineering University of Waterloo
Internal-External Member:	Nasser Lashgarian Azad, Ph.D. Professor Department of Systems Design Engineering University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contribution

The following chapters, 1, 2, 5, 6, and 7, were authored exclusively by Farima Liravi under the supervision of Prof. Ehsan Toyserkani and were not intended for publication. Portions of this thesis include two manuscripts prepared for publication:

Chapter 3:

The study was conceptualized and designed by Farima Liravi, Prof. Ehsan Toyserkani, and Dr. Mahdi Habibnejad. Farima Liravi carried out the data collection and analysis. The interpretation of results was a collaborative effort between Farima Liravi, Prof. Toyserkani, and Dr. Habibnejad. The manuscript was drafted by Farima Liravi, with contributions from Prof. Toyserkani. All authors reviewed the findings and approved the final version of the manuscript.

Citation for Chapter 3:

Liravi F, Habibnejad-Korayem M, Toyserkani E. On the performance of expert-augmented machine learning with limited experimental data collected from powder particle characteristics used in laser powder bed fusion. The international journal of advanced manufacturing technology. 2024.

Chapter 4:

The conceptualization of the study was carried out by Farima Liravi and Prof. Ehsan Toyserkani, with methodology developed by Farima Liravi and Prof. Ehsan Toyserkani. Software development was handled by Farima Liravi and Sahar Toorandaz, while validation was performed by Farima Liravi and Sebastian Soo. Formal analysis and investigation were conducted by Farima Liravi. Resources were provided by Dr. Mahdi Habibnejad-Korayem, and data curation was managed by Farima Liravi, Sahar Toorandaz, and Dr. Katayoon Taherkhani. The original draft was written by Farima Liravi, with review and editing contributions from Dr. Katayoon Taherkhani, Sahar Toorandaz, Dr. Mahdi Habibnejad-Korayem, and Prof. Ehsan Toyserkani. Visualization was done by Farima Liravi, with supervision and project administration by Prof. Ehsan Toyserkani. Funding was secured by Prof. Ehsan Toyserkani. All authors reviewed and approved the final manuscript.

Citation for Chapter 4:

Liravi F, Soo S, Toorandaz S, Taherkhani K, Habibnejad-Korayem M, Toyserkani E. Tailoring Laser Powder Bed Fusion Process Parameters for Standard and Off-Size Ti6Al4V Metal Powders: A Machine Learning Approach Enhanced by Photodiode-Based Melt Pool Monitoring. Invensions. 2024.

Abstract

The quality of parts produced by laser powder bed fusion (LPBF) is influenced by a complex interplay of factors, including process parameters, feedstock powder behavior, melt pool conditions, and environmental variables such as gas flow direction. Developing a comprehensive quality assurance system for LPBF necessitates incorporating these diverse factors into the analysis.

Despite their significance, the intrinsic characteristics of the powder—particularly powder bed quality—have rarely been the primary focus of global research. Key attributes such as particle size, morphology, and rheology are crucial in determining powder bed quality and laser-powder interaction. Therefore, accurately estimating powder rheology or flowability before printing is crucial for achieving consistent, high-quality production.

This thesis explores the use of machine learning (ML) algorithms to predict powder flowability based on simpler characteristics such as size, morphological features, and packing, without resorting to costly and time-consuming experiments. It focuses on a sample set of standard and off-size Ti6Al4V (Ti64) powders, with two main objectives: 1) assess ML regressors and equation fitting for flowability prediction, using the small original dataset, and 2) overcome data limitations' impact on regression through data augmentation techniques. Despite the limitations of the original dataset, the support vector regressor (SVR) delivered excellent performance in repeated random cross-validation (CV) analysis for the dynamic rheology characteristic of Specific Energy (SE), achieving a mean absolute percentage error (MAPE) of approximately 2.6% across six random CV iterations. Furthermore, the application of genetic programming's symbolic regression significantly enhanced the estimation accuracy of rheological characteristics, specifically the unconfined yield strength at 3 kPa (UYS3) and 9 kPa (UYS9), using the original dataset. The Synthetic minority over-sampling technique (SMOTE) proved effective in achieving admissible regression results for characteristics such as basic flow energy (BFE) and Hall flowability, whereas other explored methods of simple ML models training, and GP-SR equation-fitting had been unsuccessful.

After employing machine learning for powder characterization, it is crucial to integrate the identified powder characteristics into the optimization of LPBF process parameters. From an application perspective, selecting the appropriate set of process parameters is often influenced by the size distribution range of the powder used and the specific properties of the target part to be optimized.

Therefore, incorporating the powder size range into systems developed for tailoring and optimizing process parameters is highly beneficial.

Consequently, this thesis aimed to develop a comprehensive system for optimizing LPBF process parameters by integrating photodiode-based melt pool monitoring (light intensity signals) with ML. The proposed approach employs regression algorithms to determine optimal process parameters, including laser power, scan velocity, hatch distance, and energy density. These parameters are tailored based on the size range of the Ti64 powder, melt pool/light intensity signatures, and the desired properties of the final part, such as density, hardness, and surface roughness. Three different regressor algorithms of feed-forward neural network (FFN), random forest (RF), and extreme gradient boosting (XGBoost) were optimized and trained for the proposed task. Comparison analyses were conducted between algorithms in terms of regression accuracy and generalization capabilities. It was concluded that while all three algorithms could achieve satisfactory results on this dataset, the RF algorithm exhibited higher overall accuracy and better generalization. RF yielded an overall average MAPE of ~2% (R^2 of ~94%) in unseen samples analysis.

Acknowledgements

The completion of this research would not have been possible without the support and assistance of many individuals, to whom I would like to express my heartfelt gratitude.

I would like to express my heartfelt gratitude to my esteemed supervisor, Professor Ehsan Toyserkani, for his unwavering support, insightful guidance, valuable advice, and constant encouragement throughout the course of this work. His mentorship has provided me with the opportunity to enhance my scientific knowledge at the University of Waterloo.

I would like to acknowledge my thesis examining committee members: Prof. Ali Dolatabadi from the University of Toronto, and Prof. William Melek, Prof. Peng Peng, and Prof. Nasser Lashgarian Azad from the University of Waterloo. I appreciate their time and effort in reviewing my thesis and providing valuable feedback and suggestions.

I would also like to extend my gratitude for the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Network for Holistic Innovation in Additive Manufacturing (HI-AM), the Department of Mechanical and Mechatronics Engineering at the University of Waterloo, and Women in Engineering at the University of Waterloo.

This academic experience has highlighted the importance of interdisciplinary collaboration. I had the distinct pleasure of working with Dr. Mahdi Habibnejad-Korayem from AP&C company, as we collectively pursued a shared goal.

I would like to thank my current and past colleagues at the Multi-scale Additive Manufacturing (MSAM) Laboratory research at the University of Waterloo, Sahar Toorandaz, Sebastian Soo, Dr. Katayoon Taherkhani, and Nadia Azizi for their assistance and advice.

I would like to express my gratitude to the technical staff at the University of Waterloo — Francis Dibia, Jerry Ratthapakdee, and Grace Kurosad— for their outstanding support and assistance with the experimental tests.

Lastly, I am deeply thankful to my family for their constant love, assistance, and support throughout this journey.

Table of Contents

Examining Committee Membership ii

Author's Declaration iii

Abstract vi

Acknowledgements viii

List of Figures xiii

List of Tables xvii

List of Abbreviations xviii

List of Symbols xix

Chapter 1 Introduction 1

1.1 Research Motivations 1

1.2 Research Objectives 3

1.3 Thesis Outline 7

Chapter 2 : Literature Review 9

2.1 Laser Powder Bed Fusion 9

2.1.1 LPBF Process Parameters 10

2.1.2 LPBF Powder Feedstock Characteristics 10

2.1.3 LPBF Product Quality Assurance 13

2.1.4 LPBF Powder Characteristics and Product Quality: Literature Review 13

2.2 LPBF Process Parameters and Product Quality: Literature Review 17

2.2.1 Correlation Between LPBF Process Parameters and Physical Properties of Final Product 18

2.2.2 Correlation Between LPBF Process Parameters and Surface Roughness of Final Product 19

2.3 In-Situ Melt Pool Monitoring of LPBF 20

2.4 Application of Machine Learning in LPBF Quality Assurance Systems 23

2.4.1 Machine Learning Background....24
2.4.2 Application of Machine Learning in Powder Characterization....24
2.4.3 Application of MPM and ML for In-Situ Process Optimization....27
2.5 Summary....33
Chapter 3 : On the performance of expert-augmented machine learning with limited experimental data collected from powder particle characteristics used in laser powder bed fusion....34
3.1 Abstract....34
3.2 Methodologies and experimental setups....34
3.2.1 Machine Learning Algorithms and Data Preprocessing....37
3.2.2 Data Limitation Challenge and Remedies....38
3.3 Results and Analysis Discussion....39
3.3.1 Flowability Prediction Using the Original Dataset....40
3.3.2 Flowability Prediction Using Data Augmentation....57
3.4 Summary....66
Chapter 4 : Tailoring laser powder bed fusion process parameters for standard and off-size Ti6Al4V metal powders: a machine learning approach enhanced by photodiode-based melt pool monitoring..67
4.1 Abstract....67
4.2 Methodology and experimental setups....67
4.2.1 Design of experiment....67
4.2.2 Melt Pool Monitoring....69
4.2.3 Machine Learning Algorithms....70
4.3 Results and Discussion....75
4.3.1 Experimental Data Analysis....75
4.3.2 Machine Learning Analysis....79

4.4 Summary 90

Chapter 5 : Complementary Validation of Optimal Powder Rheology Regressors on Unseen Off-size Ti6Al4V Powder Samples 91

5.1 Abstract 91

5.2 Methodology and experimental setups 91

5.2.1 Powder Rheology Estimation 91

5.3 Results and Discussion 92

5.3.1 Dynamic Rheology 92

5.3.2 Shear Rheology 95

5.3.3 Hall Flowability 98

5.4 Summary 99

Chapter 6 : Application of the Optimal Melt Pool Monitoring-enhanced Process Parameters Tailoring Algorithms on Complex Artifacts 100

6.1 Abstract 100

6.2 Methodology and experimental setups 100

6.3 Results and discussion 101

6.4 Summary 103

Chapter 7 : Conclusions and Future Work 105

7.1 Conclusions of ML-based Powder Rheology Characterization 105

7.1.1 Dynamic Rheology: 105

7.1.2 Shear Rheology: 106

7.1.3 Hall Flowability: 106

7.2 Conclusions of ML-based Process Parameters Tailoring Using Photodiode Melt Pool Monitoring 107

7.3 Future Work 109

References 121

Appendix A 134

Appendix B : Machine Learning Algorithms 136

Appendix C 141

List of Figures

Figure 1-1: Graphical abstract of research's first objective and methodology. 5
Figure 1-2: Graphical abstract of research's second objective and methodology. 6
Figure 2-1: Schematic representation of laser powder bed fusion process [14]. 9
Figure 2-2: Schematic representation of laser powder bed fusion process parameters [15]. 10
Figure 2-3: Microtrac Camsizer powder size and shape measurement metrics [18]. 11
Figure 2-4: (a) 3D Visualization of porosities in coarse and fine powder sample produced parts built in two locations on build plate, (b) histogram of volumetric defect concentration in coarse and fine produced samples [24]. 15
Figure 2-5: Effect of hatch space on ripple [50]. 19
Figure 2-6: Melt pool monitoring in laser powder bed fusion process [51]. 20
Figure 2-7: Melt pool monitoring using photodiode sensor: (a) raw intensity signals of prints, (b) moving average smoothed intensity signal transformed to spatial map, and (c) window moving averaged intensity showing dynamic of melt pool [54]. 21
Figure 2-8: (a) Off-axis sensor mounting arrangement (sensor in chamber), (b) off-axis sensor mounting arrangement (sensor out of chamber), and (c) on-axis sensor mounting arrangement [55]. 22
Figure 3-1: Graphical abstract of the analysis (equipment images adapted from [100], [101], [102], [103]). 35
Figure 3-2; Overall comparison of various ML algorithms' performance (performed on the train samples of the original dataset) in leave-one-out CV for (a) SE, (b) BFE, (c) UYS3, (d) UYS6, (e) UYS9, and (f) hall flowability. 43
Figure 3-3: Performance of SVR regressor in prediction of dynamic rheology targets. (a) predicted SE vs. the actual SE, (b) MAPE of SE prediction in SSP mode, (c) predicted BFE vs. the actual BFE, and (d) MAPE of BFE prediction in SSP mode. 45
Figure 3-4; Assessment of the SVR regressor's performance (using the original dataset) across three performance evaluation modes for SE and BFE target regression. 47
Figure 3-5: Performance of the SVR regressor (using the original dataset) on each UYS target in SSP mode. The plot shows the MAPE of predictions averaged across all samples. 48
Figure 3-6: Estimations made by the GP-SR equations compared to the actual values of the sample for (a) UYS3, (b) UYS6, and (c) UYS9 targets in SSP mode. 50

Figure 3-7: Assessment of GP-SR performance by analyzing the error in UYS3 estimations and the uncertainty in the experimental measurements per sample. (a) GP-SR estimation errors, and (b) comparison of estimated values to the min-max interval and deviation of the experimental measurements. 51
Figure 3-8: Illustration of (a) shear behavior variations (for samples starting from the 15 μm size range) as the PSD range increases, and (b) the compressibility index of each corresponding sample. 52
Figure 3-9: Assessment of GP-SR performance by analyzing the error in UYS6 estimations and the uncertainty in the experimental measurements per sample. (a) GP-SR estimation errors, and (b) comparison of estimated values to the min-max interval and deviation of the experimental measurements. 54
Figure 3-10: Assessment of GP-SR performance by analyzing the error in UYS9 estimations and the uncertainty in the experimental measurements per sample. (a) GP-SR estimation errors, and (b) comparison of estimated values to the min-max interval and deviation of the experimental measurements. 55
Figure 3-11: Illustration of (a) shear behavior variations (for samples starting from the 45 μm size range) as the PSD range increases, and (b) the compressibility index of each corresponding sample. 56
Figure 3-12: Overall performance comparison between SVR and GP-SR techniques (using the original dataset) in estimating the shear rheology targets in SSP mode. 57
Figure 3-13: Overall comparison of the SVR algorithm's performance in BFE regression across two repetitions of 9-fold CV, using the original dataset versus the SMOTE-augmented dataset. 60
Figure 3-14: Overall performance comparison between the SVR algorithm in UYS6 regression using the SMOTE-augmented dataset and the GP-SR algorithm using the original dataset, based on two repetitions of 9-fold CV. 61
Figure 3-15: Overall comparison of various ML algorithms' performance (trained on SMOTE augmented dataset) in hall flowability regression. 62
Figure 3-16: Overall performance comparison of the SVR regressor in hall flowability prediction across three evaluation modes using the SMOTE-augmented dataset. 64
Figure 3-17: Error of hall flowability regression in unseen samples using the SVR regressor on the SMOTE augmented data. 65
Figure 4-1: Printed cubical and cylindrical parts. 68

Figure 4-2: Light intensity collection using an on-axis photodiode, processing, and clustering using SOM. 70
Figure 4-3: Graphical abstract highlighting the study workflow including application of ML on MPM data, part properties, and different powder particle size distributions for process parameters estimation. 72
Figure 4-4: Variation of parts physical characteristics in cubical ([0–53] on x-axis) and cylindrical ([54–108] on x-axis) parts with change of powder size. The process parameters per each sample in different powder batches are identical (to observe the impact of powder size change). The part characteristics variations are shown in (a) surface roughness, (b) density, and (c) hardness. 76
Figure 4-5: The average intensity level per layer is plotted for cubical and cylindrical parts with the highest and lowest surface roughness/density in three powder batches of 15–53 µm (A), 15–106 µm (B), and 45–106 µm (C). The average intensity per layer in parts is plotted for (a) cubical parts with highest and lowest surface roughness, (b) cubical parts with highest and lowest density, (c) cylindrical parts with highest and lowest surface roughness, and (d) cylindrical parts with highest and lowest density. 78
Figure 4-6: Impact of SOM cluster numbers on overall averaged performance of three regressors showed by average MAPE metric obtained in five-fold CV test implemented on train set samples. .. 81
Figure 4-7: Average (a) R^2 score and (b) MAPE of the regressor algorithms in CV analysis implemented on the train set of data. 82
Figure 4-8: Notched box plot of regression errors of three regressor algorithms obtained in CV analysis implemented on the train set. The results are shown for (a) power (P), (b) velocity (V), (c) hatch distance (HD), and (d) energy density (ED). 84
Figure 4-9: Impact of train set size on the overall performance of three regressors in the estimation of process parameters. The overall error is shown by the $\log_{10}$(MAPE) metric. 85
Figure 4-10: The impact of epoch numbers on FFN performance was tested and shown by the MSE metric. The algorithm was tested on 80% of the initial train set as the new train set and 20% of the initial train set as the validation set. 86
Figure 4-11: Prediction vs. actual target plots obtained from the unseen samples' regression in the four targets. Prediction accuracy of regressions is also shown by MAE and MAPE for (a) FFN regressor (b) RF regressor, and (c) XGBoost regressor. Power, velocity, hatch distance, and energy density are denoted by (P), (V), (HD), and (ED), respectively. 88

Figure 4-12: Probability density distribution of predicted values and actual target values in unseen samples regression resulted by (a) FFN (b) RF, and (c) XGBoost regressor. Power, velocity, hatch distance, and energy density are denoted by (P), (V), (HD), and (ED), respectively. 89
Figure 5-1: Final evaluation of the SVR algorithm in BFE regression using new off-size powder samples. The performance of (a) the developed SVR model compared to the re-tuned SVR model on the original test set, and (b) the developed model with/without the outlier sample (50-63 µm), on the new test set, in three evaluation modes of regression. 93
Figure 5-2: Final evaluation of the SVR algorithm in SE regression using new off-size powder samples. The performance of (a) the developed SVR model compared to the re-tuned SVR model on the original test set, and (b) the developed model with/without the outlier sample (50-63µm), on the new train set, in three evaluation modes of regression. 94
Figure 5-3: (a) Estimated UYS3 values by the developed GP-SR equation in three new powder samples and (b) comparison of UYS3 measures of the new samples relative to the trained range. 96
Figure 5-4: The performance of the developed model with/without the outlier sample (50-63 µm), on the new test set, in three evaluation modes of regression. 96
Figure 5-5: (a) Estimated UYS3 values by the developed GP-SR equation in three new powder samples and (b) comparison of UYS6 measures of the new samples relative to the trained range. 97
Figure 5-6: Final evaluation of the SVR algorithm in hall flowability regression using new off-size powder samples. The performance of (a) the developed SVR model compared to the re-tuned SVR model on the same test set, and (b) the developed model on the new test set, in three evaluation modes of regression. 98
Figure 6-1: Artifacts printed using each of the three batches (15-53 µm, 15-106 µm, and 45-106 µm) of Ti64 powder feedstock. 101
Figure 6-2: the performance of the RF and XGBoost regressors on the artifacts in estimation of the (a) power, (b) velocity, (c) hatch distance, and (d) energy density. 102
Figure 6-3: Performance of the XGBoost in the estimation of the power of the artifacts. 102
Figure 6-4: Comparative performance between the established RF and XGBoost regressors (trained on cubical and cylindrical parts dataset) in the estimation of process parameters of the artifacts 103

List of Tables

Table 2-1: Summarization of powder properties attained from FT4 powder characterization technology 12
Table 2-2: Literature review of studies focused on powder characteristics and final part quality correlations 15
Table 2-3: Commercial melt pool monitoring systems 22
Table 2-4: Table of literature review of machine learning and melt pool monitoring application for process inspection and optimization 29
Table 2-5: Table of literature review of machine learning application for process inspection and optimization 31
Table 3-1: Particle size distribution ranges of Ti64 powder 35
Table 3-2: Authors-selected set of powder properties used as the input features of the regressors 36
Table 3-3: Tuned hyper-parameters of SVR regressors (using the original dataset) used for the prediction of SE and BFE targets. 44
Table 3-4: Tuned hyper-parameters of SVR regressors used for the prediction of UYS3, UYS6, and UYS9 targets using the original dataset 47
Table 3-5: Tuned hyper-parameters of SVR regressors used for BFE target using the SMOTE augmented data 59
Table 3-6: Tuned hyper-parameters of SVR regressors used for UYS6 target using the SMOTE augmented data 60
Table 3-7: Tuned hyper-parameters of SVR regressors used for the prediction of hall flowability target using the original dataset 61
Table 3-8: Optimum set of features obtained by sequential feature selection for Hall flowability regression using the original dataset 61
Table 3-9: Tuned hyper-parameters of SVR regressors used for Hall flowability target using the SMOTE augmented data 63
Table 3-10: Optimum set of features obtained by the sequential feature selection for Hall flowability regression using the SMOTE augmented dataset 63
Table 4-1: Design of the experiment 68
Table 4-2: FFN-tuned hyper-parameters and the final structure of the algorithm. L1 and L2 denote the first hidden layer and the second hidden layer, respectively 74
Table 4-3: RF and XGBoost tuned hyper-parameters 75

List of Abbreviations

Adaptive Boosting..... (Adaboost)
Additive Manufacturing..... (AM)
American Society For Testing And Material .....(ASTM)
Application Programming Interface..... (API)
Artificial Intelligence ..... (AI)
Artificial Neural Networks.....(ANN)
Autoencoder ..... (AE)
Back Propagation Neural Network .....(BPNN)
Binder-Jetting..... (BJ)
Computer-Aided Design ..... (CAD)
Conditional Generative Adversarial Networks ..... (CGAN)
Convolutional Neural Network ..... (CNN)
Decision Trees..... (DT)
Design Of Experiments..... (DOE)
Directed Energy Deposition..... (DED)
Electron Beam Melting .....(EBM)
Extreme Gradient Boosting.....(Xgboost)
Feed-Forward Neural Network ..... (FFN)
Flow Function .....(FFC)
Freeman Technology.....(FT4)
Full Factorial Design..... (FFD)
Gas Atomization.....(GA)
Gaussian Mixture Model.....(GMM)
Gaussian Process Regressor.....(GPR)
Genetic Programming .....(GP)
Gradient Boost .....(GB)
Gradient Boosting Regressor ..... (GBR)
Infrared.....(IR)
Karush-Kuhn-Tucker ..... (KKT)
K-Nearest Neighbors.....(KNN)
Kolmogorov–Smirnov.....(KS)
Laser Powder Bed Fusion .....(LPBF)
Long Short-Term Memory.....(LSTM)
Machine Learning .....(ML)
Material Extrusion.....(ME)
Material Jetting.....(MJ)
Mean Absolute Error.....(MAE)
Mean Squared Error.....(MSE)
Melt Pool Monitoring .....(MPM)
Multilayer Perceptron .....(MLP)
Multiple-Instance Regression ..... (MIR)
Naïve Bayesian ..... (NB)
National Institute Of Standards And Technology .....(NIST)
Near-Infrared .....(NIR)
Partial Least Squares.....(PLS)
Particle Size Distribution.....(PSD)
Plasma Rotating Electrode Process..... (PREP)
Probability Density Function.....(PDF)
Quality Assurance..... (QA)
Radial Basis Function ..... (RBF)
Random Forest..... (RF)
Regression Tree .....(RT)
Reinforcement Learning .....(RL)
Repeated Random Split Prediction..... (RRSP)
Rotary Atomization ..... (RA)
Scanning Electron Microscopy.....(SEM)
Self-Organizing Maps..... (SOM)
Single Sample Prediction..... (SSP)
Speed Up Robust Features.....(SURF)
Stacked Sparse Autoencoder ..... (SSAE)
Support Vector Classifier ..... (SVC)
Support Vector Machine..... (SVM)
Support Vectors .....(SV)
T-Distributed Stochastic Neighbor Embedding .....(T-SNE)
The American Society For Testing And Material.....(ASTM)
The National Institute Of Standards And Technology .....(NIST)
Ti6Al4V ..... (Ti64)
Ultimate Tensile Stress ..... (UTS)
Unconfined Yield Strength ..... (FC)
Water Atomization..... (WA)
Ytterbium ..... (Yb)

List of Symbols

ε	*Error insensitive tube in support vector regression*
ξ	*Error assigned to samples outside ε tube*
W	*Coefficients of SVR fitting function*
C	*Regulation parameter*
n	*Number of samples*
b	*Bias*
γ	*Gamma parameter (samples' area of influence)*
$\alpha_i, \alpha_j, \alpha_i^*, \alpha_j^*$	*Lagrange multipliers*
D10, D50, D90	*Particle size distribution percentile values*
Span	$\frac{D90-D10}{D50}$
Carr's index	$\frac{(tapped\ density-bulk\ density)}{tapped\ density} \times 100$
$H_{ik}(t)$	*Neighbourhood kernel function*
D_{ik}	*Distance (Euclidean)*
$\sigma(t)$	*Neighbourhood size decay*
T, t	*Time*
W_k	*Weight of nodes*
$\eta(t)$	*Learning rate decay*
UYS3	*Unconfined yield strength of powder bulk at 3kPa consolidation in FT4 shear cell test*
UYS6	*Unconfined yield strength of powder bulk at 6kPa consolidation in FT4 shear cell test*
UYS9	*Unconfined yield strength of powder bulk at 9kPa consolidation in FT4 shear cell test*

Chapter 1
Introduction

1.1 Research Motivations

Additive manufacturing (AM) has been extensively adopted as an advanced manufacturing approach in recent years. AM is based on a layer-by-layer construction of parts using predetermined geometric digital models that have been designed using computer-aided design (CAD) programs. Laser powder bed fusion (LPBF) is one of the most common processing methods of AM for metal part manufacturing. The layer-wise part production in LPBF allows for the fabrication of complex geometries with accurate dimensions and high quality in terms of mechanical properties and surface roughness. Despite its advantages, LPBF still faces some challenges before it can become a viable alternative to conventional fabrication methods in mainstream manufacturing. The majority of cases where LPBF is deployed over conventional manufacturing methods involve low-volume production and prototyping [1]. Additionally, every involved factor in the LPBF process has a major impact on the final part quality due to the interdependency of the parameters. As a result of the complex relationship between all the quality-relevant parameters, the development of quality assurance (QA) platforms has become one of the most important requirements for the commercialization of AM. Developing reliable QA systems to detect, analyze, and optimize the parameters that influence final product performance and quality is key to overcoming these challenges. A proper QA system must ensure process repeatability and high-quality product reproducibility [2]. Mechanical and physical properties of LPBF-made products are key indicators in the quality compliance objectives. Mechanical properties are highly influenced by process parameters such as laser power, scan velocity, scanning strategies, and layer thickness. Additionally, feed-stock powder characteristics impact the printing process in terms of powder flowability, recoating, and powder bed compaction density, which, in turn, affect the final mechanical properties. Accordingly, it is vital to monitor both feedstock characteristics and printing process parameters to ensure optimal part quality.

Powder characteristics are critical in determining the quality of AM-made parts requiring advanced and efficient technologies for investigation. Uniformity, thinness, and high packing density are three acknowledged attributes for obtaining an optimized layer of powder in the LPBF process. Morphological factors of powder such as particle size distribution (PSD) , shape, and surface roughness induce different flowability and density in the powder spreading step of the building; therefore, the

optimal combination of powder properties should be inspected for each type of material to reach the most advantageous state in build and thereupon in the manufacturing outcome quality.

Powder characterization techniques, traditionally an experimental process, is time-consuming, costly, and prone to human error. Despite these limitations, few studies have explored smart and automated methodologies for accurate powder characterization. Machine learning (ML) applications in powder characterization remain largely untapped, with only a few studies addressing this area. Moreover, there is limited information on linking powder characteristics to the quality of LPBF-made components, thus creating a research gap.

This thesis aims to address these gaps by predicting the flowability of Ti-6Al-4V (Ti64) metal powder using ML and linking the properties to process signature identified through the melt pool monitoring. The main challenge of this procedure is dataset limitations of powder characterizations due to the time and cost-intensive process of powder sample generation, sieving, and characterization. It is attempted in this research to mitigate the lack of data by adopting ML algorithms suitable for small datasets and methods for mathematically modeling underlying data relationships. Ultimately, a simple yet practical data augmentation technique was incorporated to assess its feasibility and effectiveness in cases where previously examined methods failed to yield acceptable ML-based estimations of powder rheology or flowability.

In addition to the importance of estimating powder bed behavior through an initial yet admissible estimation of powder rheology, melt pool stability is another frequently overlooked factor in the development of LPBF QA systems. The process parameters directly impact the dynamic of the laser-powder bed interaction, the fusion of the powder particles, and subsequently the molten pool stability and characteristics. The characteristics of the melt pool would then influence the solidification of the molten pool which in turn affects the microstructural properties such as roughness and mechanical properties such as density and hardness. Accordingly, monitoring and integrating the melt pool characteristics into the QA systems would allow for more precise quality estimation and corrective decision-making.

In-situ process monitoring, and control are critical solutions for the optimization and assessment of the final product characteristics by inspection of real-time melt pool condition. Such techniques allow for the development of advanced closed-loop control systems for real-time process correction and optimization. Several technologies are being broadly used for in-situ process control that predominantly

rely on various kinds of sensing instruments such as thermal, optical, and acoustic sensors [3]. Artificial intelligence (AI) assisted inspection of light intensity irradiated from melt pool of printing parts or melt pool monitoring (MPM) is one of the promising approaches to process control. This technique is largely practiced for the detection and localization of different defects such as the lack of fusion porosity, along with examination of dimensional and geometrical accuracy of printing [4]. Optimization of process parameters is usually implemented by application of design of experiments (DOE) methods, followed by highly time-consuming experimental characterization of printed parts. These process optimization procedures are also followed by a high amount of resource and material waste, as the evaluation of the quality of the product is implemented after the completion of full part fabrication. Real-time process inspection using image-based or time-series data (3D, 2D, or 1D signals) inspires the idea of utilizing ML for the LPBF parameters optimization based on the real-time signals collected from the melt pool and surrounding areas. One of the challenges of ML adaption for 3D/2D data processing is the high volume of data which induces computational complexity. This challenge is proposed to be addressed in this research by the employment of photodiode sensors (low dimensional data) and ML model regulation for optimum data processing. To address the identified gap in existing ML-based QA systems, which usually do not involve the melt pool quality and the powder feedstock characteristics in process parameters optimization, this thesis trained and evaluated various ML regressors. These regressors were designed to predict process parameters by incorporating comprehensive quality-impacting factors, such as the statistical signatures of the melt pool and the characteristics of the powder feedstock.

1.2 Research Objectives

The goal of this thesis was to investigate the less-considered yet highly influential aspects of the LPBF process, particularly the characteristics of the powder feedstock. Rheological powder properties, like shear rheology (which indicates the powder flowability under consolidated conditions with another powder layer), enable a more precise estimation of powder bed behavior. However, measuring these properties experimentally is even more time-consuming than measuring size, morphology, and packing. Therefore, as depicted in Figure 1-1, the first goal of the study was to explore the feasibility of applying ML algorithms and data augmentation techniques to estimate the flowability and rheological properties of Ti64 based on simpler powder characteristics, such as size, morphology, and bulk packing, while working with a limited dataset.

The second goal was to develop a LPBF process parameters tailoring algorithm for Ti64 powder, enhanced by MPM. Multiple regression algorithms were trained and evaluated to predict the suitable set of process parameters—laser power, scan velocity, hatch distance, and energy density—based on melt pool signatures (extracted from light intensity signals recorded by a photodiode sensor), the Ti64 powder size range used for the print, and the optimized or customer-defined part properties of density, hardness, and surface roughness. The graphical abstract of the methodology in second goal of the research is also depicted in Figure 1-2. The steps taken to fulfill these goals are outlined below:

- Training simple regressor algorithms practical for small dataset using the available limited dataset (original dataset), to predict the powder flowability (hall flowability) and the rheological properties such as dynamic flow and shear flow.
- Exploring the innovative equation fitting algorithms to provide mathematical equations for estimation of powder flowability in cases when conventional ML algorithms trained on the original dataset failed to provide acceptable results.
- Applying data augmentation to generate synthetic data based on the original dataset to train ML regressors on a combination of original and synthetic dataset as the final explored solution to compensate the data limitation.
- Developing design of experiment (full factorial) for print of cubical and cylindrical parts with optimum and non-optimum density, hardness, and surface roughness, using three different size distribution range of Ti64 powder feedstock (three sets of prints).
- Processing the in-situ light intensity signals (MPM data) recorded by a photodiode sensor to transfer the build signals to per-part data.
- Clustering of the per-part light intensity data using self-organizing maps (SOM) algorithm.
- Extraction of multiple statistical signatures from each cluster of light intensity data (per part).
- Compiling a dataset consisting of MPM statistical signatures (in multiple clusters), Ti64 powder feedstock size range, and part properties (density, hardness, and surface roughness) as the input features of ML regressors to predict the process parameters (laser power, scan velocity, hatch distance, and energy density) as the output features of the ML regressors.

- Comparison analyses among three different trained algorithms from different aspects of regression accuracy, generalization, and statistical examination of results.

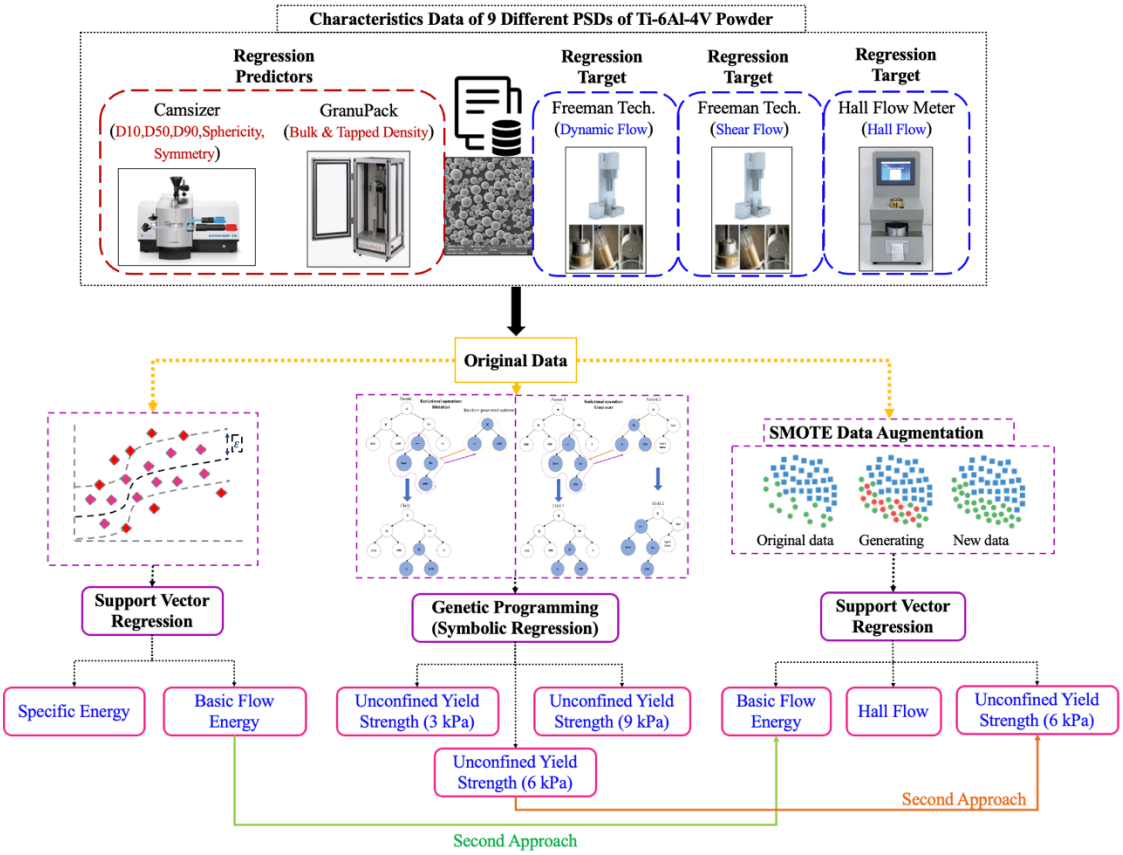


Figure 1-1: Graphical abstract of research's first objective and methodology.

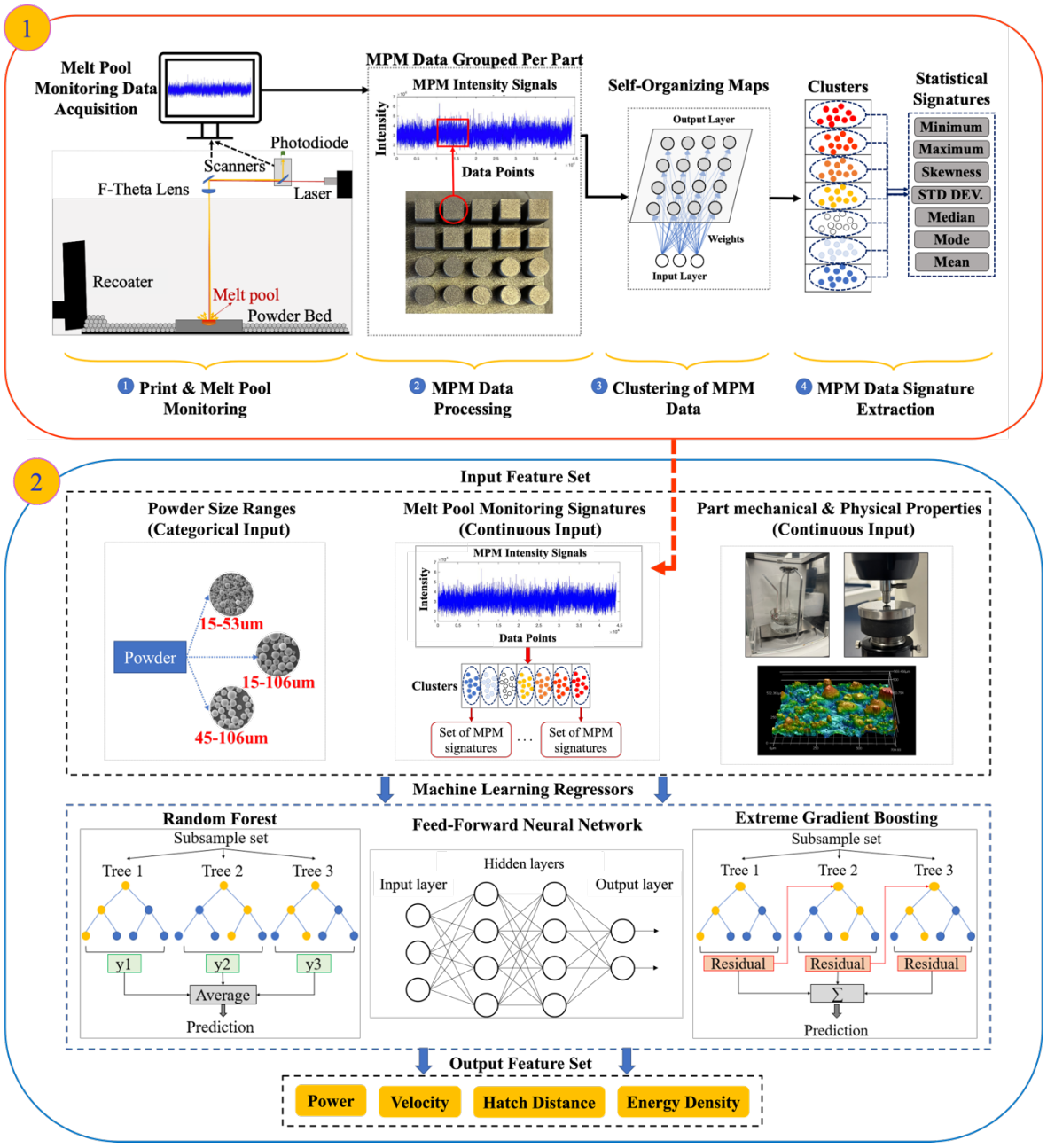


Figure 1-2: Graphical abstract of research's second objective and methodology.

1.3 Thesis Outline

The thesis is comprised of six chapters. Chapter 1 introduces the motivation for the study, the challenges faced, and the objectives outlined. Chapter 2 presents an extensive literature review, focusing on the impact of powder characteristics on the quality of Ti64 parts in LPBF and the data-driven techniques developed for metal powder characterization. This is followed by a review of the advantages of MPM in LPBF, the established MPM techniques, and studies that apply ML, MPM, or both in optimizing process parameters and quality control in LPBF. Chapter 3 examines and discusses the results obtained from exploring ML and data augmentation techniques for estimating the flowability of Ti64 powder using a highly limited dataset. Chapter 4 details the development and evaluation of regression algorithms trained and optimized to predict the process parameters for cubical and cylindrical Ti64 parts. These predictions are based on photodiode-based light intensity data (MPM), powder feedstock size range, and defined mechanical and microstructural properties of the parts. Chapter 5 assesses the practicality of the developed algorithms for predicting powder flowability on new, unseen sample sets. The powder characterization algorithms were tested on three new Ti64 powder samples with different, previously unseen particle size ranges. Chapter 6 examines the generalization capability of the established algorithms in Chapter 3 in process parameters prediction of complex artifact parts. The process parameters tailoring algorithm was evaluated by assessing the accuracy of the predicted process parameters for an artifact. Lastly, Chapter 7 outlines the conclusions of the research, addressing the defined objectives and discussing potential future work.

Chapter 3 to Chapter 4 are adapted from the author's published works and submitted manuscripts for publication, as follows:

- Chapter 3 is a journal article published in the international journal of advanced manufacturing technology:

  Liravi F, Habibnejad-Korayem M, Toyserkani E. On the performance of expert-augmented machine learning with limited experimental data collected from powder particle characteristics used in laser powder bed fusion. The international journal of advanced manufacturing technology. 2024. DOI: https://doi.org/10.1007/s00170-024-14044-2. [5]

- Chapter 4 is a journal article published in inventions journal:

Liravi F, Soo S, Toorandaz S, Taherkhani K, Habibnejad-Korayem M, Toyserkani E. Tailoring Laser Powder Bed Fusion Process Parameters for Standard and Off-Size Ti6Al4V Metal Powders: A Machine Learning Approach Enhanced by Photodiode-Based Melt Pool Monitoring. Inventions. 2024. DOI: https://doi.org/10.3390/inventions9040087. [6]

The content of these publications has been slightly modified to fit within the scope of this thesis. The license agreements are provided in the letter of copyright permissions section.

Chapter 2: Literature Review

AM has become increasingly popular for production of a wide variety of components as it permeates various industries, including healthcare and aerospace. AM uses an iterative process in which layers are added to previously fabricated layers based on predefined geometric patterns and build characteristics [7]. The layer-by-layer fabrication allows to produce lightweight parts with complex structures, high densities, and dimensionally accurate without sacrificing the physical properties and the quality of the product [8], [9]. This process will also minimize excessive resource and material waste compared to subtractive manufacturing [10]. Various AM methods can be used to make parts with particular functionality, including binder-jetting (BJ), directed energy deposition (DED), material extrusion (ME), material jetting (MJ), laser powder bed fusion (LPBF), sheet lamination, and vat photopolymerization [11]. Energy source, build procedure, hardware and device specifications, and feedstock material type vary among AM processes.

2.1 Laser Powder Bed Fusion

One of the most common AM methods used to fabricate metal parts is LPBF. As seen in Figure 2-1, the process constructs parts in a layer-by-layer approach, using a thin, dense, and homogeneous layer of powder. To fabricate a single layer of the part, a powder layer is selectively melted by laser radiation and cooled afterward. After spreading the new powder layer, successive layers of the part will be fabricated similarly. Near-net shape geometry, relatively high density, and favorable surface roughness, as well as the ability to manufacture complex shapes, are reported to be achievable by the LPBF process [12], [13].

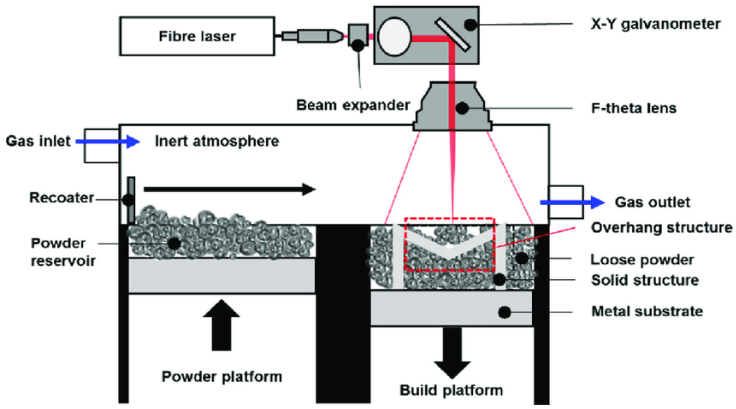


Figure 2-1: Schematic representation of laser powder bed fusion process [14].

2.1.1 LPBF Process Parameters

LPBF is governed by numerous build process parameters such as laser power, volumetric energy density, scanning velocity, hatch distance, scanning pattern, and layer thickness. Intensity of energy source required to melt the powder particles is determined by the laser power level. The laser scan pattern and velocity define the direction and speed of laser movement through the powder bed as seen in Figure 2-2.

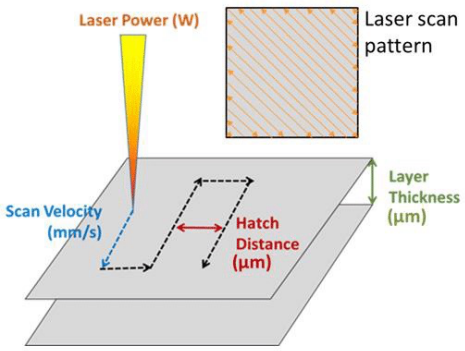


Figure 2-2: Schematic representation of laser powder bed fusion process parameters [15].

The set of process parameters determines the quality of printing and consequently the final part's physical properties, microstructure, and dimensional accuracy. Environmental factors such as chamber gas flow and build plate characteristics impact the conditions of the print and final quality, thus they are also considered as process parameters. It should however be noted that, depending on the materials, geometry, and applications, environmental parameters may not have a significant influence on the quality of the final product

The parameters involved in the process are interdependent and fully govern the mechanical properties of the final product in terms of density, hardness, surface roughness, and tensile/fatigue strength. As a result, a comprehensive process parameter set inspection, and optimization is essential to optimum print setup configurations and ultimately arrive at repeatable and reproducible production of high-quality parts.

2.1.2 LPBF Powder Feedstock Characteristics

LPBF can be applied to a wide range of metal powders that possess distinct physical, chemical, and rheological properties. The characteristics of powder feedstock are initially influenced by the

production system used, such as gas atomization (GA), plasma rotating electrode process (PREP), rotary atomization (RA), and water atomization (WA) [16]. PREP powder production system has been shown to have favorable features in terms of powder sphericity, surface smoothness, uniform particle distribution, cohesion, and flowability [16]. Powder physical/morphological properties such as PSD, shape (sphericity, symmetry, circularity, aspect ratio, … etc.), and surface texture have a significant impact on the behavior and flowability of powder in a powder bed (recoating step) and during the melting and solidification process. Thus, an effective build plan requires full powder characterization. The National Institute of Standards and Technology (NIST) has published multiple standard powder property measurement methodologies such as X-ray diffraction, and scanning electron microscopy (SEM), in addition to sieving procedures for powder size calibration [17]. Furthermore, the American Society for Testing and Material (ASTM) has developed several standards to determine powder particle shapes using its projections. Among the proposed metrics, sphericity, projected area ratio, and Ferret's diameter can be mentioned which are considered as quantitative powder shape identification measurements [17]. Some shape characteristics and standard measurement equations used for characterization can be seen in Figure 2-3.

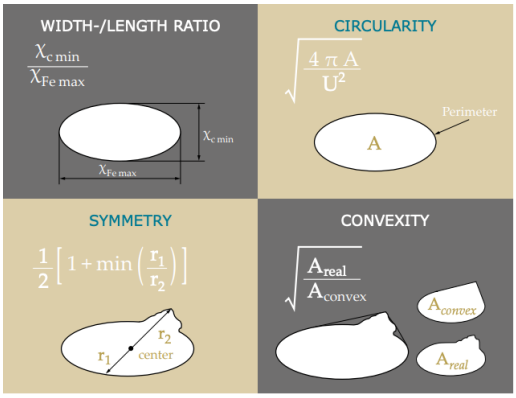


Figure 2-3: Microtrac Camsizer powder size and shape measurement metrics [18].

In addition, powder flowability/rheology and powder bulk packing characteristics are two other factors that majorly affect the powder layer homogeneity and density in LPBF. Flowability of powder varies based upon many internal and external influencing factors. Aside from powder morphological properties which impact powder cohesion and flowability, external parameters such as compression of bulk of powder also notably changes rheological behavior of the powder. Accordingly, an experimental measurement of shear resistance in consolidated powder is performed by applying yield locus curves

to the results of normal stress and shear stress measures, using the freeman technology (FT4) equipment.

Dynamic flowability (powder flowability in motion) of powder is also measured by FT4 which calibrates properties such as basic flow energy (BFE) and specific energy (SE). A brief review of FT4 tests and associated powder rheological properties is presented in Table 2-1.

Table 2-1: Summarization of powder properties attained from FT4 powder characterization technology

Powder Characterization Class	Powder Property Name	Description
Dynamic flow	**Basic flow energy**	Unconfined flow in motion (blade moving downward)
	Specific energy	Unconfined flow in motion (blade moving upward)
	Consolidation	Flow of consolidated powder in motion
	Aeration	Energy needed for inserted air flow to overcome cohesive forces between particles
	Flow rate	Sensitivity of flowability of powder to rate (speed) of flow force
Bulk	**Density**	Apparent packing of bulk of powder
	Compressibility	Change in volume of powder bulk after compression
	Permeability	Powder resistance to air flow in constrained condition provided by different levels of inserted applied normal stresses.

Shear	**Unconfined yield strength**	Resistance of powder to shear against another layer of powder after pre-consolidation and release of consolidation pressure.
	Major principal stress	Total amount of stresses applied on powder in any possible planes.
	Cohesion	Powder cohesiveness
	Angle of internal friction	Angle between normal stress and shear stress in moment of failure or flow.

2.1.3 LPBF Product Quality Assurance

Mechanical properties (density, hardness, ductility, tensile strength, fatigue, etc.…), surface roughness, potential defects, geometry and dimensional features, and residual stress of LPBF final product is highly influenced by both powder properties and build process parameters [19]. In addition to influencing the quality of the final parts, both the powder characteristics and the process parameters are closely related. It is critical for QA systems to assess the quality of AM products reliably, while promoting the repeatability and reproducibility of high-quality AM products. Consequently, automated, accurate, and efficient systems are necessary for constant monitoring and tuning of material characteristics and process parameters, individually and in conjunction. The development of a reliable and effective LPBF quality assurance system is highly dependent upon comprehensive identification of relationships between powder properties, process parameters, and mechanical properties. Many studies have been conducted addressing these objectives, utilizing different methods such as statistical, analytical, and AI-based approaches.

2.1.4 LPBF Powder Characteristics and Product Quality: Literature Review

Several studies have examined the impact of metal powder characteristics on LPBF-fabricated metal parts' properties. Haferkamp *et al.* studied the relation between powder shape and circularity, powder flowability, and powder layer density in 316L and AlSi10Mg metal powders. The study concludes that Higher powder sphericity increases the flowability and density of the powder layers in the powder bed,

and therefore, the final parts have greater density [20]. In another investigation, Riener Dipl.-Ing *et al.* implemented analysis using four different batches of AlSi10Mg, one of them manufactured using a plasma atomized process and the rest by gas atomization. The results indicated that plasma atomized powder samples have higher sphericity, surface smoothness, ultimate tensile stress (UTS), and elongation at break. Additionally, Powders with higher sphericity and bigger D10 (indicating 10% of powder particles have a diameter size lower than the D10 value) measure, demonstrate better flowability and, therefore better layer density and part density [21]. The topography and compaction density of powder bed in the LPBF process is highly affected by powder morphological and rheological properties, according to exploration carried out by Mussatto *et al.* which used AISI 316 L stainless steel powders for the experiments. This study implies that powder particles with less satellites have better flow during spreading process. It is also concluded that homogenous powder layer is achieved by powder samples which have wider PSD range and contain more fine particles, while it should be considered that powders with mentioned features usually demonstrate more powder segregation and poor flow [22].

To specifically interpret the correlation between Ti64 powder characteristics, the LPBF process, and final part's properties, Brika *et al.* implemented powder characterization of three samples of Ti64 powder, provided by two separate producers that used gas atomization and plasma atomization approaches. The authors of this study have stated that powder samples with higher sphericity and less fraction of fine particles demonstrate better density and flowability in the melt pool, and consequently, superior density, surface roughness, and mechanical properties in the printed parts [23]. Another study focusing on the influence of Ti64 powder indicated that machined LPBF-made parts using coarse powder samples (15-53 μm) exhibit slightly better fatigue performance. According to this study, the produced parts using a finer powder have a smaller number of defects, although the biggest defects were seen in fine powder-made parts due to more spatters originating from fine particles [24]. Visualization of pores and volumetric quantity of defects obtained in this study can be seen in Figure 2-4.

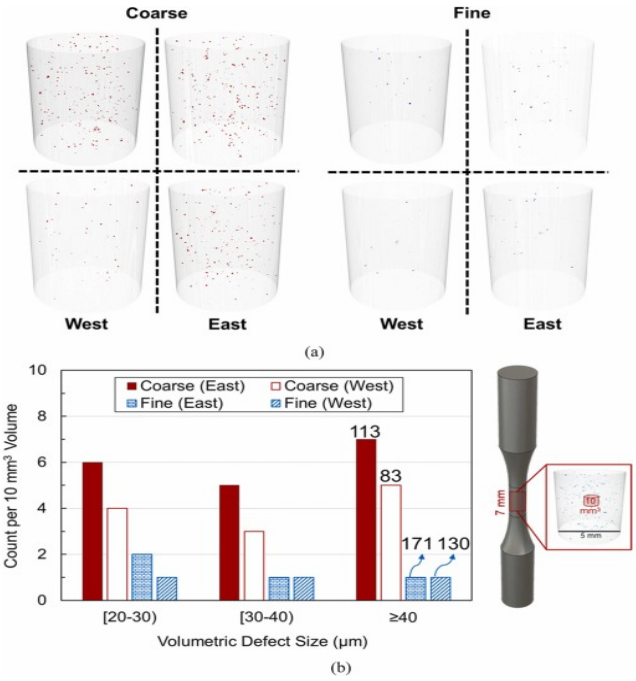


Figure 2-4: (a) 3D Visualization of porosities in coarse and fine powder sample produced parts built in two locations on build plate, (b) histogram of volumetric defect concentration in coarse and fine produced samples [24].

Further review of other literatures exploring impact of powder properties on the LPBF-made part quality is represented in Table 2-2.

Table 2-2: Literature review of studies focused on powder characteristics and part quality correlations

Sources	**Powder Properties**	**Part Properties**	**AM Process**	**Powder Material**
[25]	Morphology, density, flowability	–	–	316L Stainless Steel – AlSi10Mg
[26]	Powder fine particle fraction, flowability	–	–	–

[27]	Powder size distribution (PSD), sphericity, surface chemical composition, surface texture, flowability	_	_	Inconel718
[28]	Powder PSD, dense packing, flowability, spatter, build height	_	Selective laser melting (SLM)/LPBF	Ti-6Al-4V
[29]	Powder flowability indexes (shear, dynamic flow, compressibility)	_	_	Titanium alloys-Stain less Steel-Aluminum alloys etc.
[30]	Powder size, shape, surface roughness	Density-surface roughness-Porosity-laser/material interaction-	Selective laser melting/sintering (SLM) (SLS)	Multiple alloys
[31]	Powder PSD, median size, morphology, flowability	Density-melt pool dimensions	Powder bed fusion process	316L stainless steel
[32]	Powder PSD	Density/porosity (energy density)	LPBF	316L stainless steel
[33]	Powder production system, morphology, powder PSD, flowability	Yield strength-elongation-ultimate tensile strength	LPBF	304L stainless steel
[34]	Powder morphology (agglomeration, satellite), powder PSD	Porosity-surface roughness-microstructure	LPBF	AlSi10Mg

[35]	Powder PSD	Hardness-density-tensile strength	LPBF	17-4 PH stainless steel
[36]	Powder in various PSDs	Density, surface roughness, mechanical properties	LPBF	Ti64
[37]	Powder PSD, average size, flowability, apparent density	Surface roughness and dimensional accuracy	LPBF	Ti64
[38]	Powder PSD	Mechanical properties of the lightweight features	LPBF	Ti64
[39]	Off-size powder PSD	Mechanical, surface roughness, and dimensional accuracy	LPBF	Ti64
[40]	Coarse powder PSD	Density and porosity	LPBF-LBM	Ti64
[41]	Bimodal powder (coarse and fine)	Flow and compaction	LPBF	AISI 420 stainless steel

2.2 LPBF Process Parameters and Product Quality: Literature Review

The quality of final part is usually evaluated by assessment of factors such as part's physical properties, dimensional accuracy, and surface roughness [42]. The mechanical properties such as tensile strength, fatigue, hardness, and ductility [43] differ by any variation in the build process parameters. Many studies aim to optimize the quality of parts by analyzing the influence of process parameter variations on the mechanical properties. These studies are usually focused on specific types of metal powders

since the results of investigations are unique to the type of material examined. Accordingly, the literature review presented here regarding the effect of laser power, scan speed, hatch distance, and layer thickness on quality of LPBF products is mainly focused on investigations using Ti64 metal powder as feedstock material.

2.2.1 Correlation Between LPBF Process Parameters and Physical Properties of Final Product

Layer fusion state and porosity defects have significant impact on the final part mechanical properties such as strength of the part according to a study carried out by Malekipour *et al.* [44]. It is stated in the study that the fine layer thickness and higher scan overlap increases the layer bonding. Also, the layer orientation is highly impactful on the generation of porosity. The variation in scan velocity also impacts the mechanical properties of Ti64 parts in the LPBF process. For instance, in an exploration, the effect of scan velocity variations on hardness, tensile, and fatigue performance of printed parts is assessed in terms of energy density parameter (power, hatch distance, and layer thickness are fixed). It is concluded that the energy density lower than the standard caused by high scan velocity negatively impacts the aforementioned physical properties of the final parts due to lack of sufficient melting, lack of fusion and consequently generation of voids, pores, and fractures as the result [45]. It should be considered that process parameters impact on mechanical property is highly dependent on interactions between process parameters. Another investigation conducted by Khorasani *et al.* It is pointed out by the authors that tensile strength of parts (fabricated from Ti64 powder) is mainly dominated by effect of laser power process parameter [46]. A high laser power and low scan velocity cause weak tensile strength, while a low laser power and high scan speed leads to the lack of fusion and low tensile strength. It is indicated in this study that higher scan speeds with the optimum laser power increase tensile strength. Hardness is another mechanical property which reflects the quality of final parts in the LPBF process. The effect of laser power and scan velocity on hardness is demonstrated in a study by Khorasani *et al.* in which a higher laser power and lower scan speed (higher energy density) provide parts with better melting of tracks on top of one another, which in turn produces better hardness [47]. Also, it is mentioned that lower scan pattern angle will benefit the hardness of parts since it provides the chance of re-melting of previously fabricated layers. Porosity defects is another major factor that contributes to the densification degradation of final parts. This topic is also studied by Khorasani et al. in another investigation in which heat treatment factor was implied as the most dominant factor in improving parts' mechanical properties; however, scan speed and hatch space were also shown to have significant effects on density [48]. In summary, it is reported by the authors that the lower scan velocity and the higher hatch distance

(in standard hatch space interval) leads to sufficient wettability and less overlap, respectively. these conditions resulted in a higher density/less porosity in final parts.

2.2.2 Correlation Between LPBF Process Parameters and Surface Roughness of Final Product

An investigation conducted by Pegues *et al.* examined the effects of process parameters on the microstructure and formation of defects in parts manufactured with LPBF and electron beam melting (EBM) [49]. The authors claimed that in the LPBF process, the laser power and scan velocity affect the surface roughness of the top and side surfaces of the part. Furthermore, the increase of laser power reduced surface roughness in top surface and side surface due to the generation of flatten surface and less balling, respectively. Increase of scan velocity in the same fixed hatch distance setting was also reported to cause poor surface quality since this set of process parameters generates lower overlap between tracks of prints. The same connections between the laser power and scan velocity are reported by Khorasani *et al.* for the parts manufactured by LPBF using Ti64 metal powder [50]. It is also stated by the authors of this study that surface roughness is influenced both positively and negatively by the hatch distance as shown in Figure 2-5. This was explained in the study to be due to the surface tension elevated in overlapped areas which leads to worsen surface roughness, but on the other hand, the overlap of scans results in fewer empty areas in the melt pool (less perturbation in the melt pool), making the surface quality higher.

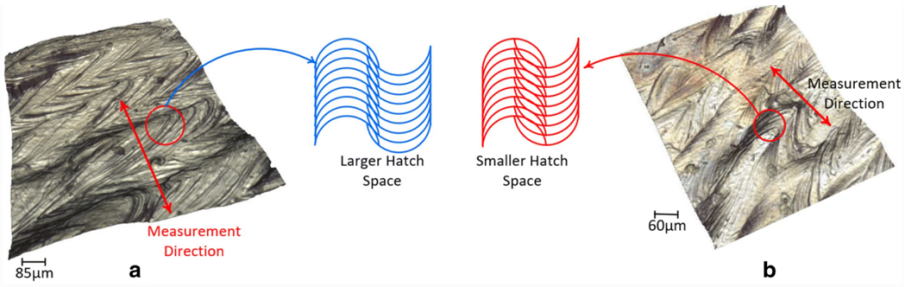


Figure 2-5: Effect of hatch space on ripple [50].

Research has shown that variations in the laser power greatly influence surface roughness of parts manufactured by LPBF using Ti64 powder. An optimal laser power range is required for an optimal surface quality, and any out-of-range laser power may lead to either the lack of fusion or keyhole defects [46].

2.3 In-Situ Melt Pool Monitoring of LPBF

Layer-by-layer manufacturing in LPBF, coupled with advanced additive manufacturing machines with built-in sensors, allows for in-situ inspection of melt pool and build processes, as seen in Figure 2-6 [3]. Using fast and precise analytical tools, real-time information obtained from various machine sensors, process parameters, and the final product's quality can be analyzed and controlled during the printing process.

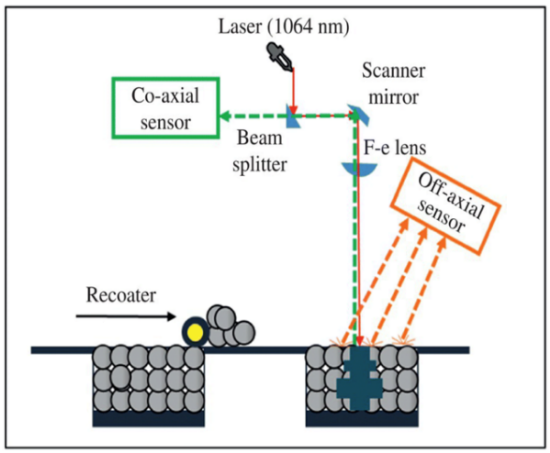


Figure 2-6: Melt pool monitoring in laser powder bed fusion process [51].

But it should be considered that the build process of LPBF can be controlled on-line as long as the sensor system is functioning properly. Therefore, the right sensor types and set-ups determination is crucial for accurate investigation of the LPBF process evolutions.

Recent explorations of in-situ AM process control have incorporated multiple sensor types and sensor configurations into the process, each of which provides specific advantages for some inspection applications. In the following sections, various sensor types and sensor set ups which are commonly applied for in-situ process control in LPBF are reviewed.

Various sensors in melt pool monitoring

Several types of sensor tools are available, each of which relies on specific signal sources. Acoustic, optical, and thermal sensors have been in the center of attention in recent years for in-situ melt pool monitoring and defect detection [3].

Examining/estimating melt pool temperatures is an essential factor to predict abnormally in AM products. The irradiated light off melt pool and part’s body normally in near-infrared (NIR) and infrared

(IR) wavelength ranges would be recorded by the non-contact sensors. It is imperative to adopt non-contact-based sensors, such as infrared cameras, photodiodes, and pyrometers, to obtain accurate light emissions and melt pool temperatures to capture the process signatures in real-time and in a non-intrusive manner [52]. Photodiode is a non-contact thermal sensing device which besides favors in terms of cost and sensitivity, it has the considerable asset of representing the light or thermal intensity of the melt pool in the form of single data point. It is attributed to the photovoltage measure related to the light captured by the sensor [53]. Various forms of analysis commonly applied on photodiode sensor signals are depicted in Figure 2-7.

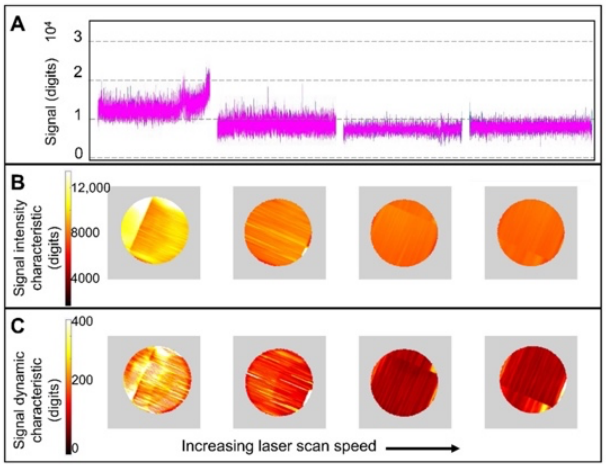


Figure 2-7: Melt pool monitoring using photodiode sensor: (a) raw intensity signals of prints, (b) moving average smoothed intensity signal transformed to spatial map, and (c) window moving averaged intensity showing dynamic of melt pool [54].

The widely utilized in-situ sensing set ups for LPBF are discussed in the following section.

Various sensor set ups in melt pool monitoring

Two sensing mounting approaches of co-axial and off-axial have been widely employed for in-situ process and melt pool monitoring in metal additive manufacturing. The sensors are mounted coaxially when the optical path of the radiated light coincides with power source axis. Off-axial sensor systems, on the other hand, are mounted outside the power source's optical path, allowing them to observe the evolution of melt pool from a different view. In each of these sensing setup arrangements, specific information can be extracted about melt pool dynamics, print quality, and hence final product quality. Accordingly, a properly configured sensor arrangement is necessary for achieving the measurement

objectives that can be direct or indirect defect detections. Sensor mounting set ups are illustrated in Figure 2-8.

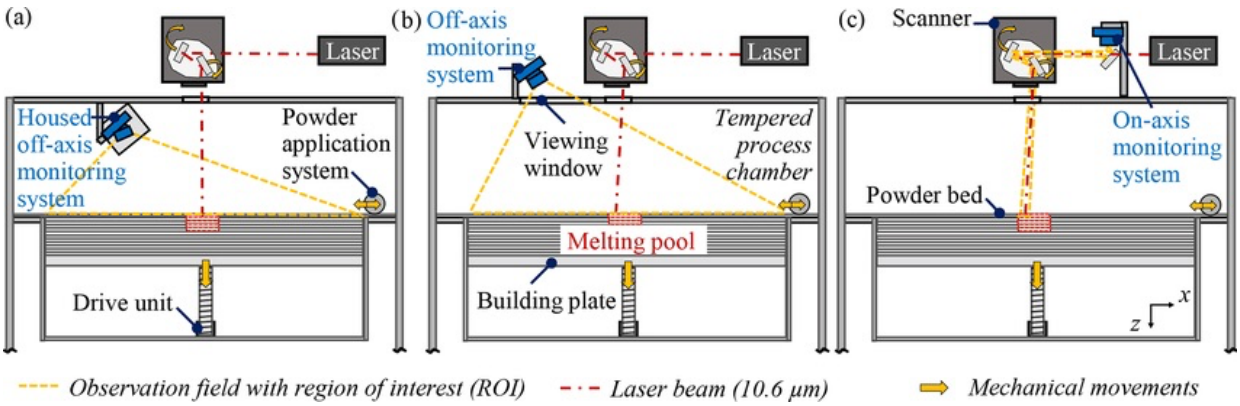


Figure 2-8: (a) Off-axis sensor mounting arrangement (sensor in chamber), (b) off-axis sensor mounting arrangement (sensor out of chamber), and (c) on-axis sensor mounting arrangement [55].

Various melt pool monitoring commercial equipment:

Several commercial printing systems have incorporated sensor technology into their systems, each with specific sensor types, mounting setups, and functionalities. One of the advanced monitoring systems for LPBF has been developed by EOS in two different systems with specialized analytical software for interpretation of acquired data from melt pool. EOSTATE Melt Pool Monitoring and EOSTATE Exposure Optical Tomography are two melt pool monitoring systems established by EOS.

Other advanced melt pool monitoring commercial systems such as Renishaw InfiniAM [56], GE additive concept laser [57], SLM Solution [58], and Trumpf monitoring [53] have also been widely employed in various academic studies for in-situ process inspection and control. Details of commercial systems including sensor types and sensing setup arrangements of sensors are reviewed in Table 2-3.

Table 2-3: Commercial melt pool monitoring systems

Manufacturer	System Name	Monitored Item	Sensors	Sensor Set Up
EOS	EOSTATE Melt Pool Monitoring	Melt pool	Photodiode 1	Co-axis
			Photodiode 2	Off-axis

	EOSTATE Exposure Optical Tomography	Melt pool	sCMOS camera	Off-axis
SLM solution	Melt pool monitoring	Melt pool	Two photodiode pyrometers	Co-axis
	Laser power monitoring	Laser	Photodiode	Co-axis
	Layer control system	Powder layer, printed layer	Camera	Off-axis
GE additive (concept laser)	QM melt pool 3D	Melt pool	Photodiode	Co-axis
	QM coating		Camera	Co-axis
Renishaw	InfiniAM	Melt pool	Photodiode 1	Co-axis
			Photodiode 2	Co-axis
Trumpf	Truprint Monitoring	Powder bed and part geometry	Camera	Off-axis
	Truprint Monitoring	Melt pool	photodiode	Co-axis

2.4 Application of Machine Learning in LPBF Quality Assurance Systems

As LPBF products are affected by a large number of interdependent factors, the development of an effective holistic QA system is extremely challenging. According to a recent study carried out by Seo *et al.,* the lack of integrated QA systems with authentic predictive models to examine the correlation between material, process, and part characteristics is an evident gap in AM quality assurance studies [59]. Smart systems/models capable of predicting powder characteristics in different materials and ranges of particle size distributions coupled with an in-situ build process inspection and modification will provide the opportunity for significant quality improvement that will contribute significantly to fill this gap. An in-situ LPBF control system should be able to inspect the quality state of the build based

on process and melt pool signatures (melt pool shape, depth, width, thermal intensity, etc.…). This process allows for real-time control and establishment of modification commands in cases of non-standard build conditions or aborting the process in cases of severe failure during the print.

In order to develop smart automated QA systems, ML models are recognized as the most practical tool due to their ability of accurate data interpretation and rapid decision making regardless of the type and volume of the data. These models can be used to evaluate and modify the printing process in real time.

2.4.1 Machine Learning Background

Machine learning is defined as computer programs that automatically learn how to perform tasks based on the collected data to improve the task performance metrics. In general, ML algorithms are usually divided into three classes of supervised, unsupervised, and reinforcement learning (RL). Supervised learning comprises of ML models that learn based on set of samples/examples with predefined labels or true output states. Algorithms such as support vector machine (SVM), decision tree (DT), random forest (RF), and some neural networks are supervised learning techniques. In contrast, unsupervised learning operates on a set of examples without specified labels and outputs. The algorithm learns to retrieve the underlying patterns and similarities between the unlabeled data. Clustering algorithms such as self-organizing map (SOM) are the instances of unsupervised ML algorithms. RL operates based on actions and the response to each action in each state. The model adopts an acting strategy or "policy" which determines how to act in each state such that the action rewards are maximized.

Numerous machine learning algorithms have been developed which differ in areas such as the procedure of optimizing task performing program identification which would be based on well-known optimization models or continuous testing and performance evaluation of randomly generated programs. The latter case is explored in genetic programming which includes various algorithms such as symbolic regression (GP-SR).

2.4.2 Application of Machine Learning in Powder Characterization

Thus far, the development of AI-based models for powder characterization has been limited, due to a lack of sufficient and reliable data of metal powder morphology and rheology. Metal powder property datasets is normally selected only through experimental measurements which are majorly inefficient in terms of time and expenses. Therefore, there are challenges associated with the integration of AI to powder characterization systems due to a lack of descriptive powder material data.

The investigations which are concentrated on the AI-based powder characterization have employed creative tools and approaches to resolve the data limitation obstacle. In an investigation Valente *et al.* have used each particle of 21 samples of powders as an individual observation in the dataset to tackle the material data limitation issue [60]. Purpose of this study was to classify each powder sample of data in two classes of good and bad flow, based on Hall flow metric measured for each powder sample. The ML model of decision tree yielded classification accuracy of 98.04% on known samples and 77.31% accuracy in flow classification of a combination set of known and new powder particles. Powder characteristics datasets compiled from multiple materials were used in a few studies to classify powder samples into various powder flow categories using AI and ML. Tkachenko *et al.* operated an investigation around the classification of material quality (manufactured part potential quality) based on morphological powder characteristics in three various titanium alloys and IMI 125 [61]. The machine learning algorithm ensembled of random forest tree and Kolmogorov-Gabor polynomial method is applied for classifying the samples to four classes of part quality. The 96.88% accuracy in test samples classification implies superior accuracy of the ensembled algorithm to logistic regression, support vector machine and random forest models.

Additionally, images of powder samples have also been used in several studies to eliminate the data restriction obstacle by alternating the numerical powder characteristics with image data. Multiple materials properties have been incorporated in some studies, in addition to applying SEM images as the dataset on which ML, computer vision, or a combination of both is used for powder characterization. For instance, Decost *et al.* employed a combination of computer vision and machine learning approaches to generate numerical presentation of powder micrographs which are then used in ML models for classification of the powder descriptive vectors once based on material type, and once based on powder production system [62]. As authors of the survey has stated, SVM resulted in 95% powder material type classification accuracy on test set. Also, Principal Component Analysis and t-Distributed Stochastic Neighbor Embedding (t-SNE) methods could accurately separate the feature vectors based on their production systems.

Many of the studies adapting ML models to powder characteristics datasets (of various types) are used for classification purposes, whereas very few studies have offered regression ML models for powder characteristics prediction in the AM field. In recent research completed by Zhang *et al.*, Convolutional Neural Network (CNN) deep learning model with VGG-16 architecture and computer vision approaches along with a single layer Artificial Neural Networks (ANN) are adopted to predict powder

properties such as Hausner ratio, Carr index and angle of repose as the flowability indicators of three different powder materials (AL-Si-Mg, Ti64, and Inconel625) [63]. The results implied that computer vision technique founded on speed up robust features (SURF) feature extraction approach yielded superior performance in regression of target powder properties with 12∓7 MAPE. Particle shape analysis using deep CNN and digital image processing approaches is also explored by Liang *et al.* to extract the shape of the particles from images with complex backgrounds using the trained CNN algorithms [64]. This is followed by modification of the extracted shapes' edges by image processing techniques to eventually determine the shape distribution of the particles. The shape distribution estimations from examined method is assessed using manually obtained shape distribution of the samples which confirmed the successful performance of the proposed method. In a more recent investigation, powder particles are classified into six shape groups (spherical, pearl, connected, etc.) using deep neural network and cloud point data retrieved from 3D x-ray computed images of Ti64 powder [65]. Bakas et al. in another study applied mask-R CNN algorithm as well, to estimate morphological properties such as high sensitivity circularity, circular equivalent diameter, spherical equivalent volume, etc. of AISI H13 tool steel powder [66].

The studies with similar mentioned objectives in the AM field are mainly restricted to classification of powders in terms of microstructure, size, shape, and flowability majorly using high dimensional data such as image which causes high computational complexity. Studies using regression models and numerical powder property data have been mostly conducted in pharmaceutical field. For instance, Sandler *et al.* used a statistical regression model called Partial Least Squares (PLS) which was trained on a numerical and continuous dataset comprised of powder size and shape distribution properties as input, and powder packing and flowability measures as the target properties [67]. According to the authors of the study the model R^2 was above 85% in prediction of packing properties when evaluated on unseen test samples, while this value decreased for powder flow properties (68% and 84% for flow function and UYS, respectively). Schuliger et al. applied a data of powder morphological properties (size, shape, etc.) in trained multiple-instance regression (MIR) and decision tree classifier algorithms to estimate the hall flow rate of multiple materials of metal powder [68]. In another similar study the authors developed multiple algorithms of mask R-CNN, linear regressor, and RF regressor to estimate the hall flow rate and apparent density of S7 and A8 tool steels powders based on the morphological properties and then the proportion of particle clusters in powder [69].

2.4.3 Application of MPM and ML for In-Situ Process Optimization

Although sensing technology in printing machines has achieved good precision in data acquisition from build processes and melt pools, accurate on-line analysis of data gathered from sensors, and extraction of the correlation between the data to the final quality of parts has remained a major challenge [3]. However, it is now possible to quickly and cost-effectively analyze a variety of sensor data types using AI models (machine learning and deep learning) to facilitate online decision making. For instance, in-situ defect detection can be achieved by using sensors that detect process signatures during print. The real-time process correction can then be performed based on analyses produced from analytical and AI models to prevent defects in subsequent layers [4]. A review of the most recent innovative in-situ melt pool monitoring investigations for LPBF, in which machine learning and deep learning methods are utilized is presented in the following sections.

In a study, two on-axis photodiode sensors of SLM 280^{HL} system are used to assess the thermal emission sensitivity to the energy density variation and drift defects denoted as lack of fusion and overheating. A support vector machine classifier is also trained for detection of drift layers in different geometrical shapes of AlSi7Mg0.6 [70]. Similarly, in another investigation relation between optical emissions (multispectral) recorded by two photodiode sensors and final quality of parts in terms of volume of voids is reported in [71]. The optical emissions of prints of Inconel 718 parts are used for melt pool monitoring in another study which are gathered by two off-axis photodiode sensors (multispectral monitoring) on which the graph Fourier transform coefficients are extracted to be used instead of conventional statistical signatures (mean, median, etc. …) of melt pool emission signals. Multiple classification models (for porosity level classification) such as support vector machine (SVM), DT, k-nearest neighbors (KNN), shallow neural network etc. … along with a regression shallow neural network (for porosity volume estimation) are employed, which all resulted in achieving higher accuracy in porosity severity detection/estimation using the graph Fourier transform coefficients as input rather than usual statistical MPM signals signatures [72]. Harbig *et al.* explored the potential of a system consisted of various sensor types and sensing systems in defect detection of the LBF-LB/M process [73]. They used two melt pool monitoring systems of EOS M290 and a test bench with 316L and Scalmalloy® materials, respectively. On-axis and off-axis photodiodes and on-axis high-speed camera sensors are used in EOSM290, while an off-axis thermographic camera and an off-axis high-speed camera are utilized in the test bench printing system. The results indicated that combining the data of monitoring systems increased the balling defect detection sensitivity by about 20% due to the suitability

of each single monitoring system for detection of specific sets of defects, while the combination of them leads to a more extensively accurate defect detection in melt pool. The melt pool light intensity signal attained from photodiode sensors coupled with machine learning models employed in a study for the prediction of the microhardness level of parts printed from K438 alloy powder feed stock. The frequency domain of intensity signals is used to extract 9 MPM signal signatures (power spectrum density per each layer of build), to be used as input of ML models such as XGBoost, RF, and LightGBM, in addition to process parameters measures (power, scan speed, hatch distance). XGBoost yielded the best overall performance in terms of train and prediction time and hardness prediction accuracy [74]. Unsupervised machine learning algorithms have also been explored for in-situ photodiode MPM data interpretation and quality evaluation. For instance, Jayasinghe *et al.* employed the Renishaw 500 M equipped with three on-axis photodiode sensors to retrieve melt pool signatures from MPM data using the singular value decomposition method. The signatures are then used as input to two unsupervised K-mean clustering and gaussian mixture model (GMM) to cluster the samples based on the final part density measure. Also, a supervised gaussian process model using the same signatures is practiced in predicting the density, which results indicated superior accuracy of photodiode-based signatures in comparison to reliance on only process parameters data as input of density prediction models (in both supervised and unsupervised algorithms) [75]. The inversion of photodiode 1D signal (MPM data) to grayscale images has been implemented by Cao *et al.* to take advantage of the full potential of the CNN for LPBF track molten state classification (lack of melting, normal melting, and over melting). Off-axis photodiode is applied to track melt pool light intensity data acquisition, and images are generated from signals. Classification accuracy of 94.2% was reached by the proposed method [76]. In a recent study, the development of an unsupervised ML algorithm for in-situ defect detection and the establishment of an online control system for defect reduction has been explored. Taherkhani *et al.* used a SOM clustering ML model on on-axis photodiode MPM signals (using EOS M290 and EOSTATE Melt pool monitoring system) for lack of fusion defect detection of parts fabricated from Hastelloy powder [77].

In addition to the review of the most innovative and state-of-the-art research in photodiode-based in-situ monitoring coupled with machine learning for LPBF quality assurance, other similar works are presented briefly in the Table 2-4.

Table 2-4: Table of literature review of machine learning and melt pool monitoring application for process inspection and optimization

Sources	ML Model	Material	Sensor	ML Input	ML Output
[78]	MLP, SVM, KNN	In625	High-speed camera, thermal camera, photodetector	Melt pool image, thermal image, intensity signals	Overhang and bulk thermal signatures identification
[79]	GMM	IN718	Photodiode	Melt pool intensity	quality based on ultimate tensile strength condition
[80]	CNN, long short-term memory (LSTM)	SS316L	Photodiode sensors, acoustic sensor	Optical, thermal, acoustic signals	Lack of fusion, conduction mode, keyhole
[81]	Logistic regression, support vector classifier (SVC), RF, and DT	AlSi10Mg	Photodiode	photodiode statistical signatures	Nominal or overheating anomaly in layer
[82]	CNN	–	High-speed cameras, photodiodes, and microphones	Monitoring signals turned to 2D grayscale images	Quality class based on density
[83]	Back propagation	–	Photodiode	Photodiode signal/temperature	Temperature/flaw detection

	neural network (BPNN), Stacked parse auto-encoder (SSAE), LSTM				
[84]	CNN	_	Photodiode, camera	Fusion of signals	Classifying Various sizes of defect
[85]	Auto-encoder (AE), deep CNN model (ResNet)	_	Acoustic, photodiode	Signals	Melt pool characteristics visualization/LoF detection
[86]	XGBoost	_	Photodiode	Signal in time and frequency domain	Classifying abnormal process parameters

As reviewed previously, major proportion of the studies conducted for quality assurance of the LPBF process adopting melt pool monitoring (photodiode sensor as the focus of this thesis) and ML are focused on abnormality/defect detection. Meanwhile, tailoring process parameters for LPBF is still substantially implemented through experimental and conventional statistical methods of DoE. Very few investigations have explored the application and advantages of a technique enhanced by the combination of the ML and in-situ monitoring as a fast-paced and data-driven alternative or complementary aiding tool for design of process parameters in LPBF. According to the state-of-the-art of the studies carried out with the purpose of data-driven process parameters optimization, it could be observed that in recent years the major focus has been invested to application of ML to optimize process parameters solely based on physical and mechanical properties of the parts while melt pool monitoring, as one of the main quality-impacting factors, is not concerned in these inspections. A brief review of similar studies is presented in Table 2-5.

Table 2-5: Table of literature review of machine learning application for process inspection and optimization

Source	ML Model	Material	ML Input	ML Output
[87]	Gradient boosting regressor (GBR), XGBoost, adaptive Boosting (AdaBoost), RF, gaussian process regressor (GPR)	Multiple	Process parameters	Microhardness and wear rate
[88]	Linear regressions, DT regressions, RF, multilayer perceptron (MLP)	Ti64	Process parameters	Melt pool geometry, density
[89]	SVR, MLP, regression tree (RT), RF, AdaBoost, GBR, XGBoost	SS 316L	Process parameters, thermal properties, beam diameter	Density and surface roughness
[90]	Reinforcement learning (RL)	SS316L	Rewarding based on Melt pool depth	Power-velocity pair
[91]	MLP	AA2024	Melt pool dimensions/process parameters/combination of melt pool dimensions and one of the process parameters	Process parameters/melt pool dimensions/the other process parameter

[92]	Naïve Bayesian (NB), SVR, logistic regression	Ti64	Process parameters/part properties	Part properties/process parameters
[93]	KNN, ensemble learning decision tree, dendrogram clustering algorithm, SVM	316L stainless	Process parameters	density, hardness, yield strength, and tensile strength
[94]	XGBoost	stainless steel 316 L, AlSi10Mg, and Fe60Co15Ni15Cr10 medium entropy alloy	material properties of powder and process conditions	density
[95]	Extra trees Regression, GPR	Ti64	Process parameters, layer thickness, and powder porosity	Melt pool geometry

Few studies have been published in recent years which adopted melt pool monitoring and ML for process parameters optimization. For instance, some studies have developed close-loop quality controlling systems based on process parameters optimization according to feed-back provided by ML algorithms [96], [97], [98]. However, to the best knowledge of the author of this thesis, one recent study carried out by Lapointe et al. has attempted to apply ML and melt pool monitoring (photodiode-based) to provide process parameters for better dimensional accuracy in parts as a non-feed-back-control system and solely as a technique for priori process parameter optimization. the authors worked on development of two 1D convolutional neural networks. The first model for prediction of scan track photodiode signals using laser power, scan velocity and five geometry-related features (extracted from CAD files of prints), and the second model for prediction of laser power and scan velocity as the output of a combination input set of photodiode signals and geometrical features. The first model resulted in the relative success of signal estimation (not in signal fluctuations), while the second model

demonstrated notable performance in the prediction of process parameters [99]. The three major physical and mechanical properties to be optimized in each LPBF application are density, hardness, and surface roughness, which have never been the focusing point for process parameters tailoring by ML, relying on melt pool monitoring data. This led to the motivation of the second objective of this research to train and assess different regressor algorithms for accurate prediction of process parameters which were used to fabricate parts with varying density, hardness, and surface roughness qualities. Moreover, another distinguishing attribute in design of this analysis was to involve standard and off-size powder size ranges in prints and collection of the data, and ultimately in the algorithms' input feature set.

2.5 Summary

Powder feedstock features, process parameters changes, and melt pool dynamics are known to have major impact on quality of final LPBF-made parts. Therefore, extensive efforts have recently been put forward to develop and assess automated quality assurance (QA) systems using machine learning (ML) and melt pool monitoring (MPM). In this chapter first, a comprehensive review of impact of powder characteristics (mainly focused on Ti-6Al-4V (Ti64)) and process parameters on final laser powder bed fusion (LPBF) components is provided. It is then followed by review of state-of-the-art literatures concentrated on powder characterization and optimization of LPBF process using machine learning algorithm. MPM methodology and its functionality in in-situ process control and quality optimization is also surveyed. As the literature in this chapter exhibits, there is a lack of knowledge on holistic relationships between feedstock powder characteristics, LPBF process parameters, and LPBF-made final product's physical properties (hardness, density, and surface roughness), since most studies are mainly concentrated on control systems and defect detection. But it should be considered that in the event of slight or significant changes in any of the factors discussed above, the quality of the finished product will be significantly altered. In LPBF, this phenomenon emphasizes the necessity of in-situ holistic QA systems that incorporate feedstock characteristics and process parameters concurrently. In-situ QA systems demand high-speed analyzing tools capable of big data processing and accurate decision making to achieve real-time quality control. Accordingly, ML-based powder characterization and estimation of optimized process parameters based on MPM intensity signals are pursued in this research for addressing the existing gaps in current state of the LPBF quality assurance studies.

Chapter 3: On the performance of expert-augmented machine learning with limited experimental data collected from powder particle characteristics used in laser powder bed fusion

3.1 Abstract

This chapter investigates the feasibility of using machine learning (ML) regressors and equation fitting algorithms, optimized for small datasets, to estimate dynamic rheology (specific energy (SE) and basic flow energy (BFE)), shear rheology (unconfined yield strength at 3kPa (UYS3), 6kPa (UYS6), and 9kPa (UYS9), and Hall flowability of Ti6Al4V (Ti64) powder. The goal is to develop predictive models for these critical rheological characteristics across various particle size distributions (PSD) relevant to laser powder bed fusion (LPBF) applications, enabling prior estimation of powder bed quality. A modified synthetic minority oversampling technique (SMOTE) is applied to address sample scarcity where ML or equation-based methods underperform. Morphological powder characteristics are used as inputs for regressors such as support vector regression (SVR), k-nearest neighbors (KNN), decision tree (DT), and random forest (RF). After algorithm tuning and feature selection, the best regressor for each target is selected based on leave-one-out cross-validation and further evaluated using an unseen sample set. Results indicate that while conventional ML regressors like SVR are sufficient for dynamic flowability prediction, more complex targets such as shear and Hall flowability require innovative approaches like genetic programming and SMOTE.

3.2 Methodologies and experimental setups

The goal of this study is to predict the flowability and rheological properties of the Ti64 powder based on the powder's physical and morphological properties (shown in Figure 3-1). Various ML algorithms that are known to be practical on limited datasets are optimized, trained, and evaluated for predicting powder flowability targets in various standard and off-size PSD ranges. The goal is to analyze the feasibility of ML application for powder flowability indicators estimation; which are highly influential in the LPBF process; relying on small datasets. Since powder characteristics data are generally extremely limited, costly, and difficult to obtain, it is of high importance to resolve the lack of data problem without major dependency on attaining a high number of experimental samples. Accordingly, the data limitation challenge is addressed in this study first by applying simple conventional ML regressors and equation-fitting (genetic programming (GP)) on the limited data sets. Lastly, the data

augmentation technique was employed when the previous remedies subjected to the original dataset resulted in poor generalization of regression. Hence, the priority of this study is to achieve adequate accuracy in the estimation of powder rheology relying on the first two techniques since they operate on the available limited data.

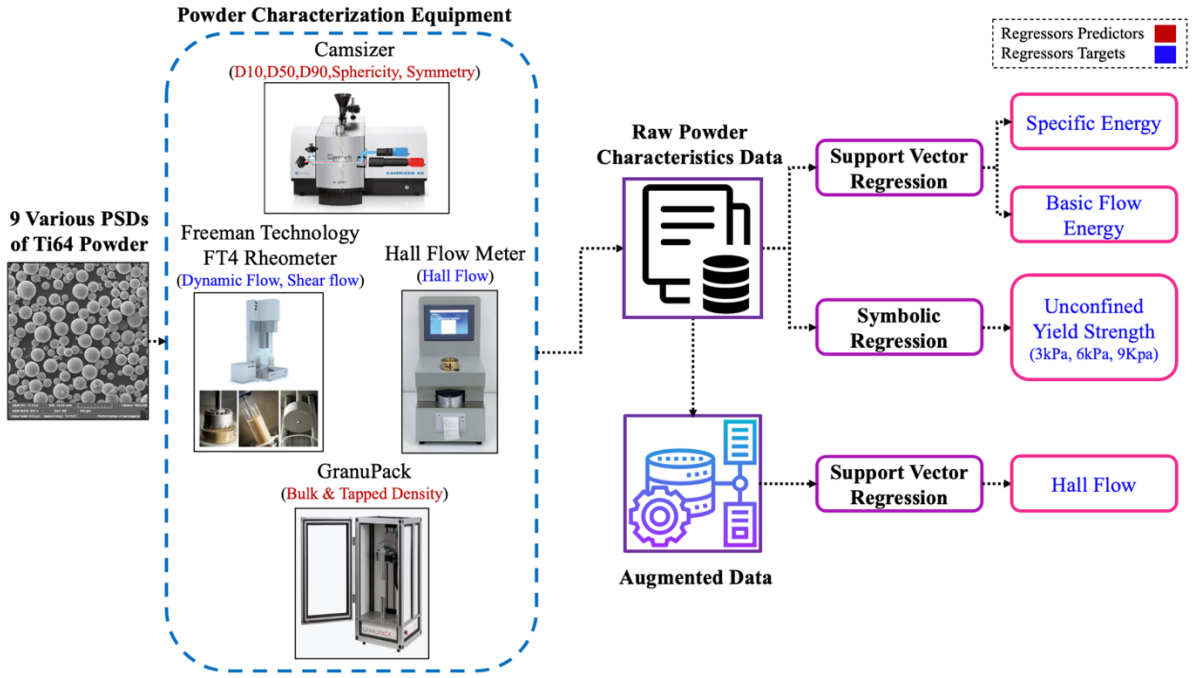


Figure 3-1: Graphical abstract of the analysis (equipment images adapted from [100], [101], [102], [103]).

Nine different PSD ranges of Ti64 (grade 23) were provided by AP&C (GE Additive, Montreal, QC, Canada) to be used as samples of the analysis. Standard PSD ranges of powder in LPBF such as 15-45 µm and 15-53 µm are included in the dataset along with off-size and wider PSD ranges such as 15-106 µm and 45-106 µm. The list of 9 samples is provided in Table 3-1.

Table 3-1: Particle size distribution ranges of Ti64 powder

Sample	Ti64 _1	Ti64 _2	Ti64 _3	Ti64 _4	Ti64 _5	Ti64 _6	Ti64 _7	Ti64 _8	Ti64 _9
PSD	15-45μm	15-53μm	15-63μm	15-75μm	15-90μm	15-106μm	45-75μm	45-90μm	45-106μm

Powder physical and morphological properties such as size distribution (indicated by three PSD percentile values of D10, D50, and D90), the span of the size range (the ratio of D90 minus D10 to D50), and powder particle shape (average symmetry and sphericity over the distribution) are used as input variables/predictors. The rationale for choosing these inputs is based on their significant influence on powder behavior, such as flowability. Furthermore, it should be noted that the focus of the investigation was on bulk-scale powder characteristics and particle-scale characteristics were not included in the scope of this study. Powder size and morphology are characterized and quantified using the Retsch Technology Microtrac Camsizer X2 (X-Dry with X-Jet dispersion) device which uses the Dynamic Image Analysis tool technique (the technique is described in the latest draft of ISO 13322–2:2021 [104]).

In addition, bulk packing characteristics of powder such as bulk/apparent and tapped density, and the compressibility index of powder; which is retrieved using both values; (denoted by Carr's index) are the second group of powder properties selected as predictors. Powder bulk density and compressibility determine the homogeneity and denseness of the powder layer in the LPBF process. Furthermore, compressibility impacts the rheological behavior of the powder (shear and dynamic flow) due to the involvement of compression/consolidation in rheological characterization techniques. Therefore, this set of powder properties is also included as the input variables. Bulk density and tapped density are obtained using the GranuTools GranuPack device. The set of these 9 input features (authors-selected) is shown in Table 3-2.

Table 3-2: Authors-selected set of powder properties used as the input features of the regressors

Input Feature Category	**Particle Size Distribution**				**Particles Morphology**		**Packing/Density**		
Input Feature Name	D10	D50	D90	Span	Sphericity	Symmetry	Bulk	Tapped	Carr's Index

The targets/outputs of the ML models in this study are three classes of hall flowability, dynamic flow, and shear flow that will be predicted by optimized ML models (trained for each target separately). Hall

flowability *(s)* is determined by the time (seconds) that it takes for the bulk of powder to flow through the funnel of the hall flowmeter and is the conventional measure of the general flowability of powders. The rheological properties of powder (dynamic and shear flow) on the other hand indicate the bulk flowability of powder under more sophisticated conditions, which are closer to the actual behavior of powder in LPBF printing. The dynamic flowability of powder is determined by two measures of basic flow energy (BFE (*mJ*)) and specific energy (SE *(mJ/g)*) which measure the energy required for forced/downwards and unconfined/upwards blade motion in the bulk of powder, respectively. Powder shear flowability in this study is determined by unconfined yield strength metric (UYS (*kPa*)) at three various consolidation levels of 3kPa, 6kPa, and 9kPa. UYS implies the level of resistance to shearing/flow against another layer of powder in the consolidated powder. Both rheological property groups of powder are measured by Freeman Technology FT4 equipment.

3.2.1 Machine Learning Algorithms and Data Preprocessing

Data normalization was carried out using the standardization technique to transform the data into the same scale in feature space. The data limitation challenge in this study requires inspecting the ML models that operate well on small datasets. Conventional ML models such as DT, SVR, RF, GB, and KNN are all investigated in this study to assess their performance and identify the most accurate and robust model for powder flowability prediction using limited datasets. The SVR foundation is similar to the SVM classification algorithm but modified for regression purposes. With SVR, the best-fitting hyperplane is found using optimum support vectors (SV) while satisfying two constraints: the minimization of prediction error and the minimization of complexity of fitting hyperplane. For detailed introduction to SVR algorithm refer to appendix B. KNN algorithm predicts the target for unseen data based on similarity of it to train samples. The similarity is determined by distance metrics and the average or weighted average of the K (needs to be optimized and defined by user) train samples close to the unseen data is calculated to predict the target. DT algorithm divides the dataset into various groups (parent node splits into two child nodes which contain groups of data points) based on selection of a partitioning condition which will be determined at each node of the tree on the feature space of data. The dataset splits according to whether each data point satisfies the parent node condition or not. The purity of dataset at each node will be then measured based on present variation in the set. The partitioning of data continues (along the tree length) until no further improvement is made on the purity of last partitions of dataset. The unseen datapoints prediction is estimated from the average of target value corresponded to train data partitioned in the same leaf (node) as the unseen sample.

Random Forest (RF) and Gradient Boosting (GB) are both advanced algorithms derived from the Decision Tree (DT) technique. These methods share a common foundation in partitioning data samples and repeating this process until no further improvement is achieved based on a chosen impurity metric. The key distinction between RF and DT lies in RF's use of an ensemble of multiple trees, each trained on random subsets of both features and data samples. This ensemble approach enhances the robustness of regression compared to a single decision tree.

While GB also employs an ensemble of trees, it differs from RF in that it builds the trees sequentially, with each new tree focused on correcting the errors or residuals of the previous one. In other words, GB improves regression in a step-by-step manner. This sequential learning process makes GB computationally more complex but allows it to capture more intricate patterns in the data.

The hyper-parameter optimization of each model was performed by the grid search and leave-one-out CV method on an author-defined set of values. This was then followed by sequential feature selection (based on the mean absolute error metric) before the ML model training and prediction. The performance results of these models and comparison analysis between models are reported in the results section.

GP-SR is employed to address the limitations of conventional ML models in the accurate modeling of complex rheological targets, such as shear flowability (UYS). In contrast to conventional ML models, genetic programming is an evolution-based algorithm that trains models without a prior assumption (non-parametric) about the fitting function and data's variables space. Therefore, it allows for training and generation of a more accurate equation in datasets with complex underlying relations. GP-SR produces mathematical equations and formulates the relations between input and output variables. For detailed introduction to GP-SR algorithm refer to Appendix B. The resulting equations in this study are used to estimate the UYS target for all 9 samples of data. ML algorithms training and analysis are implemented on an open-source Python programming platform.

3.2.2 Data Limitation Challenge and Remedies

Although various ML models applicable to small datasets were trained, optimized, and tested during the investigations, the lack of sufficient samples' effect on the accuracy and generalization still persevered in some of the targets' modeling. The data generation method of SMOTE was employed to generate synthetic samples from the original data. It involves creating new samples whose differences from the original samples are not statistically significant but have slight deviations from the original

samples. The new samples thereby can promote the reduction of regression overfitting (increased generalization) to some extent. This declines the overfitting of the model to limited samples of the train set. The SMOTE technique, which is usually applied to eliminate the imbalanced classes in the classification ML algorithms, has been expanded by researchers in a study for regression applications [105]. In this work, the KNN algorithm is first used to find two neighboring samples to each of the original samples. Then, the new sample is generated using a mathematical equation and the two neighboring samples. After the generation of the new dataset, it will be tested statistically to ensure it is not significantly different from the original data. This data augmentation approach was adapted in this study with minor changes in number of neighboring samples (which was changed to one here due to limitations of the raw dataset) and therefore the mathematical equations used for data generation. Alternative generative approaches, including variational autoencoders (VAE) and conditional generative adversarial networks (CGAN), were also investigated during this study. However, SMOTE demonstrated superior performance in augmenting the dataset, given the constraints imposed by the limited availability of raw data. The limited raw dataset made it difficult to properly train the generative neural networks, which ultimately made these methods unsuitable for the intended applications in this research.

3.3 Results and Analysis Discussion

Throughout this section, in the first part, the performance of each optimized ML regression model, which is trained and tested on the original Ti64 dataset, is reported and discussed for each target in the dynamic and shear flowability categories. An overall comparison between candidate conventional machine learning algorithms (SVR, DT, KNN, etc.) is presented to select the best regressor for each target. The equation-fitting approach of GP-SR is then applied and reported (used for shear flowability estimation) as a separate technique practical for limited data applications and fitting of complex relationships. A concluding discussion is then provided to evaluate the feasibility and performance of all practiced techniques which rely only on the highly limited original dataset but still could provide satisfactory results.

The second part is devoted to the demonstration of results achieved by the application of data augmentation methods used for hall flowability target regression in addition to the shear and dynamic targets which previous approaches failed to accurately model. The ML algorithms were once again optimized and trained on the newly generated datasets to assess the models' performances and selection

of the best regressor. A final discussion regarding the impact of data augmentation on the regression accuracy of the algorithms is then provided.

3.3.1 Flowability Prediction Using the Original Dataset

3.3.1.1 Machine Learning Optimization and Performance Evaluation Discussion

Using the original dataset with only 9 samples (Ti64 powders with 9 various PSD ranges, including standard LPBF ranges to off-size ranges) challenges the ML algorithm developments and performance. Furthermore, obtaining and using a high number of samples is not feasible for this problem set due to its nature and cost concerns. Considering the importance of powder PSD variation in the estimation of flowability and the fact that only a limited number of powder PSDs are practical for industrial/commercial LPBF production systems, this study highlights the significance of training and adjusting conventional machine learning regressors to perform effectively with a constrained quantity of experimental samples.

Additionally, the generation of each of these powder samples in a specific PSD range and experimental powder characterization is highly time-consuming and expensive, making the collection of big data on metal powder characteristics almost infeasible. Accordingly, to the authors' best knowledge, this investigation is the first attempt to train and test ML algorithms for metal powder flowability estimation using a restricted number of samples that vary based on PSD ranges established for LPBF production systems. Furthermore, multiple off-size PSD ranges of powder are also included in the dataset for expanding the application of ML-based powder characterization systems in powder production industries.

Various approaches have been used in this study to train and test the machine learning algorithms and draw logical conclusions regarding their performance on highly restricted material samples. One main technique that could be practical when using small datasets is CV which performs based on iterating a train-test process multiple times using varying portions of the present dataset.

Firstly, the leave-one-out CV method is performed on train samples to determine the optimum hyper-parameters of ML algorithms. It is then followed by backward sequential feature selection (per each target).

After the determination of optimum hyper-parameters and input feature set/predictors, the performance of the conventional ML models (SVR, DT, GB, RF, etc.) is explored through three modes of 1- single

sample prediction (based on leave-one-out CV) denoted by 'SSP' (flowability estimation of each of these established PSD ranges based on the information of other 8 PSD ranges), 2- repeated random split prediction (based on repeated K-fold CV) denoted by 'RRSP' (random selection of train and multiple test samples in multiple iterations to assess the generalization performance of the ML models), and 3- predictions made on two authors-selected test/unseen samples denoted by 'train-test split' (conventional approach of ML regression model evaluation). 7 samples of original data are kept as training samples, and two remaining samples of 15-63 μm and 45-90 μm are chosen as test samples since these two are representative of fine powder group samples and coarse off-size group samples, respectively. While the last mode is useful for evaluating the model's performance and generalization in applications with a large dataset with more representative test samples, here, however, it is only used to examine the regressors' performance on authors-selected unseen samples. In the other two modes (SSP and RRSP), the entire dataset (all nine samples) is used, in which performance evaluation is assessed on all the standard and off-size samples, based on CV techniques. In limited sample cases, these modes could serve as the best means to assess the model's performance and generalization.

The SSP method operates in a similar leave-one-out CV technique. The main reason this evaluation was utilized here is to have a precise understanding of ML regressors' performance in the prediction of each PSD range of powder solely. This is important because, in the LPBF process, one optimum PSD range (according to parts application) will be used for printing, and it is the priority for the customers to be able to estimate the powder behavior/rheology for single powder batch in hand. Therefore, the SSP approach satisfies the ML performance evaluation in terms of AM requirements and customers' interests.

Regarding the RRSP evaluation mode importance, it is essential to consider that the performance of regression should also be evaluated from a ML science perspective. This would be addressed mainly by generalization testing. The generalization of regression algorithms is usually assessed through the common method of 'train-test split' since enough data samples are normally available. But in this case with data restrictions, it is important to employ a valid generalization evaluation method by applying random various train sets, and unseen samples (in contrast to the 3rd evaluation mode which operates on the authors-selected train and unseen samples). Furthermore, the generalization should also be tested based on how accurately multiple random unseen samples could be predicted, which the SSP mode doesn't support. Accordingly, the RRSP mode would be the best technique to test the regressor's generalization to address the randomness of test samples (satisfied by the random split of data in each

iteration) and regression accuracy for multiple test samples (considered in the portion specification of the random split of data).

Lastly, the GP-SR method was applied in this study in the estimation of shear flow targets, as an alternative approach to conventional ML models. However, it should be considered that the GP-SR resulting equation (equation fitting method) was then adopted to estimate the targets for each sample. Accordingly, in this case, the performance evaluation modes mentioned beforehand were replaced by the common mean absolute error percentage calculation between the equation's outcome (estimation of target) and the actual value.

3.3.1.2 Flowability Prediction Results

Considering the small dataset on which the ML regression model should be developed, simpler ML algorithms such as SVR, KNN, and DT, and other ensemble algorithms, that are known to be effective with small data, such as RF and GB were explored in this study. As mentioned in the previous section, each model is firstly optimized per each target. This is followed by selecting the best regression algorithm for each target by testing all the algorithms on the train set (7 samples preselected by the authors) using the leave-one-out CV method. It should be noted that the full set of input features shown in Table 3-2, predefined by the authors, is used during the leave-one-out CV, and the feature selection process was carried out later using the selected model in this step. We used the leave-one-out CV in this study due to the restriction of data. Moreover, it makes the accuracy of the predictions in each single test sample a higher priority. Therefore, the regressor demonstrating the best performance in leave-one-out CV is initially selected and subsequently utilized in further analyses and performance evaluations, including multi-test sample regression. The result of the leave-one-out CV of all tested algorithms for each target is shown in Figure 3-2. The green arrows represent the average negative mean absolute errors (MAE) obtained in each CV turn. The box represents the interquartile range of obtained MAE values (between the 1st quartile and the 3rd quartile) and the caps represent the minimum and maximum MAE values. Empty dots show the outliers in the corresponding MAE population's box plot. The hyper-parameters of the five tested regression models for each target are provided in Appendix A. The hyper-parameters of the selected algorithms for each target are reported in the respective target results sections.

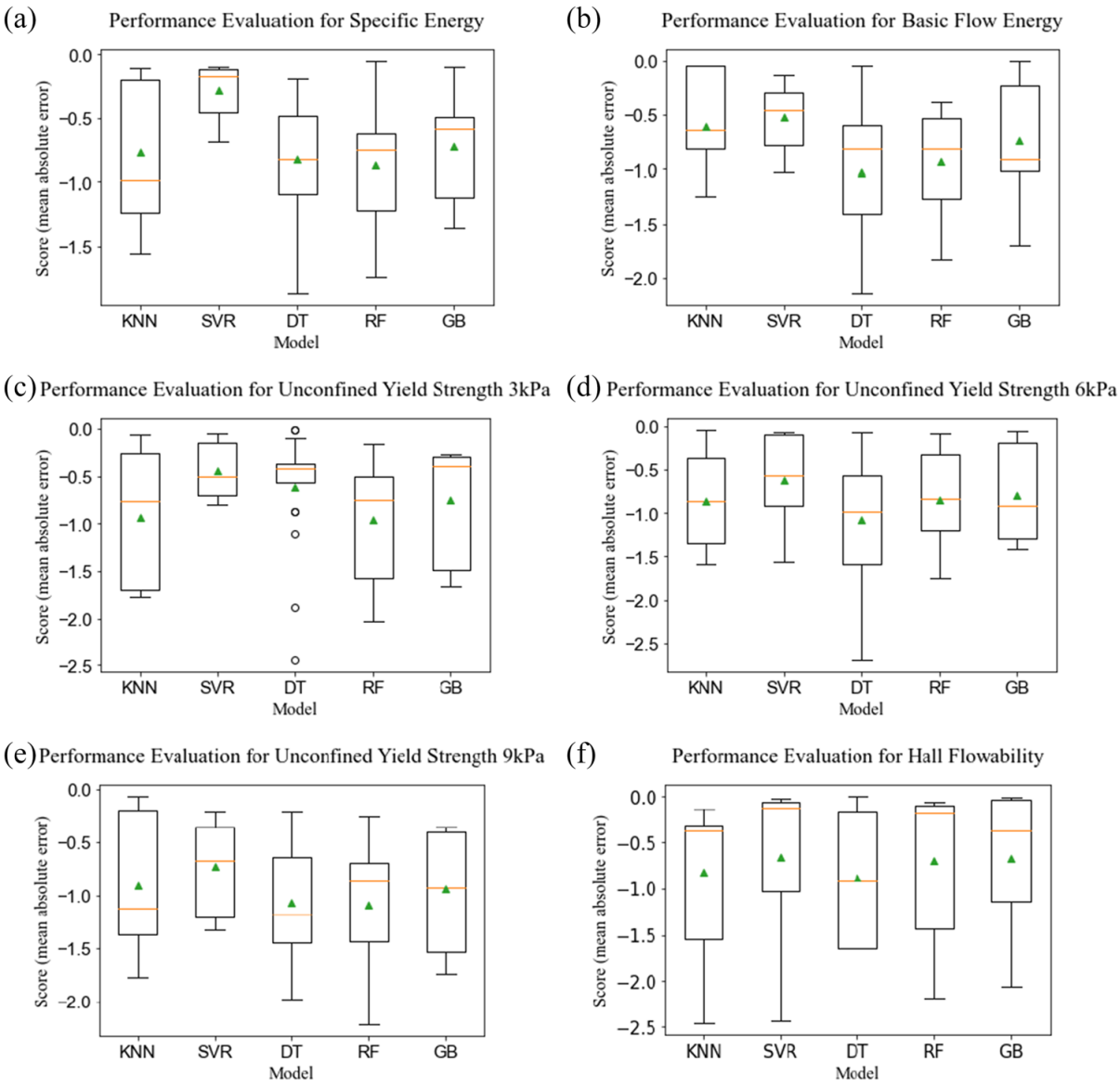


Figure 3-2; Overall comparison of various ML algorithms' performance (performed on the train samples of the original dataset) in leave-one-out CV for (a) SE, (b) BFE, (c) UYS3, (d) UYS6, (e) UYS9, and (f) hall flowability.

Dynamic Rheology (Specific Energy and Basic Flow Energy):

According to Figure 3-2(a) and Figure 3-2(b), the SVR algorithm outperforms other regressors in the prediction of dynamic rheology targets (SE and BFE). To analyze the model's generalization on all the 9 PSD ranges of Ti64 powder, SVR was trained and tested in all three performance evaluation modes

(RRSP, SSP, train-test split). The feature selection analysis for dynamic rheology targets indicated that the original set of input features (Table 3-2) consisting of size, shape, and density indicators perform well in the SVR algorithm for SE and BFE prediction. Therefore, the analysis proceeded with the same input feature arrangement. The tuned hyper-parameters of SVR algorithms applied for SE and BFE regression are illustrated in Table 3-3.

Table 3-3: Tuned hyper-parameters of SVR regressors (using the original dataset) used for the prediction of SE and BFE targets.

Target	C	Degree	Epsilon	Gamma	Kernel	Tolerance
SE	1000	_	0.01	0.001	RBF	0.1
BFE	0.1	2	0.1	1	Poly	0.1

According to the results of the SSP evaluation of the SVR model shown in Figure 3-3(a) and Figure 3-3(c), the correlations between the input features set of powder properties (PSD, shape, and packing) and the SE and BFE targets have been successfully determined, and the trend of the predictions aligns with the actual trend in both targets' measurements.

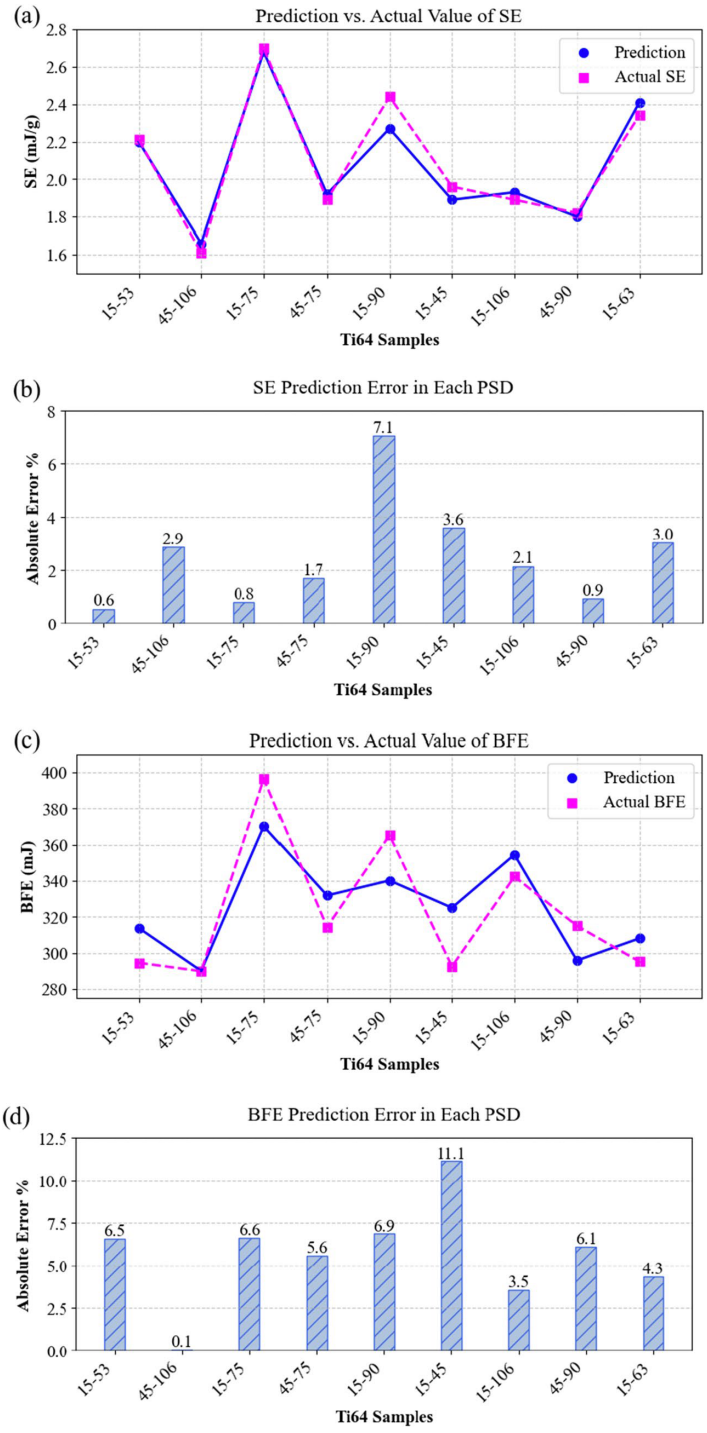


Figure 3-3: Performance of SVR regressor in prediction of dynamic rheology targets. (a) predicted SE vs. the actual SE, (b) MAPE of SE prediction in SSP mode, (c) predicted BFE vs. the actual BFE, and (d) MAPE of BFE prediction in SSP mode.

Figure 3-3 (b) indicates that the prediction error for SE (SSP mode) is around 3% for most samples except the 15-90 μm Ti64 sample with the error of 7.07%, which still counts as a relatively accurate performance. After that, 6 iterations of random splits (RRSP implemented by 6 repeats of random 3-fold CV) were also conducted, which yielded an overall MAPE of 2.62% that confirms the good generalization of the SVR model on the current dataset.

It is apparent from the error percentages of BFE target prediction shown in Figure 3-3(d) that SVR performance on BFE is not as accurate as SE performance. Complications associated with the confined dynamic flow measurements could lead to more complex underlying relationships between physical properties and basic flow energy, making the regression more challenging for ML algorithms. Nevertheless, given a bigger train set of samples was available (SSP and train-test split modes of evaluation), the SVR model achieved acceptable overall performance, as the obtained MAPE level of 5.64% (SSP) and 4.77% (train-test split) listed in Figure 3-4 implies. Further assessment of the model's generalization capability was conducted through the implementation of the RRSP mode.6 iterations of random splitting of the dataset (6 repeats of 3-fold CV) resulted in an overall performance of 12.78% MAPE for BFE prediction. Although train sample reduction affected the overall performance here by causing overfitting in the model, the model's generalization remains relatively acceptable considering the data limitation. The results in all three modes indicate that data expansion through the addition of new experimental data can reduce the current overfitting.

Generally, the predicted errors associated with SE regression analysis are below 3% for all three modes, signaling the good generalization of the SVR model in SE prediction of 9 described PSD ranges of Ti64 and potentially in unseen samples with different PSD ranges.

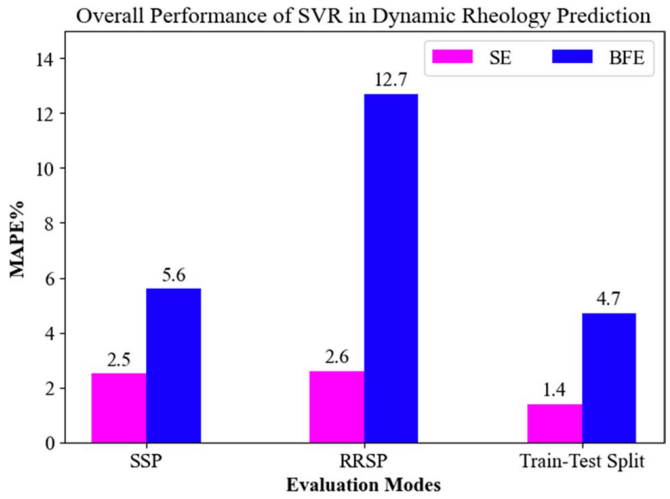


Figure 3-4; Assessment of the SVR regressor's performance (using the original dataset) across three performance evaluation modes for SE and BFE target regression.

Shear Rheology (Unconfined Yield Strength):

UYS characteristics are retrieved from yield locus calculations applied on shear stress measurements of powders in three initial compression stresses of 3 kPa, 6 kPa, and 9 kPa to investigate powders' shear behavior in different consolidations/compressions. The results of hyper-parameters tuning of the SVR model by Leave-one-out CV are shown in Table 3-4. The same set of 9 input features (Table 3-2) is also used for shear rheology targets regression.

Table 3-4: Tuned hyper-parameters of SVR regressors used for the prediction of UYS3, UYS6, and UYS9 targets using the original dataset

Target	C	Degree	Epsilon	Gamma	Kernel	Tolerance
UYS3	10	_	0.5	'scale'	Sigmoid	0.1
UYS6	1000	_	0.1	0.0001	Sigmoid	1e-4
UYS9	1e-5	4	0.1	10	Poly	0.01

It is evident from Figure 3-2(c)-(e) that the SVR model demonstrates a higher prediction accuracy than other assessed regressors. Nevertheless, it should be noted that the MAPE metric values obtained from SVR here are twice and three times those obtained from the dynamic rheology targets regressions.

Furthermore, since performance in SSP mode guarantees the LPBF users' interest, it is required to be adequately accurate. However, as the results demonstrated in Figure 3-5 indicate, the poor performance of the SVR model in the regression of UYS targets is also observed in the SSP mode of performance evaluation.

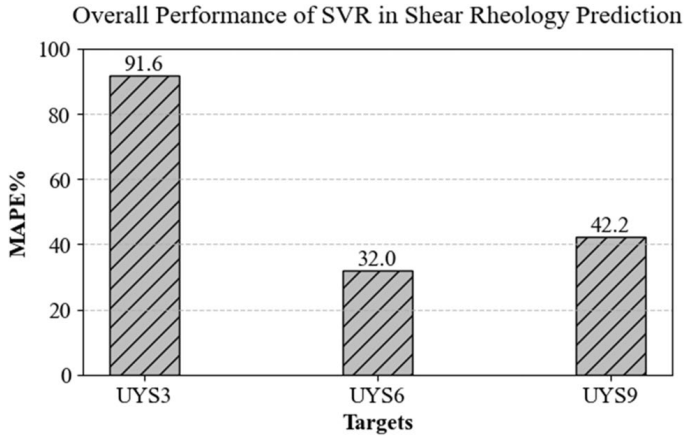


Figure 3-5: Performance of the SVR regressor (using the original dataset) on each UYS target in SSP mode. The plot shows the MAPE of predictions averaged across all samples.

Following a thorough investigation into identifying suitable models or techniques for formulating the complex relationship between input and output features using small datasets, GP-SR was determined to be practical and was employed to predict the UYS targets. Regression using the GP-SR technique aims to identify and derive the underlying relationships within a dataset without making any prior assumptions about the mathematical expression of the data-fitting model. It operates by exploring various mathematical fitting functions (programs) rather than optimizing the coefficients of an initial function generated based on a priori model assumptions. The fitness of each generated population/program is evaluated by the defined error metrics such as mean squared error, mean absolute error, etc.

Since conventional ML algorithms such as SVR could not accurately derive the correlation between powder physical properties and UYS targets, the GP-SR was employed for the purpose. The derived fitting equations were then practiced for the estimation of targets for each sample.

The GP-SR model is first trained on the train set (7 authors-selected samples) to identify the best fitting function. The resulting mathematical equations for UYS3, UYS6, and UYS9 targets are shown in equations (3-1), (3-2), and, (3-3) respectively.

$$UYS3 = sin(-D90) \quad (3\text{-}1)$$

$$UYS6 = -D90 \times \sqrt[2]{Carr's\ index} \quad (3\text{-}2)$$

$$UYS9 = (D10 \times D90) + \sin(Span - \sin D90) \quad (3\text{-}3)$$

It could be interpreted from the equations that D90 has the highest impact on the estimation of UYS targets. It is also seen that a sinusoidal trend of the UYS targets in current data is accurately determined by the GP-SR equations in two out of three targets (shown in Figure 3-6(a) and (c)). Although Sinus was available for all three targets as one of the mathematical units upon which the GP-SR model generates programs, however, sinus was not passed to the final mathematical expression fitting the UYS6 target. Multiple trial runs were tested using candidate mathematical equations consisting of sinus, but acceptable accuracy was not obtained.

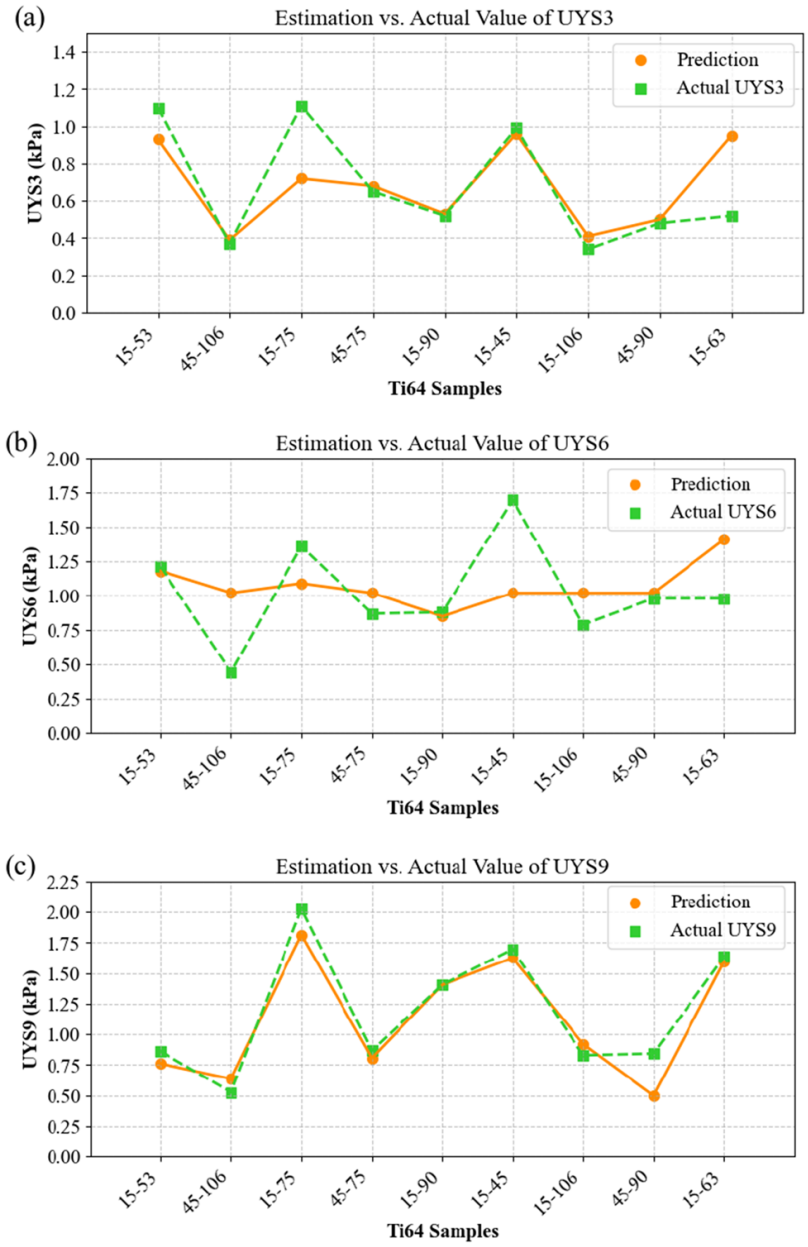


Figure 3-6: Estimations made by the GP-SR equations compared to the actual values of the sample for (a) UYS3, (b) UYS6, and (c) UYS9 targets in SSP mode.

The results illustrated in Figure 3-6 indicate that the GP-SR method was able to extract a relatively accurate fitting equation for the estimation of UYS at 3kPa and 9kPa. The trend of the plotted estimations of these two targets exhibits high similarity to the variations observed in the actual values' trend.

Although the performance of GP-SR is notably better on the lowest and highest consolidation states, it was not able to draw an accurate fitting function for UYS in middle-level consolidation (6 kPa). Similar observations are apparent in Figure 3-6(b) suggesting that the trend of the estimated values remains relatively stable and does not closely mirror the trend of the actual values. Further investigation into the performance of GP-SR for the UYS6 target and potential causes of poor accuracy are discussed in the following sections.

The application of GP-SR for UYS3 estimation across all nine samples yielded a MAPE of 18.93%, representing a significant improvement over the MAPE of 91.61% obtained from the SVR model. (The MAPE results of the analysis are depicted in Figure 3-5).

In Figure 3-7(a), it is evident that the sample with a PSD range of 15-63 µm exhibits the highest proportion of error in the performance of the GP-SR model. In general, three main factors were identified to impact the accuracy of shear flow estimation: 1- level of uncertainty in experimental measurements (standard deviation of measures in each target), 2- extreme points of the data, and 3- the complex correlation between the compressibility (Carr's index) and the consolidation factor involved in shear rheology test.

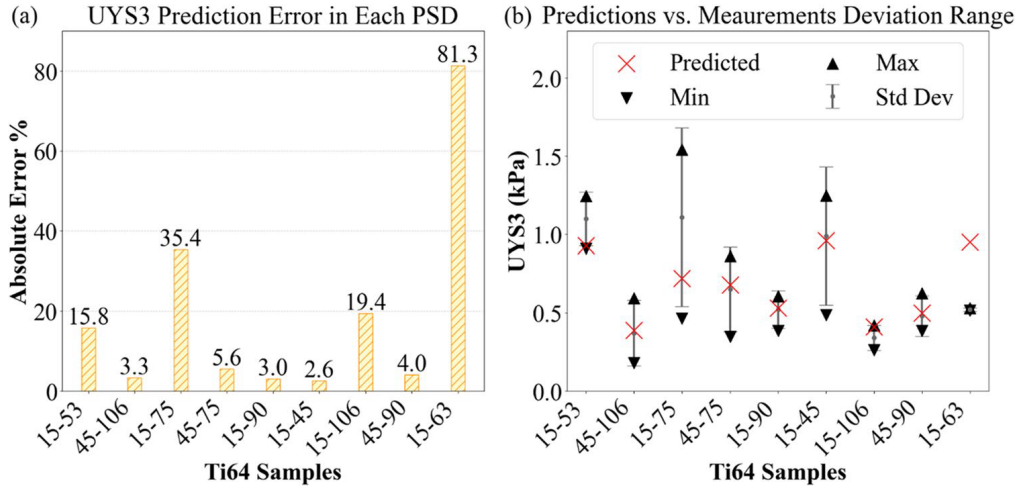


Figure 3-7: Assessment of GP-SR performance by analyzing the error in UYS3 estimations and the uncertainty in the experimental measurements per sample. (a) GP-SR estimation errors, and (b) comparison of estimated values to the min-max interval and deviation of the experimental measurements.

Firstly, it should be noted that the significant difference between the 15–75 µm sample behavior and the rest of the samples affects the prediction of the 15-63 µm sample because it has similar physical/morphological characteristics to the 15-75 µm sample while exhibiting different rheology (shown in Figure 3-8(a)). To provide further clarification, a sudden drop in the UYS level of the 15-63 µm sample is depicted in Figure 3-8(a). This deviation is not consistent with the UYS levels observed in samples with smaller PSD ranges nor with the subsequent sample (15–75 µm). This discrepancy suggests that either the 15-75 µm sample should exhibit a lower UYS level, or conversely, the 15-63 µm sample should demonstrate a considerably higher UYS level. However, considering the high level of uncertainty associated with the measurements for the 15-75 µm sample, which is not observed in the 15-63 µm sample, it can be inferred that imprecise data for the 15-75 µm sample contributed to the poor estimation performance for the 15-63 µm sample. Additionally, as Figure 3-8(b) shows, the powder samples of 15-63 µm and 15-75 µm both have high compressibility, which causes complex and rather unjustifiable rheological behavior in each consolidation state. It is due to the amount of air trapped in bulk powder in various consolidations which impacts fine and coarse powder samples' shear rheology differently. Also, as the compressibility level increases in powder, in higher consolidation levels more air trapped in the powder will be released, which ultimately causes more dramatic complexity in the shear behavior of the powder. This even leads to a notable level of uncertainty in the experimental shear measures of these samples.

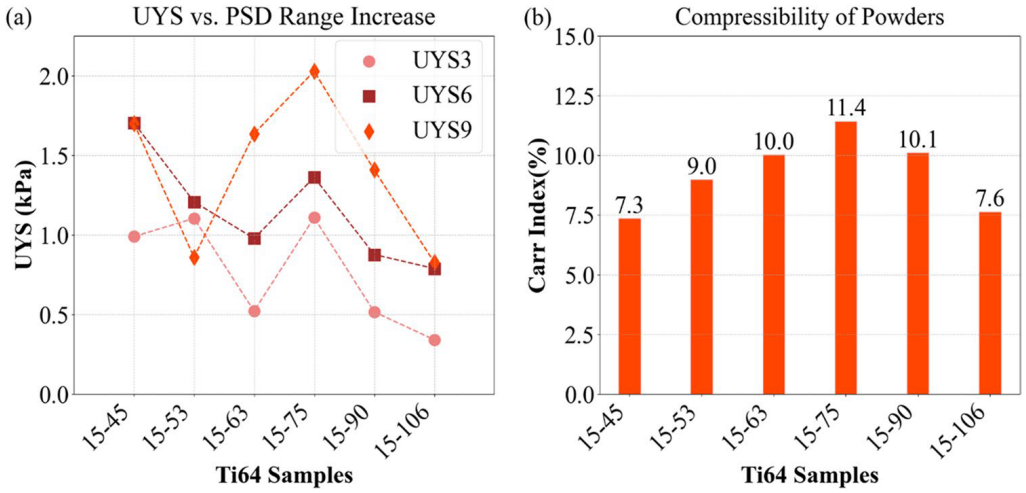


Figure 3-8: Illustration of (a) shear behavior variations (for samples starting from the 15 µm size range) as the PSD range increases, and (b) the compressibility index of each corresponding sample.

Since measurement uncertainty in experimental data affects the performance of regressors and the equation-fitting models, it was taken into account for performance analysis as well. Accordingly, the GP-SR estimations were compared with the measurements' standard deviation interval and the minimum and maximum measures obtained per sample. This is to clarify whether the obtained results are in the logical range to be acceptable. As shown in Figure 3-7(b), although notable uncertainty of measurements (high standard deviation) could be seen in the behavior of most samples from the 15 μm starting size range (except 15-90 μm), in most cases (except for the 15-63 μm) the estimations of the UYS3 target are included in the target's experimental measures interval (min-max range), demonstrating the acceptable performance of the model given the measurement uncertainties.

All these arguments could lead to low prediction accuracy, especially in cases of limited data samples. Adding new powder samples with similar PSD ranges and new different PSD ranges to the recent data would significantly reduce the impact of individual samples on the overall performance of regression models, resulting in more reliable generalizations. However considering the recent dataset challenges and uncertainty regarding the repeatability of UYS measurements obtained by FT4, GP-SR yielded equations are showing relatively accurate estimations for UYS targets, which could not be obtained by other ML models.

The GP-SR model resulted in inadequate accuracy in UYS6 data fitting and estimation. One of the primary factors contributing to this may be attributed to the fact that UYS6 measurements were conducted twice, as opposed to UYS3 and UYS9, which were repeated three times. This occurred due to the inherent settings of the FT4 measuring device during the experimentation period. Consequently, the increased level of measurement error and uncertainty associated with UYS6 resulted in less precise outcomes from the GP-SR method.

As shown in Figure 3-5, the MAPE (32.84%) obtained in UYS6 estimation using the GP-SR method is significantly inaccurate compared to the results obtained on UYS3 and UYS9. Figure 3-6(b) demonstrates a correlation between the trend fluctuations of the actual measures and the estimations, but the magnitude of the fluctuations in estimations and actual value trends does not match. Besides the lack of sufficient samples, the major challenge of predicting extreme points may be responsible for this phenomenon.

As illustrated in Figure 3-9(a), Ti64 samples of 45-106 μm and 15-45 μm demonstrate high estimation error percentages. This is due to their positions as the maximum and minimum extreme points within

the dataset. According to Figure 3-9(b), among these three samples, 15-45 µm also have a high level of uncertainty in FT4 measurements which explains the high level of error. Sample 15-63 µm also accounts for a high level of error in UYS6 estimations, due to the similar potential causes mentioned in the UYS3 results section. However, it could also be drawn from the figure that most of the estimated values (except for the 45-106 µm and 15-75 µm) are within the min-max range per sample. The unusual rise of the UYS6 level (shown in Figure 3-8(a)) in the 15-75 µm sample; potentially due to the highest level of compressibility; negatively affected the accurate equation fitting here as well.

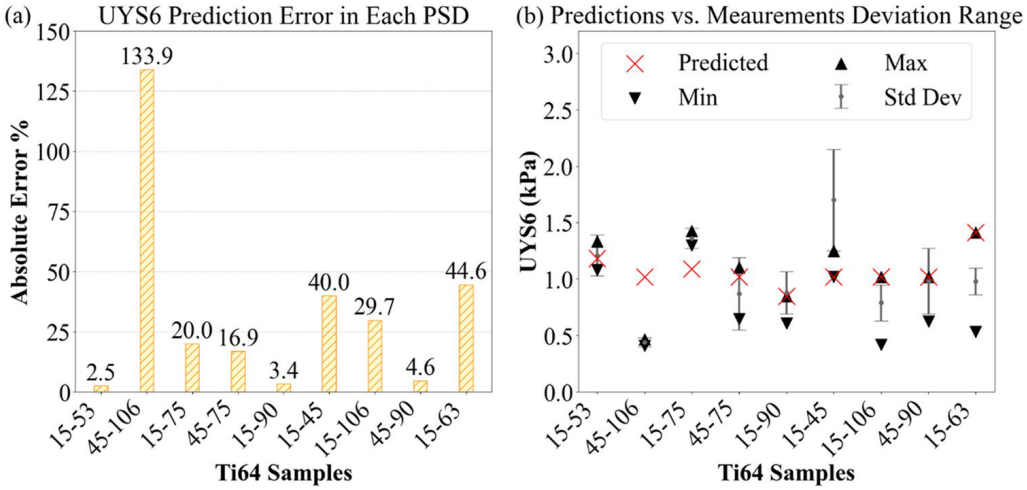


Figure 3-9: Assessment of GP-SR performance by analyzing the error in UYS6 estimations and the uncertainty in the experimental measurements per sample. (a) GP-SR estimation errors, and (b) comparison of estimated values to the min-max interval and deviation of the experimental measurements.

In conclusion, prioritizing the inclusion of new actual samples to reduce the sensitivity of models to individual samples in UYS6 modeling should be the foremost objective in future research. Additionally, further investigation is necessary to extract a relevant and effective set of input features, potentially derived from the shear rheology test factors, to address the dynamics of powder compressibility and the applied consolidation levels. Ultimately, the equations derived in this study could potentially replace some of the input features in the machine learning models for future work.

GP-SR method achieved the best performance in the estimation of the UYS9 target. The resulting mathematical equation accurately estimated UYS9 measurements for all nine samples. This superior performance is likely due to the correlation between compressibility and trapped air in the samples and

the impact of consolidation on these factors. It is expected that UYS measurements with less uncertainty can be obtained at the highest consolidation level (9 kPa), as most trapped air is released, and the powder behavior can reach greater stability at this level. Conversely, the greatest instability is observed at the lowest and mid-level consolidations, particularly in samples with higher compressibility, as noted in previous results, specifically for the UYS6 target.

Figure 3-6(c) illustrates the similarity of the UYS estimation trend to the actual measures and the small difference between the estimations and the actual values. Based on Figure 3-10(a), the GP-SR model yielded an acceptable level of error in estimation for most of the samples, indicating good accuracy in estimating the correlation between physical properties and UYS9. The highest error is associated with the 45-90 µm and 45-106 µm samples. This could be attributed to the imbalanced data, as there are fewer samples in the coarse PSD range (45-106 µm) compared to the number of samples in the fine PSD interval (15-106 µm). This condition could cause the model to not properly account for these samples in the modeling or result in poor accuracy in learning the patterns between the inputs and targets.

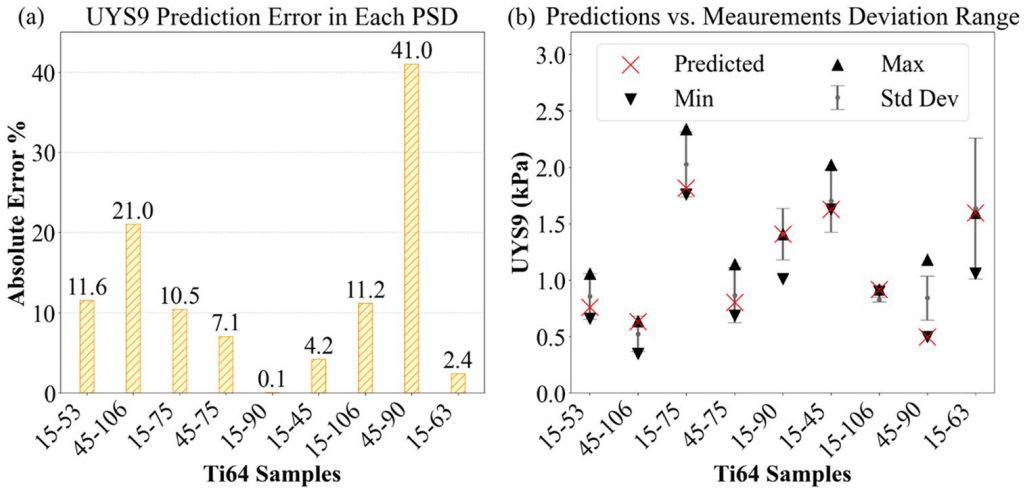


Figure 3-10: Assessment of GP-SR performance by analyzing the error in UYS9 estimations and the uncertainty in the experimental measurements per sample. (a) GP-SR estimation errors, and (b) comparison of estimated values to the min-max interval and deviation of the experimental measurements.

The correlation between compressibility and consolidation could play a major role in the high error of the 45-90 µm sample too. According to Figure 3-11(b), this sample has the highest Carr's index

(compressibility) among other samples in the coarse size range, therefore causing more irregular shear behavior as the consolidation increases. The trend of UYS6 variations of this sample in Figure 3-11(a) confirms this too as the unexpected rise is seen in the UYS6 level of 45-90 µm sample.

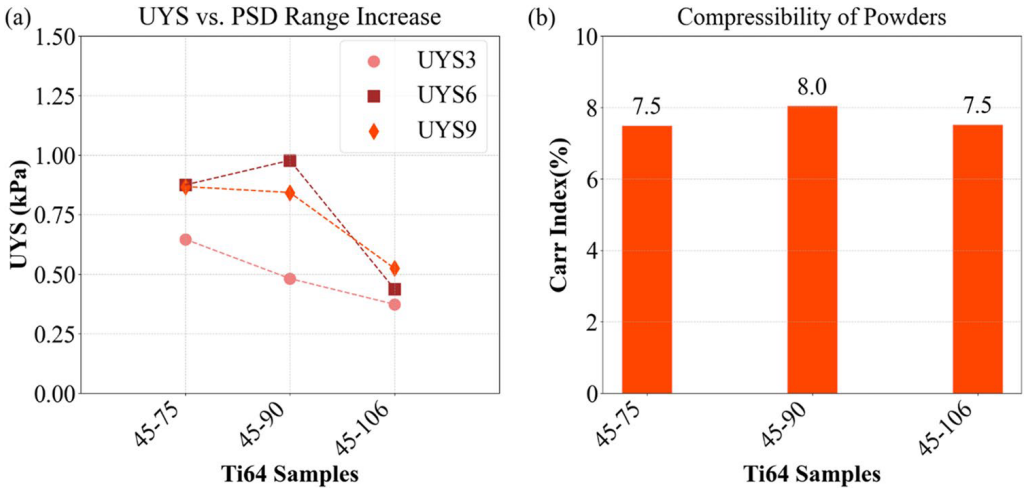


Figure 3-11: Illustration of (a) shear behavior variations (for samples starting from the 45 µm size range) as the PSD range increases, and (b) the compressibility index of each corresponding sample.

Nevertheless, since the GP-SR estimations for all samples fell in the min-max interval of experimental measurements, an acceptable level of accuracy was reached in the results. The estimations were also contained in the deviation interval of the actual values except for the 45-90 µm sample, as expected. Furthermore, Figure 3-12 shows the 12.1% overall MAPE obtained by the GP-SR fitting equation (on the whole data set), which compared to the SVR model's performance (42.2% MAPE) implies the superior performance of the GP-SR method in the UYS9 target. This could be considered a satisfactory result despite the challenges of data limitation and potential powder characterization measurement uncertainty affecting the regression models' accuracy and generalization.

In general, as it is reviewed in Figure 3-12, it was concluded that the GP-SR formulation of physical properties and UYS targets could impressively enable an admissible shear rheology modeling despite the limitations of the dataset and the discussed complexities present in the shear rheology data (the measurement uncertainties and the compressibility-consolidation dynamic factor). GP-SR technique yielded notable improvement in estimations except in UYS at 6kPa compression, which is almost as accurate as the results obtained by the SVR model. The lack of reliable performance for UYS6 could serve as a subject for future studies.

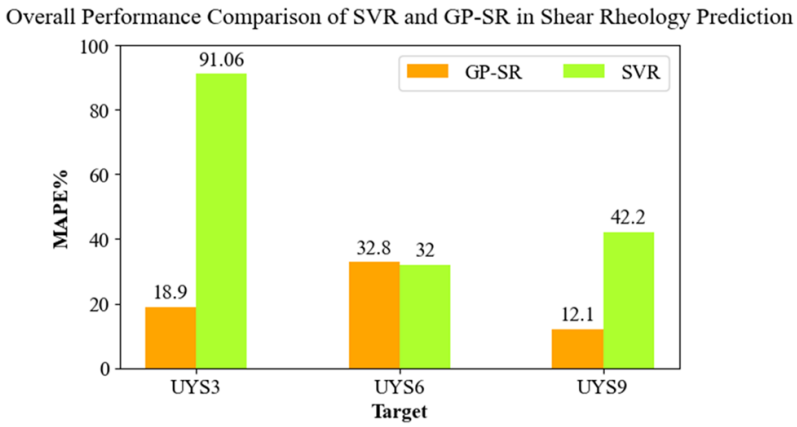


Figure 3-12: Overall performance comparison between SVR and GP-SR techniques (using the original dataset) in estimating the shear rheology targets in SSP mode.

3.3.2 Flowability Prediction Using Data Augmentation

Data augmentation was investigated in this study as the last solution to compensate for the effect of data limitation on the performance of predictive algorithms. The earlier methods of training simple ML regressors, using various evaluation modes, and developing a GP-SR equation-fitting algorithm achieved satisfactory rheology predictions for Ti64 powder in PSD ranges of 15 to 106 μm in most targets. However, these methods relied on the limited original dataset.

Since these techniques couldn't yield the same level of accuracy in BFE and UYS6 as in SE, UYS3, and UYS9, the SMOTE data augmentation technique was practiced here as the ultimate solution. First, the original data was expanded with synthetic samples, then the best regressor algorithm was trained for these two targets to assess whether data augmentation could enhance the performance of the regressors. The SVR algorithm was trained on the augmented dataset (with the pre-defined arrangement of 9 input features) and its generalization accuracy was evaluated through RRSP along with train-test split mode in BFE regression, and RRSP was assessed in UYS6 regression.

Moreover, hall flowability regression was also carried out by relying on SMOTE data augmentation. Figure 3-2(f) demonstrates that the SVR regressor exhibits superior accuracy in hall flowability prediction relative to the other models when the original dataset is utilized for algorithmic development. However, the SVR regressor showed poor generalization after hyper-parameter tuning, model optimization, and feature selection, leading to the application of data augmentation as the last examined solution in this study for hall flowability regression.

During the synthetic data expansion process for hall flow, BFE, and UYS6 targets, SMOTE generated 28 new samples using the 7 training samples of the original dataset. The two authors-selected unseen samples mentioned beforehand were kept as unseen data and were not included in the data generation process. In the end, the 28 new synthetic samples and the original 7 samples were combined to be used as the new train set (35 samples).

3.3.2.1 Machine Learning Optimization and Performance Evaluation Discussion

The ML models' performance evaluation and comparison using the augmented dataset was conducted in the same approach as the previous analysis utilizing the original dataset. However, the augmented data offers more opportunities for performance evaluation across multi-sample sets while it was not as feasible in dynamic and shear flow analysis with the original dataset restrictions. Moreover, relying on the leave-one-out CV for the selection of the best regression model for the hall flowability target failed to provide acceptable results. Hence, the repeated K-fold CV (3 repeats of 9-fold) was used for model selection, aiming to amplify the significance of the models' generalization capability (across multiple unseen samples).

The same hyper-parameter tuning and feature selection procedure (similar to the approach carried out for the previous targets' algorithm optimization) was undergone, resulting in SVR having the best performance using the full set of pre-defined input features/predictors (size, shape, and density indicators) for BFE and UYS6 targets.

As mentioned before, among the techniques used previously in the dynamic and shear flow targets prediction evaluations, repeated K-fold CV was carried out here (instead of leave-one-out) but with some user-defined modifications. The reason for the modifications was the high resemblance of the newly generated samples to the original samples' values and statistical characteristics. This means that each of the original samples is reproduced with some deviations (obtained by analytical calculations of the SMOTE method) every time the SMOTE process is iterated (4 times in this case). Therefore, counting the model's performance on both the new synthetic sample (that is generated based on an original sample) and the original sample itself would be redundant and cause misjudgment. In other words, predicting hall flowability, BFE, or UYS6 for a test sample could become unrealistically accurate and biased when the training set contains multiple highly similar samples. As an alternative, we modified the repeated k-fold CV to address this issue and enable a proper and unbiased performance and generalization evaluation. The generated synthetic samples (not the samples from the original

dataset) which were generated from the test samples at hand were removed from the train set. This approach allowed us to test if the inclusion of augmented samples obtained from the rest of the original samples (which are not in the test set) could truly cause any accuracy improvement in the prediction of the intended unseen/test samples (in which their corresponding reproduced samples (synthetic samples) are eliminated from the train set).

3.3.2.2 Flowability Prediction Results

Basic Flow Energy:

Data expansion here allowed for a more reliable RRSP and train-test split evaluation mode in order to assess whether SVR could generalize the BFE prediction. The tuned SVR model (hyper-parameters shown in Table 3-5) was lastly employed on the augmented dataset to regress the BFE target once again.

Table 3-5: Tuned hyper-parameters of SVR regressors used for BFE target using the SMOTE augmented data

Target	C	Degree	Epsilon	Gamma	Kernel	Tolerance
BFE	10	_	0.1	0.1	RBF	1e-3

The obtained results implied that SMOTE data augmentation notably improved the SVR algorithm's accuracy in BFE regression with a MAPE of 3.40% (in 18 times testing on random train-test subsets). As Figure 3-13 demonstrates, in a similar effort (18 iterations of algorithm random testing) using SVR and the original dataset a MAPE of 12.78% was achieved. This error reduction indicates the effectiveness of data augmentation in performance improvement of SVR regressor. This was further confirmed by the conventional train-test split mode which achieved a MAPE of 3.13% (4.77% was achieved in the previous attempt using the original dataset).

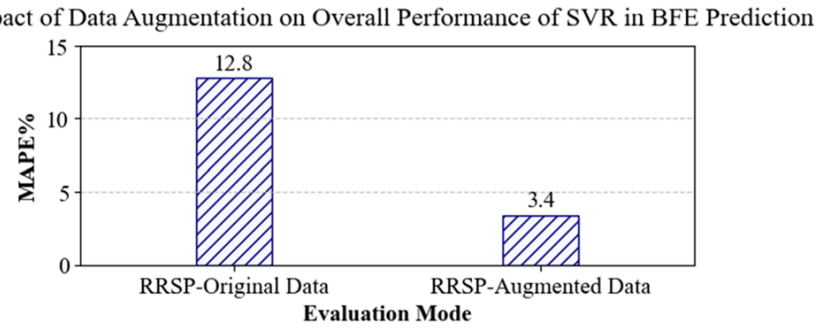


Figure 3-13: Overall comparison of the SVR algorithm's performance in BFE regression across two repetitions of 9-fold CV, using the original dataset versus the SMOTE-augmented dataset.

Unconfined Yield Strength at 6kPa:

Although there was a significant improvement observed in BFE modeling, the utilization of data augmentation did not lead to as pronounced overfitting reduction in UYS6 regression. SVR was tuned (shown in Table 3-6) and optimized on the expanded data and achieved a MAPE of 27.9% in the RRSP mode (18 times of random testing).

Table 3-6: Tuned hyper-parameters of SVR regressors used for UYS6 target using the SMOTE augmented data

Target	C	Degree	Epsilon	Gamma	Kernel	Tolerance
UYS6	1000	_	0.1	0.01	RBF	1e-3

This showed a slight improvement (around 5% error reduction) compared to the GP-SR result (result comparison shown in Figure 3-14). This concludes that data expansion could be a relatively effective option to attain more accurate results in UYS6 regression too, but major improvement could potentially be obtained when actual new data points/measures could be added to the data. Further investigation on more robust and accurate shear flowability estimation, particularly in middle range consolidation (6kPa), could be developed focusing on data expansion through experimental powder characterization and more advanced data augmentation techniques in the future. Moreover, other feature engineering and data processing methods could be employed for generation of more effective input futures to compensate the limitation of the data samples.

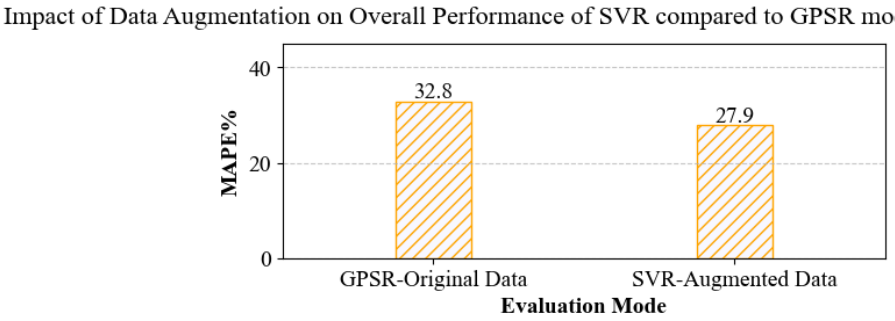


Figure 3-14: Overall performance comparison between the SVR algorithm in UYS6 regression using the SMOTE-augmented dataset and the GP-SR algorithm using the original dataset, based on two repetitions of 9-fold CV.

Hall Flowability:

As illustrated in fig 2 (f), utilizing the original dataset, SVR regressor resulted in the best accuracy among other models for hall flowability modeling. Accordingly, hyper-parameters tuning (shown in Table 3-7) and feature selection was then implemented by SVR model.

Table 3-7: Tuned hyper-parameters of SVR regressors used for the prediction of hall flowability target using the original dataset

Target	C	Degree	Epsilon	Gamma	Kernel	Tolerance
Hall Flow (Original Data)	0.1	2	0.01	Scale	Poly	1e-3

Regarding the input feature selection, as demonstrated in Table 3-8, the best result was obtained by the selection of 5 input features out of the 9 author-defined input features.

Table 3-8: Optimum set of features obtained by sequential feature selection for Hall flowability regression using the original dataset

Input Feature Category	Particle Size Distribution			Particles Morphology	Packing/Density
Input Feature Name	D10	D50	D90	Symmetry	Carr's Index

SVR model resulted in a ~26% MAPE in the SSP mode. This concludes the high overfitting of the SVR model due to the lack of sufficient training samples and makes it ineligible for the aimed powder characterization in LPBF applications task (SSP priority in LPBF applications). Similar to the same approach adopted for the shear rheology estimation, GP-SR technique was also trained and tested for hall flowability which resulted in an overall estimation error of ~25.5% MAPE. Therefore, due to the poor outcome, the GP-SR technique was also rejected as a reliable approach for hall flowability regression.

Accordingly, to examine the last proposed methodology in this study, the generation of synthetic samples/data augmentation was adopted in this case, as well. The SMOTE technique was employed here to generate synthetic samples. The resulting synthetic data were expected to allow for the reduction of the model's sensitivity to the original samples and consequently reduce the overfitting of the regressors.

As Figure 3-15 indicates, the SVR regressor was selected as the best ML algorithm for hall flowability regression using the augmented dataset.

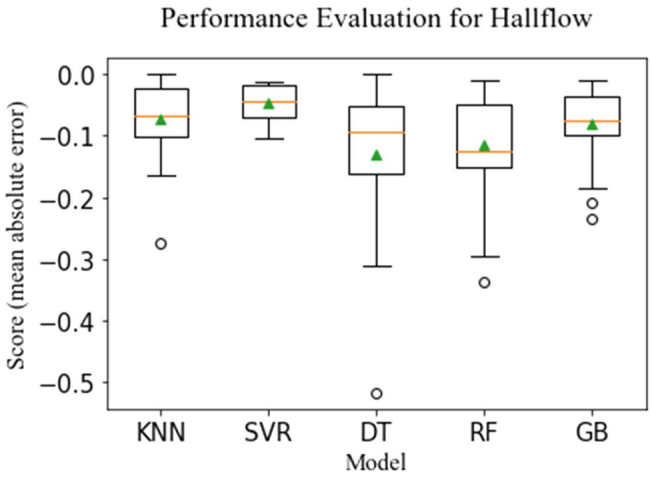


Figure 3-15: Overall comparison of various ML algorithms' performance (trained on SMOTE augmented dataset) in hall flowability regression.

After performing the SMOTE data augmentation, the hyper-parameters tuning of the SVR model (listed in Table 3-9) was followed by the sequential feature selection step.

Table 3-9: Tuned hyper-parameters of SVR regressors used for Hall flowability target using the SMOTE augmented data

Target	C	Degree	Epsilon	Gamma	Kernel	Tolerance
Hall Flow (Data Augmentation)	10	_	0.01	0.01	RBF	1e-4

Similar to SVR regressor feature selection results using the original dataset, SVR once more achieved better accuracy of regression with a subset of 5 input features (selected input features demonstrated in Table 3-10) using the augmented data.

Table 3-10: Optimum set of features obtained by the sequential feature selection for Hall flowability regression using the SMOTE augmented dataset

Input Feature Category	**Particle Size Distribution**		**Particles Morphology**		**Packing/Density**
Input Feature Name	D50	Span	Sphericity	Symmetry	Tapped Density

The performance evaluation in hall flowability modeling was analyzed through the same modes of 1-'SSP', 2-'RRSP' (using the modified repeated k-fold CV on the whole dataset of 37 samples composed of the original train set, synthetic, and authors-selected test samples), and 3-'train-test split' (two authors-selected unseen/test samples which were not involved in any conducted data augmentation and model optimization steps).

As shown in Figure 3-16, using SMOTE data, SVR yielded ~20% MAPE in SSP mode, indicating that model could reduce the impact of overfitting by SMOTE data to some extent, compared to ~26% MAPE obtained when SVR was optimized and trained using only original samples. In addition, the results obtained from the first mode (RRSP) with 2 repeats of 9-fold CV (in general testing 18 randomly selected combinations of train and test samples) yielded the final MAPE of 10.13%. This implies that the MAPE rate is reduced by around 8% (performance comparison illustrated in Figure 3-16 compared to when the SVR regressor was applied only to the original dataset. According to the outcome of SSP

and RRSP evaluation modes, this also could be interpreted that inclusion of higher numbers of synthetic samples could enable the model for better regression accuracy. However, it should be noted that generating more synthetic samples from a limited number of original samples could also lead to an ineffective repeat of similar information which could make the model biased. Therefore, further improvement of performance relying on data generation requires the availability of new actual samples to generate effective synthetic samples.

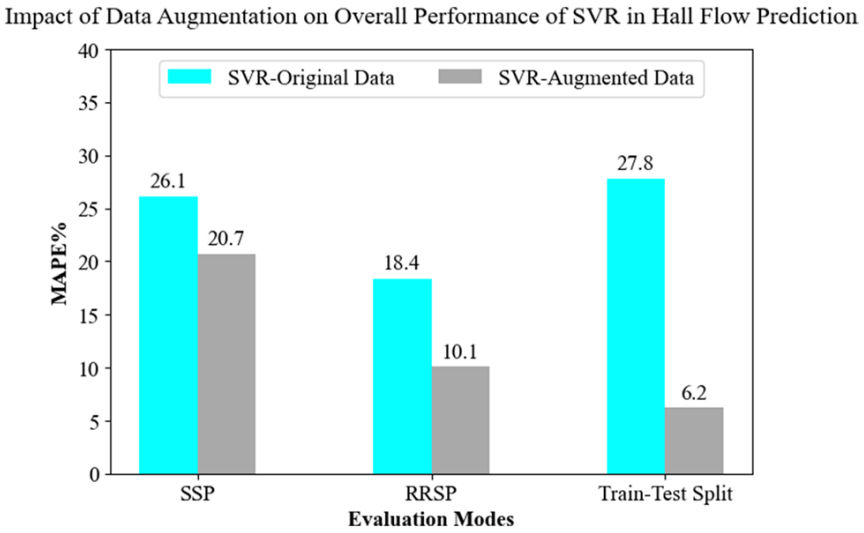


Figure 3-16: Overall performance comparison of the SVR regressor in hall flowability prediction across three evaluation modes using the SMOTE-augmented dataset.

Data augmentation facilitated a more accurate evaluation of the model by using a train-test split approach. The training set comprised synthetic data along with seven original samples, while two pre-defined, unseen samples were used as the test set. The tuned SVR regressor operating on the optimum selected set of inputs (shown in Table 3-10) was trained on the train set. The trained model was then examined on the test samples of 15-63 µm (fine standard powders), and 45-90 µm (coarse off-size powder) which according to Figure 3-16, an average error of 6.2% (MAPE) was ultimately obtained. According to Figure 3-17, the error primarily stemmed from the prediction of the sample in the 15-63 µm range, which exhibited a 9.7% MAPE error. Notably, this sample represented an extreme point with the highest level of hall flow, even within the augmented dataset. Additionally, it was not augmented by SMOTE, as unseen samples were excluded from data augmentation. These factors likely contributed to the reduction in regression accuracy to some extent.

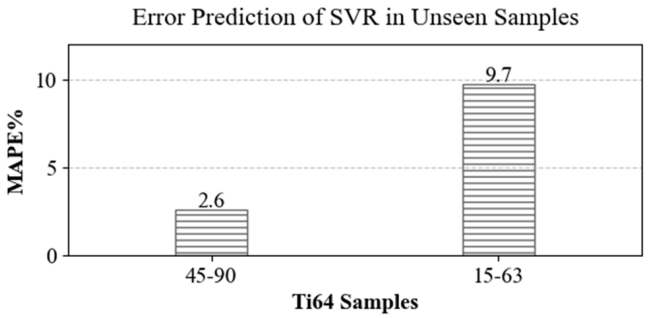


Figure 3-17: Error of hall flowability regression in unseen samples using the SVR regressor on the SMOTE augmented data.

In summary, regression using conventional ML algorithms and the GP-SR while relying on the original dataset was incapable of retrieving accurate underlying correlations between the powder physical and packing properties and the hall flowability target. The causing reason could be related to two significant factors: 1-highly limited data samples and 2- potentially complex underlying physics of hall flowability measurements, which could not be apparent to the users. However, as a solution, synthetic data generation using SMOTE reproduced the available information in the form of new samples with slight deviations to decrease the overfitting of regressors. Although no new actual samples with new measures were generated, the overall accuracy of the SVR regressor (specially in unseen data test mode) appeared to be rather successful in regressing the hall flowability target.

It is essential to note that although data augmentation was employed in this study, the dataset size still counts as highly restricted for ML regressors. In other words, synthetic data generation only partially reduced model overfitting to the samples. Moreover, the reported results are specific to the defined powder set (Ti64 in the 15-106 μm size range). Accurate regression of hall flowability in SSP mode was not achieved by any of the explored methods. The potential causes for this limitation, based on the available dataset, were discussed, but further investigation is recommended for future research. Employing advanced data augmentation techniques, such as conditional generative adversarial networks (CGAN), could address the insufficient information for reliable data fitting. Furthermore, reducing hall flowability measurement errors through higher repetition can yield data with more reliable underlying patterns. For powder flowability and rheology targets such as hall flowability and UYS6, the lack of sufficient data appears to be the primary source of poor performance in ML algorithms. Consequently, aside from other aiding solutions such as employing effective feature engineering methods, further data collection in new PSD ranges remains the essential future step for achieving

satisfactory data fitting and regression. However, the results indicate that such predictive systems can be developed for various material/powder sets using small datasets. This approach offers a highly cost-effective and time-efficient method for initially estimating powder behavior, allowing for modifications to the LPBF printing process accordingly.

3.4 Summary

In this chapter, our goal was to investigate the feasibility of machine learning (ML) regressors in predicting three flow indicators: dynamic, shear, and hall flowability. Regression models were trained using various particle size distribution (PSD) ranges of Ti6Al4V (Ti64) powder (with 9 practical PSD ranges in LPBF). To address the challenge of limited data, we employed multiple approaches, such as using simple/conventional ML regressors, equation fitting using genetic programming (GP), and data augmentation. As a summary, the proposed techniques of ML regressors (SVR) and genetic programming's equation fitting (GP-SR) operating on the available limited data could achieve sufficient accuracy in estimation of the dynamic and shear rheology targets for Ti64 samples in 15-106 μm size range. Data augmentation was then shown to be effective in improvement of the ML regressors' performance in situations where techniques relying on original dataset failed such as basic flow energy (BFE), unconfined yield strength at 6kPa consolidation (UYS6), and Hall flowability.

Chapter 4: Tailoring laser powder bed fusion process parameters for standard and off-size Ti6Al4V metal powders: a machine learning approach enhanced by photodiode-based melt pool monitoring

4.1 Abstract

In this chapter, machine learning (ML) models were trained, optimized, and evaluated using melt pool monitoring (MPM) data from photodiode sensors to customize laser powder bed fusion (LPBF) process parameters for both standard and off-size Ti6Al4V (Ti64) powder feedstocks. The study utilized three particle size distributions (PSD) of 15–53 μm, 15–106 μm, and 45–106 μm to fabricate cubical and cylindrical samples, enabling the development of a process parameter optimization system adaptable to varying powder sizes. Self-organizing maps (SOM) were employed to cluster light intensity signals into multiple classes for each part, from which statistical signatures were extracted. These signatures, along with powder size range, density, surface roughness, and hardness, served as inputs for Feed-Forward Neural Network (FFN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) regressors. The regressors predicted laser power, laser velocity, hatch distance, and energy density for the parts.

4.2 Methodology and experimental setups

4.2.1 Design of experiment

A commercial LPBF system (EOS M290, EOS GmbH, Krailling, Germany) was utilized for printing Ti64 parts. The powder bed recoating is handled by an in-chamber automatic system. The machine has an Ytterbium (Yb) fiber laser with 400 W nominal power and a laser wavelength of 1060 nm. Argon gas is employed in the chamber for Ti64 parts printing. Ti64 metal powder (AP&C, Boisbriand, QC, Canada) in three size ranges of 15–53 μm (fine batch/batch A), 15–106 μm (broad batch/batch B), and 45–106 μm (coarse batch/off-size/batch C) is used as the feedstock material to print 10 mm × 10 mm × 10 mm cubical and cylindrical samples with a height and diameter of 10 mm × 10 mm (a subset of printed parts is shown in Figure 4-1) using the Full Factorial Design (FFD) technique.

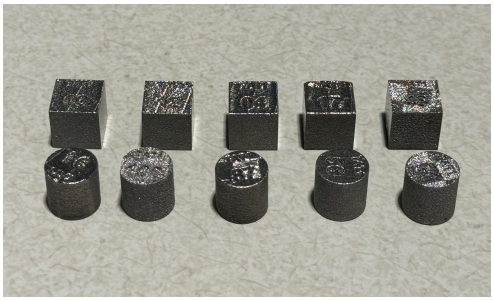


Figure 4-1: Printed cubical and cylindrical parts.

Three separate sets of prints were carried out for the printing of cuboid and cylindrical parts of 15–53 µm, 15–106 µm, and 45–106 µm powder batches, i.e., 54 cuboid and 54 cylindrical parts of each powder size range were printed in a single print.

Laser power (W), laser energy density (J/mm^2), hatch distance (mm), and scanning strategy are varied in the ranges listed in Table 4-1. The layer thickness of 40 µm is kept fixed in all prints.

Table 4-1: Design of the experiment

Tagging Class	**Power (W)**	**Energy Density (J/mm^2)**	**Hatch Distance (mm)**	**Scan Strategy**	
Low	150	1.5	0.08	Stripe	Chess
Mid	250	2.75	0.09		
High	350	4	0.1		

Laser velocity (mm/s) was determined according to specified energy density and hatch distance using equation (**4-1**). Parts geometry (cubical/cylindrical) and powder size range are the other changing variables in the design of the experiment.

$$Energy\ density = \frac{Laser\ Power}{Laser\ Velocity \times Hatch\ Distance} \quad (4\text{-}1)$$

In total, 324 parts were printed which consists of 108 parts (54 cubical and 54 cylindrical samples) printed with each of the three powder size ranges. The same variation of the process parameters is repeated for each set of 108 parts, allowing for inspection of the powder size range changes' influence on the final parts' quality and MPM data alterations.

The density, surface roughness, and hardness of the cuboid and cylindrical parts are measured by Archimedes, Keyence laser profilometer, and Phase II Rockwell hardness tester. The hardness and surface roughness are measured on the vertical sides. Nine points across the side surface of the cuboid and cylindrical parts were measured, controlled for potential noisy dataset elimination, and averaged to obtain the side surface roughness measures per part. Regarding the hardness measurements, the hardness in five points on the side surface of the parts were tested by the Phase II Rockwell equipment. The averaged hardness over the five measured points per part was then used as the hardness input feature of the algorithms.

4.2.2 Melt Pool Monitoring

Light intensity signals irradiated from the melt pool area were collected by an on-axis photodiode sensor during the printing of 108 parts of each powder batch (each single run of print), as shown in Figure 4-2. The on-axis photodiode sensor records the visible to NIR range of light intensity reflected parallel to the laser radiation direction. The on-axially recorded light intensity signals are gathered in the form of raw data containing the laser x-y location on the build plate, the modulation of the laser (on-off logs of the lasers), and the light intensity at each corresponding x-y location. Geometry and light intensity data corrections were applied to the raw MPM data to compensate for the chromatic aberration and position-related errors that appear in the light-intensity data recording due to the placement of the sensor and the difference between the optical elements of monitoring and the light wavelength (for more information about the melt pool monitoring technique and the data corrections, refer to [106]). Additionally, it is important to note that printer's bult-in system of photodiode signal acquisition performs some data correction and noise reduction measures automatically. Further noise reduction and outlier data removal (such as outlier minimum and maximum light intensity data points) were implemented during the signal data preprocessing and the data clustering performed by the self-organizing maps (SOM) algorithm.

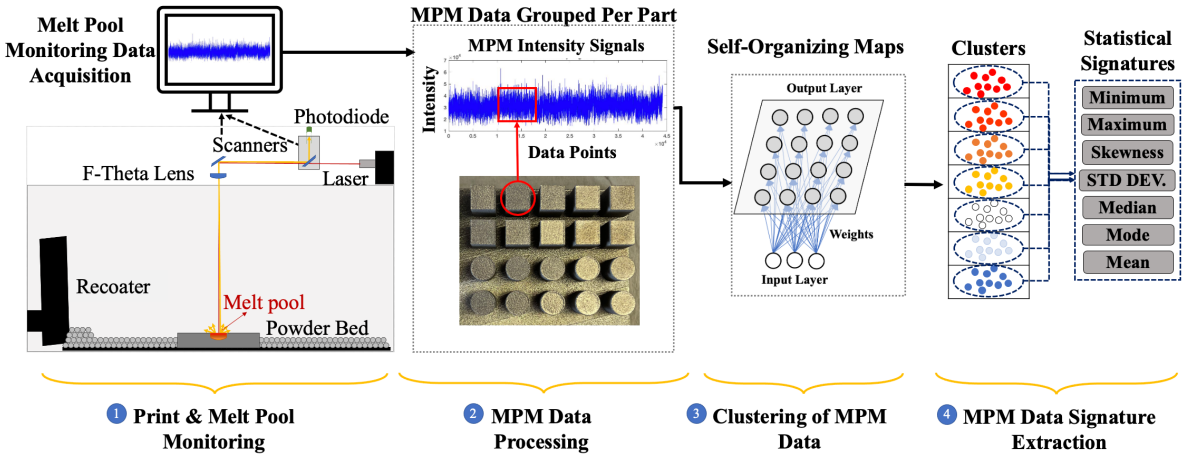


Figure 4-2: Light intensity collection using an on-axis photodiode, processing, and clustering using SOM.

4.2.3 Machine Learning Algorithms

4.2.3.1 Self-Organizing Maps

Each part's MPM data were then clustered into seven groups (other cluster numbers were tested and results are shown in the following sections) based on the intensity level using the unsupervised SOM algorithm. As a result of changes in heat conduction, heat accumulation, and emission rate across different regions of a part, this technique was adopted to extract process signatures from multiple classes/clusters of light intensity data of the parts rather than extracting one general measure (e.g., one average intensity signature for the entire part) per each part. It also benefits the ML algorithms' performance due to the larger feature set.

The SOM algorithm offers several advantages over other clustering methods, such as k-means clustering. The primary reason for its application in this study is its robustness to noise. MPM datasets, i.e., the light intensity signals, are highly susceptible to noise. The SOM algorithm can effectively discern the true underlying patterns within noisy data. This capability arises from the neural network nature of the algorithm, wherein learning occurs through multiple iterations and continuous improvement. Moreover, a previous study conducted by our team [107] has demonstrated that the SOM algorithm can effectively cluster in situ light intensity signals obtained using the same printer and sensor employed in this study.

As illustrated in Figure 4-2, seven statistical metrics were chosen to be extracted from each cluster of MPM signals (per part) as the printed part's signatures.

As described in a study published by Taherkhani et al., SOM training is based on competition while other neural network algorithms' optimization is based on back-propagating [107]. The SOM nodes with the same dimension as the input data have weight vectors that are randomly initialized. In general, the SOM training process could be summarized in three steps: competition, cooperation, and adaptation. In the competition step, the winning node for each of the input data points is identified based on the lowest distance between the node weight values and the input data points. In the cooperation step, the neighboring nodes to the winner node which have a high similarity with it (low distance) are identified using the neighborhood kernel function. The last step of adaptation consists of the modification of node weights regulated with a learning rate decay parameter. For a detailed description of the method, see [107] and refer to Appendix B. The weight optimization continues until no further changes are seen in the winning nodes and then the final map of the clustered data set can be generated using the final weight values of the nodes.

The processed MPM datasets, including the clustering, signature extraction, and scaling, in addition to categorical input features such as powder size ranges and continuous input features such as parts' density, hardness, and surface roughness, are then fed to three various supervised regression algorithms, FFN, RF, and XGBoost, to predict the process parameters (algorithm outputs/targets), as shown in Figure 4-3.

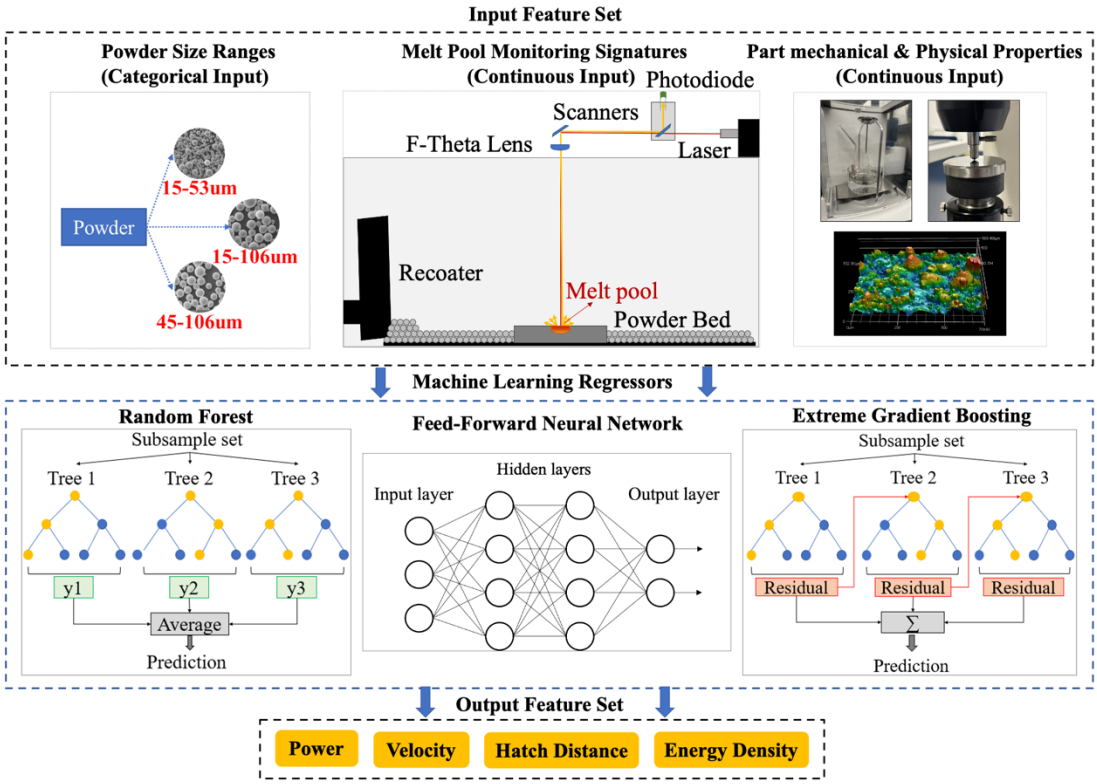


Figure 4-3: Graphical abstract highlighting the study workflow including application of ML on MPM data, part properties, and different powder particle size distributions for process parameters estimation.

Ultimately, the design of the study was planned on providing information regarding the powder size range, the mechanical and roughness properties of the parts, and the statistical representation of the light intensity signals (light intensity signatures) to feed the regressor algorithms. The algorithms were intended to estimate the proper set of process parameters associated to the given powder, part, and melt pool properties.

The MPM light intensity signals, recorded per layer, were initially clustered using SOM. This clustering step aimed to group signals obtained from the entire part into multiple clusters, considering various underlying patterns influenced by factors such as part geometry (e.g., cuboid/cylindrical), process parameters (e.g., power, hatch distance), gas flow, and recoater direction. These factors affect light intensity differently based on the location of the melt pool on the build surface and the height of the printed layer within the part (layer number), making clustering essential for accurate representation of

light intensity signals. SOM incorporates topological characteristics of the signals, generating clusters with similar behavioral patterns in the data.

Subsequently, signals classified into each cluster were characterized using statistical signatures. For each cluster, the average, standard deviation, mode, and median of light intensity were computed. Additionally, the minimum and maximum intensity levels were calculated by averaging the intensity over 10 melt pool points with the lowest and highest intensities, respectively, to mitigate the impact of outlier points potentially affected by noise, particularly in extreme intensity levels. This approach yielded seven statistical signatures from each cluster of photodiode signals recorded in each part, effectively reducing the dimensionality of the MPM data.

These seven light intensity signatures per cluster were combined with categorical indicators of powder size and part properties (density, hardness, and side surface roughness) to serve as input features for the algorithms. In total, 55 input features were utilized in this study, comprising seven signatures from each of the seven clusters of MPM data, one-hot-encoded indicators for three batches of powder sizes, and three features representing part physical properties. The algorithms aimed to predict the four major quality-impacting process parameters: laser power, scan velocity, hatch distance, and energy density, serving as the outputs of the models.

4.2.3.2 Feed-Forward Neural Network

FFN is a supervised ANN that is comprised of the input layer, hidden layers, and the output layer. Each layer is constructed from multiple information processing units denoted as neurons/nodes. Information transfers between the neurons of the subsequent layers by weight parameter. Weights connect the neurons, and their value determines the importance of the information conveyed between each pair of neurons to the accurate calculation of the known target of the network. FFN algorithm training operates using the back-propagation technique which optimizes the user-defined error/loss functions (difference between network calculations and actual outputs) by moving along the steepest slope of error decrease in the loss function space at each training iteration [108]. Training and loss function updating continue until no further improvement is made in the calculations. In this study, the FFN regressor is developed and trained using the tensor flow library and Keras API on the Visual Studio Code interface. Hyper-parameters such as the number of layers, neurons, drop-out rate, and initial learning rate (used in the Adam model optimizer) are tuned in this study by the application of the Keras-tuner tool in Keras application programming interface (API). Adam optimizer, tanh activation function, and the inverse

time decay model optimization technique were chosen for the model based on error and trial attempts operated on the train (0.8 fraction of initially selected train set) and validation data (0.2 fraction of initially selected train set). The model's final structure was derived from the tuner outcome demonstrated in Table 4-2. The activation function of tanh was used in the input layer and hidden layers. Linear activation was employed in the output layer for the regression task.

Table 4-2: FFN-tuned hyper-parameters and the final structure of the algorithm. L1 and L2 denote the first hidden layer and the second hidden layer, respectively

Algorithm	Layers	Nodes		Drop Out		Initial Learning Rate
FFN	2	L1	311	L1	0.2	0.01
		L2	119	L2	-	

4.2.3.3 Random Forest

The RF regressor simultaneously generates several DTs on the various subsets of the dataset. Each tree starts with a parent node that splits into child nodes, which contain subgroups of data points, based on a partitioning condition which is determined at each node of the tree. The dataset splits according to whether each data point satisfies the parent node's partitioning condition or not. The impurity reduction between parent and child nodes of the dataset at each tree is then measured based on the defined metrics (e.g., variation and error). The partitioning of data continues (along the tree length) until no further improvement is made on the impurity reduction in the last partitions of the dataset [109]. An ensemble (average) of random DT learners is then used to implement the final prediction.

4.2.3.4 Extreme Gradient Boosting

XGBoost regressor is also a modified RF algorithm that generates DT learners on random subsets of samples and features but in successive order. Each DT (weak learner) is developed to predict the residual. The prediction of each weak learner is improved compared to the previous weak learner by modification of features' coefficients in the error equation based on the previous learner's results [110]. XGBoost enables a more powerful model regularization by employing L1 and L2 methods and can avoid overfitting more efficiently with the tree-pruning tool. The trained RF and XGBoost algorithms'

hyper-parameters in this study are tuned by implementing the CV and grid search on the train set. The main fine-tuned parameters are listed in Table 4-3.

Table 4-3: RF and XGBoost tuned hyper-parameters

Algorithm	Estimators	Learning Rate	Max Depth	Max Leaf nodes
RF	50	_	9	50
XGBoost	200	0.1	4	_

To summarize, according to the graphical abstract shown in Figure 3, three powder size ranges (15–53 µm, 15–106 µm, and 45–106 µm), the melt pool signatures (mean, mode, median, minimum, maximum, skewness, and standard deviation), and the printed part physical properties (density, surface roughness, and hardness) comprise the input features set of the algorithms. The categorical variables of powder size are transformed into numerical variables through one-hot encoding. Ultimately, three algorithms of RF, FNN, and XGBoost are then trained to predict the outputs which are the process parameters of power, velocity, hatch distance, and energy density of the laser.

4.3 Results and Discussion

4.3.1 Experimental Data Analysis

4.3.1.1 Correlation between Powder Size Distribution and Part Properties

The alteration in part properties (surface roughness, density, and hardness) against the powder batch shift from the finest size range (15–53 µm) to the widest (15–106 µm) and the coarsest range (45–106 µm) are illustrated in Figure 4-4. According to Figure 4-4(a), the surface roughness of cylindrical parts (samples 54–108) exhibits a relative rise in comparison to cubical parts (samples 0–53), particularly accentuated in the coarser batch ranging from 45 to 106 µm. For example, for the cubical sample 2 in this batch, the surface roughness is ~23.01 µm whereas for cylindrical sample 2 in the same batch, which is printed with the same set of process parameters as the mentioned cubical sample, the surface roughness increased to ~29.02 µm. Considering the same set of process parameters are used for the cylindrical and cubical parts, this therefore implies the impact of geometry on the surface roughness of parts. Meanwhile, this trend of variation was not observed in density and hardness characteristics

meaning the change of geometry in this design mainly affected the surface roughness. For all three studied part properties (Figure 4-4(a–c)), it may be concluded that parts in all three powder batches mostly behave the same in response to process parameters variations and have similar fluctuations. As expected, Figure 4-4(a) shows that the parts printed with the coarsest powder batch resulted in the highest roughness while parts printed with the fine powder batch had the lowest roughness. According to Figure 4-4(b) and Figure 4-4(c), the best quality of density and hardness are respectively obtained with parts printed with the fine powder batch as well but much less difference is seen between the powder batches for these characteristics (the least difference in density). Although the set of all influential factors should be addressed in the inspection of LPBF parts quality, it could generally be concluded that the fine powder tends to promote a higher density packing, powder bed uniformity, and energy absorption, thus, less perturbation in the melt pool dynamics. This could lead to improved part properties.

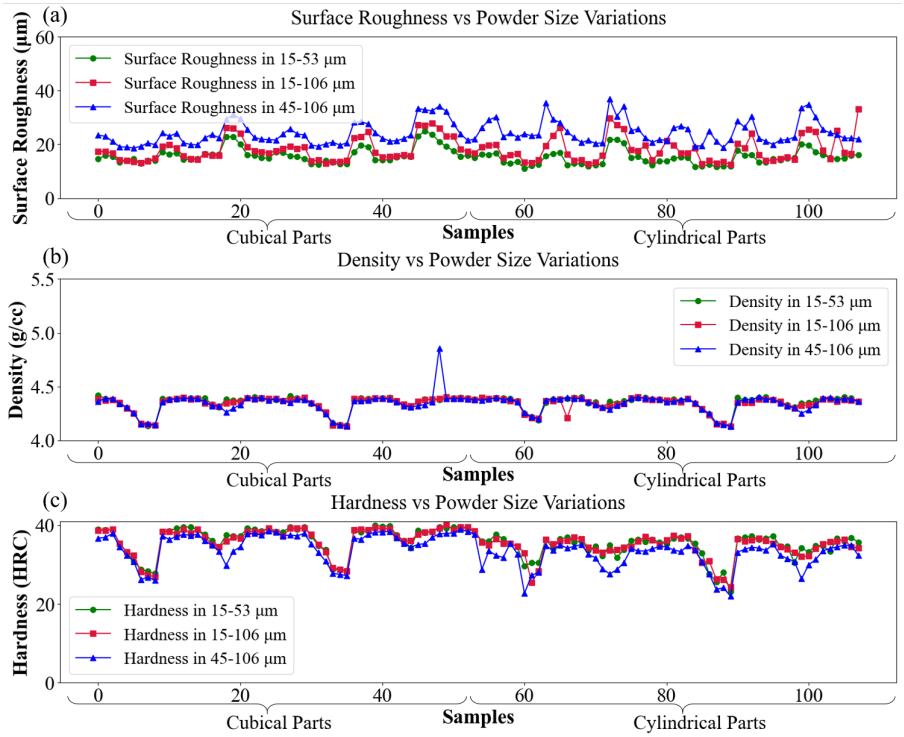


Figure 4-4: Variation of parts physical characteristics in cubical ([0–53] on x-axis) and cylindrical ([54–108] on x-axis) parts with change of powder size. The process parameters per each sample in different powder batches are identical (to observe the impact of powder size change). The part characteristics variations are shown in (a) surface roughness, (b) density, and (c) hardness.

The mentioned features of the fine powder have led to the application of the finer powders for higher quality production in the industry (15–53 μm is a standard size range for LPBF applications). However, the optimization of LPBF for wider and coarser ranges of powder is a vital requirement in LPBF since it is more cost-effective due to the notably lower production cost of wider-range powder. The coarse/wider powder normally inherits better flowability which can significantly alter the powder bed characteristics positively. The process parameter optimization for unconventional/off-size powder sizes is a major gap in the recent state of LPBF research which inspired this study to integrate the standard and off-size powders into the ML-based design of process parameters. The results of this study enable data-driven process parameters arrangement for wider adaptation of off-size Ti64 powder in the LPBF applications.

4.3.1.2 Correlation between Melt Pool Light Intensity, Process Parameters, and Part Properties

Aside from the inspection of relationships between part properties and powder size ranges, the correlation between the MPM data and these factors is also highly influential in the LPBF process optimization. This inspection was also carried out in this study, and it was observed that the MPM data were highly impacted by power and energy density levels rather than powder size range. The correlations between part properties (density and roughness), powder size, and MPM signatures (the average intensity per layer) are demonstrated in 3D plots in Figure 4-5, for parts with the highest and the lowest surface roughness as well as density in all three batches of powder. For clarification on the mapped information in this figure refer to Figure C- 1 in Appendix C section. As seen in Figure 4-5(a) and Figure 4-5(d), regardless of the powder batch used for the prints, higher power causes a higher level of intensity in parts. In cases with equivalent power levels, energy density should be inspected to interpret the intensity level. This means that either hatch distance or velocity would influence the level of light intensity irradiated from the melt pool. In Figure 4-5(b), the highest and lowest density for cubical samples in batch A (pink square and diamond markers) are both attributed to the laser power of 150 W; but the higher energy density in the latter caused the higher level of intensity. Both laser velocity and hatch distance also become relevant even when both power and energy density are equivalent. As such condition could be seen in Figure 4-5(b) between the lowest density cubical samples in A and B (pink and gray diamonds), the smaller hatch distance of the latter caused higher intensity due to heat accumulation within a certain volume of process zone. The same applies to the sample with the lowest surface roughness in batch A (pink circles shown in Figure 4-5(c)) which has the highest light intensity compared to the two samples with the lowest surface roughness in batch B

and C (gray and blue circles) due to the smaller hatch distance. In the same plot, the sample with the highest surface roughness in batch B (gray triangle) shows a lower intensity regardless of having an equivalent power but a higher energy density due to a lower velocity. The significantly lower velocity and lower hatch distance in this sample could generate a keyhole in the melted material causing light scattering within the process zone, which in turn could decrease the intensity which also enables the detection of potential flaws (keyhole in this case) based on the MPM data.

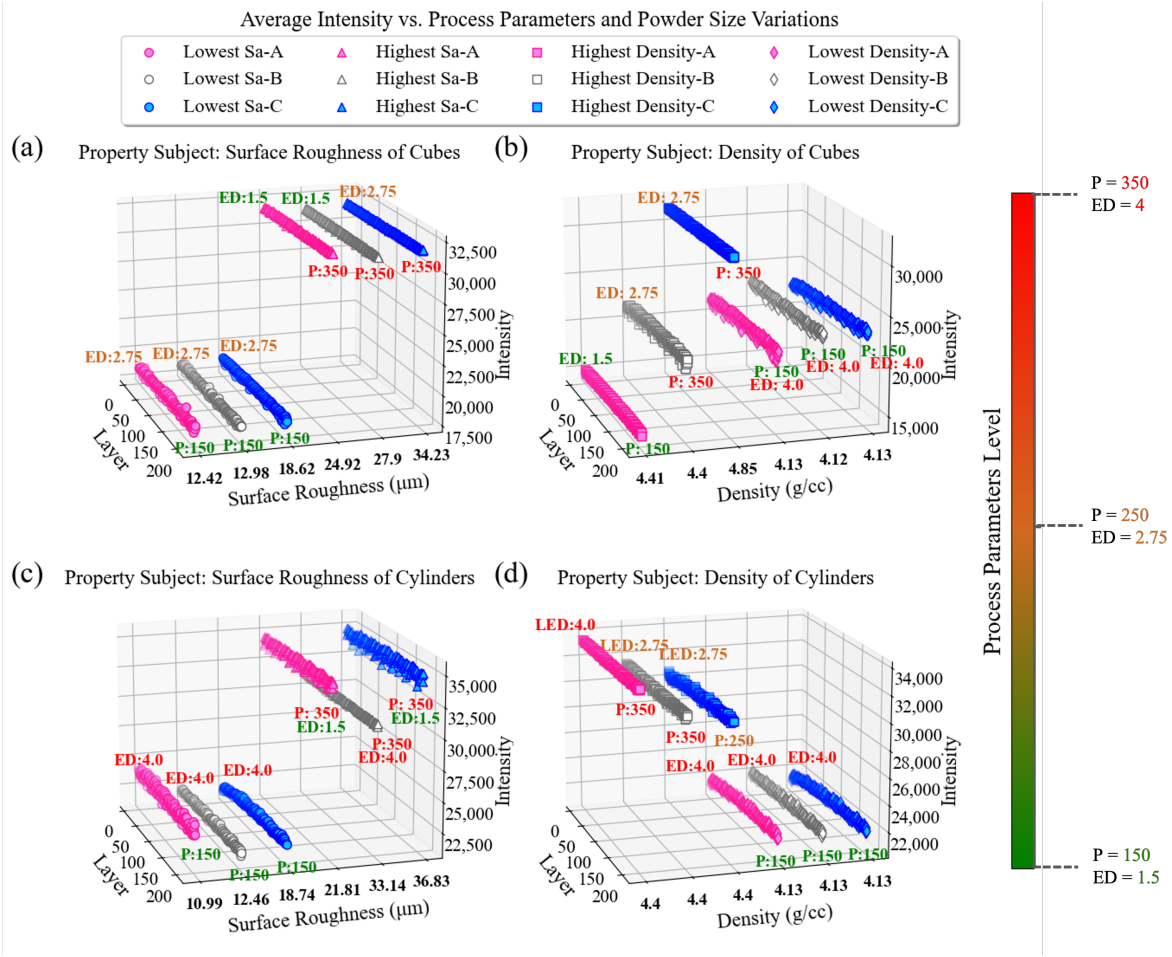


Figure 4-5: The average intensity level per layer is plotted for cubical and cylindrical parts with the highest and lowest surface roughness/density in three powder batches of 15–53 µm (A), 15–106 µm (B), and 45–106 µm (C). The average intensity per layer in parts is plotted for (a) cubical parts with highest and lowest surface roughness, (b) cubical parts with highest and lowest density, (c) cylindrical parts with highest and lowest surface roughness, and (d) cylindrical parts with highest and lowest density.

The light intensity data recorded in each layer, particularly at the edges of the parts, directly links melt pool stability to the resulting roughness, including the side surfaces where roughness measurements were taken. These light intensity signals are closely related to the temperature and thermal distribution on the build surface. Instabilities in the melt pool, caused by thermal variations, are reflected in the light intensity signals. Anomalies such as spatter and balling manifest as perturbations in these signals. Since phenomena like balling and spatter affect the roughness of the parts, the light intensity signal signatures can approximate the resulting roughness in layers and, ultimately, the entire part. Similarly, the melt pool condition in each layer is recorded by the light intensity, with the overall extracted light intensity signatures (per part) statistically representing the light intensity across all layers, correlating to the density and hardness of the final fabricated parts. Light intensity can also indicate potential lack of fusion or keyhole defects based on signal perturbations caused by thermal disturbances during these defects. These defects determine the quality of fusion and ultimately define the part density. Lastly, the cooling rate and thermal history of the melt pool, which significantly impact the microstructure of the parts, can also be captured by the light intensity data. This is due to the light intensity's ability to reflect temperature and thermal conditions within the melt pool, as previously discussed.

These analyses collectively underscore the importance of powder PSD, process parameters, and melt pool characteristics in shaping the physical properties of LPBF-made parts. It should be also noted that these factors are intricately interconnected through complex and nonlinear relationships, which often defy unraveling through simple statistical modeling or experimental studies alone. The volume of informative data from these factors further urges for leveraging advanced data processing tools capable of complex data learning and modeling, such as ML and deep learning algorithms.

4.3.2 Machine Learning Analysis

4.3.2.1 Evaluation Metrics and Techniques

In the first approach of model assessments, the algorithms were trained and evaluated using the repeated CV method. Twelve samples with optimum levels of surface roughness and density were selected from the original dataset to be used as unseen samples. Therefore, the train set on which CV was performed is comprised of the remaining 312 samples. It should be noted that the targets of the analysis contain low variance. Hence, the regression results analysis could not be solely based on the conventional error measuring metrics, such as Mean Squared Error (MSE) or MAE. These metrics mainly normally imply the difference between the actual and predicted values, where they are highly sensitive to scale and

outliers in data. Due to the low variance in the targets, the study assessed regression performance using the MAPE metric. This metric, being scale-independent, was employed for analyzing accuracy results in CV and predictions on unseen samples. Furthermore, due to the low variance in the targets, the deviation span of the errors in the CV analysis was also inspected to attain a reliable insight into the performance of the models. By this approach, an estimation of the uncertainty in predictions could be obtained. The coefficient of determination (R^2 score) is reported in both CV and unseen samples regression results to assess the extent to which the models explain variation in the targets or how effectively a regressor can model the overall target data population (goodness of fit). Additional statistical examinations, such as the Kolmogorov–Smirnov (KS) test and comparison of probability density distributions of prediction and actual targets, are conducted in the unseen samples regressions to provide a comprehensive assessment of regressors' generalization on the unseen samples. Integrating all these assessments helps address the low variance in the targets and provides unbiased insight accordingly.

4.3.2.2 Self-Organizing Maps Clusters Optimization

The number of SOM clusters on which the dataset was generated was also optimized using the CV technique. Initially, seven clusters were chosen to be applied in SOM clustering and generation of the baseline dataset, according to the results of a previous similar study involving a smaller dataset with similar DoE. Therefore, the initial dataset was generated with seven SOM clusters and was subsequently used for regressor algorithms hyper-parameters tuning. For additional assessment, the tuned algorithms were then re-examined with datasets derived from various SOM cluster numbers, using CV with five folds. In CV, the data are divided into multiple folds to make predictions once on each left-out fold while being trained on the rest of the folds. While there are no strict rules regarding the number of folds and it is often dependent on the characteristics of datasets and modeling requirements, five folds of CV were opted for in this study, as this aligns with a common practice where 5 or 10 are typically chosen for the CV implementation [111]. The overall MAPE of prediction obtained by the regressors in the CV results was averaged to choose an appropriate number of clusters among 5, 7, and 10 clusters for this application.

The results of this test are illustrated in Figure 4-6, indicating that the dataset generated using seven clusters of the SOM led to a marginally better average performance in the regression models for all four targets, achieving an average MAPE of approximately 3.7%. Although the difference in performance

between using five and seven clusters was slight, both configurations were deemed valid choices. However, based on previous preliminary investigations with different datasets (cuboids and cylindrical parts), the seven-cluster SOM consistently yielded the best performance. This suggests better generalization of algorithms with this clustering approach across various datasets, leading to the decision to proceed with the seven-cluster SOM dataset.

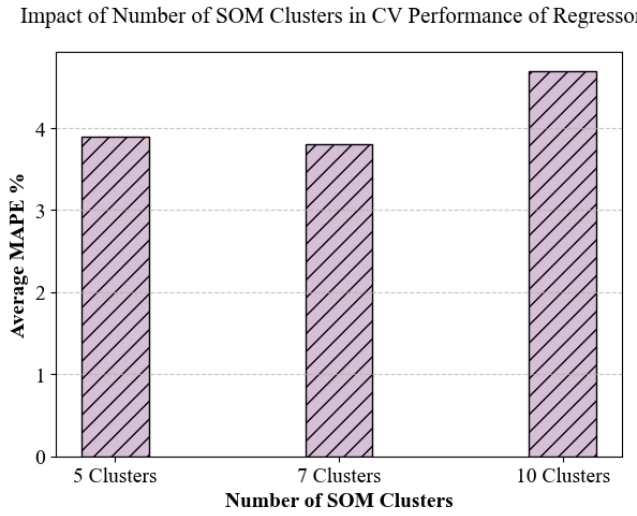


Figure 4-6: Impact of SOM cluster numbers on overall averaged performance of three regressors showed by average MAPE metric obtained in five-fold CV test implemented on train set samples.

4.3.2.3 Cross-Validation on Train Set Analysis

Regression Accuracy Evaluation

Five repeats of the five-fold CV were implemented on the train set meaning that the iterative process of CV is repeated five times to randomly change the set of samples in each fold, resulting in a total of 25 various validation sets. The FFN algorithm was trained here with 50 epochs in each round of CV.

As shown in Figure 4-7(a), the R^2 score obtained by all three models in the regression of validation folds indicates that the models can fit the dataset very well. The obtained R^2 scores show that above 90% of the target variance is explained by the models in all 25 configurations for power, velocity, and energy density targets. However, the results vary for hatch distance, with XGBoost showing the best modeling with a ~0.57 R^2 score, while RF and FFN could not achieve meaningful modeling of the data. This shows that although the XGBoost model exhibits relatively better capability in learning the hatch distance variance, it could still lack a reliable level of fitness due to a low R^2 score. It should also be

considered that low target variability of hatch distance (~9% coefficient of variability in the train set while other targets have above 30%) impacted the R^2 score, regardless of the level of error in prediction. By inspecting the actual values and prediction values obtained in each CV, it was observed that a low level of variability in hatch distance relative to cumulative error significantly decreased the R^2 score in each CV split compared to other targets with high variability. R^2 score mainly reflects how well a model fits the data and how effectively it estimates the variation of the target based on inputs/predictors and low variability of targets could cause bias in its result. On the other hand, error metrics, such as MAPE, indicate the proximity of model predictions to actual data values on a point-by-point basis. For a comprehensive inspection of regression performance, both aspects should then be analyzed. The average MAPE (average of five iterations of random five-fold CV) of models in the estimation of targets is shown in Figure 4-7(b). The MAPE level obtained by the FFN for all four targets exceeds those of the XGBoost and RF regressors. Comparing RF and XGBoost, RF demonstrates marginally more accurate predictions in energy density and power. However, the difference in performance between the two regressors for hatch distance and velocity predictions is minimal. It could also be interpreted from the hatch distance results that, as it was expected by R^2 scores, XGBoost resulted in a lower prediction error. Additionally, although the R^2 score was low, the model regressed with good accuracy (MAPE of ~0.047). This again implies that low target variability could have influenced the low R^2 score.

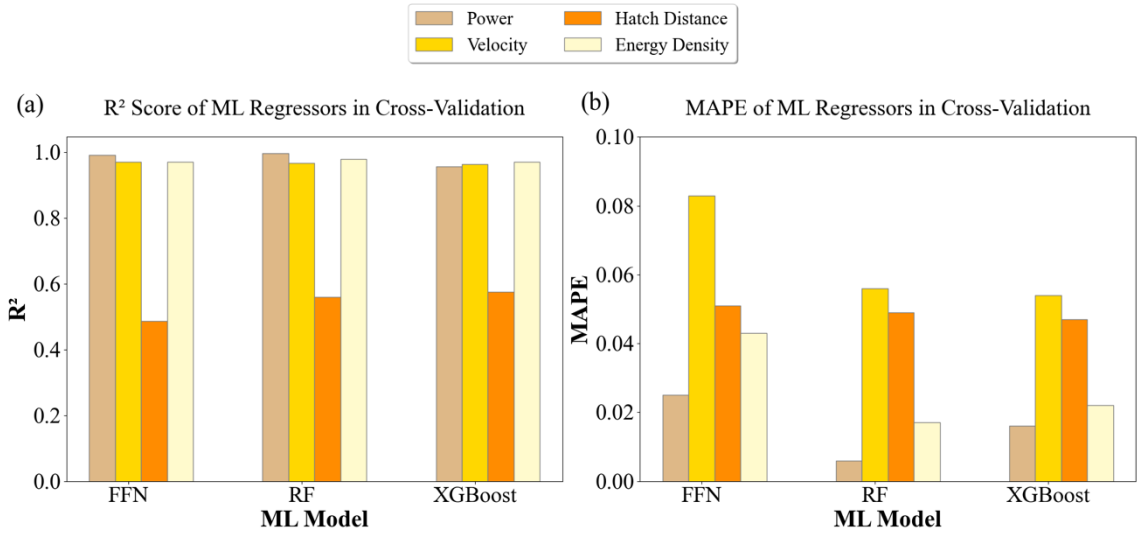


Figure 4-7: Average (a) R^2 score and (b) MAPE of the regressor algorithms in CV analysis implemented on the train set of data.

Statistical Evaluation

The accuracy of the models' prediction is further assessed from a statistical perspective, to identify which model exhibits more reliable regression accuracy. The notched box plot of the yielded average MAPE of models in all 25 configurations of five-fold CV is illustrated in Figure 4-8. Both XGBoost and RF regressors show favorable process parameters regression results in the CV analysis for all four targets. The box plots in Figure 4-8(a–d) show that RF, with small superiority to XGBoost, achieves the highest accuracy for all targets. This marginal difference, compared to the RF model, is also seen in the prediction accuracy results shown in Figure 4-7(a). Additionally, considering the interquartile ranges and whiskers caps, it can also be interpreted that RF exhibits a smaller range of variation in error levels throughout the 25 CV combinations for all four targets, this underscores the reliability of this regressor, suggesting its higher robustness to data perturbations and potential for improved generalization. In contrast, the FFN regressor shows a wider range of errors in the CV analysis for all four targets, indicating its less robustness to noise or any potential data variation. Hence, it is anticipated to result in poorer generalization to new unseen samples. The overlapping notches in the RF and XGBoost boxes illustrated in Figure 4-8(b–d) show the median in these two populations is not significantly different, showing a similar level of prediction accuracy, in contrast to FFN. In summary, it can be concluded that XGBoost and RF are the better regressors, in the CV analysis focused on the training set, in terms of goodness of fit and prediction accuracy.

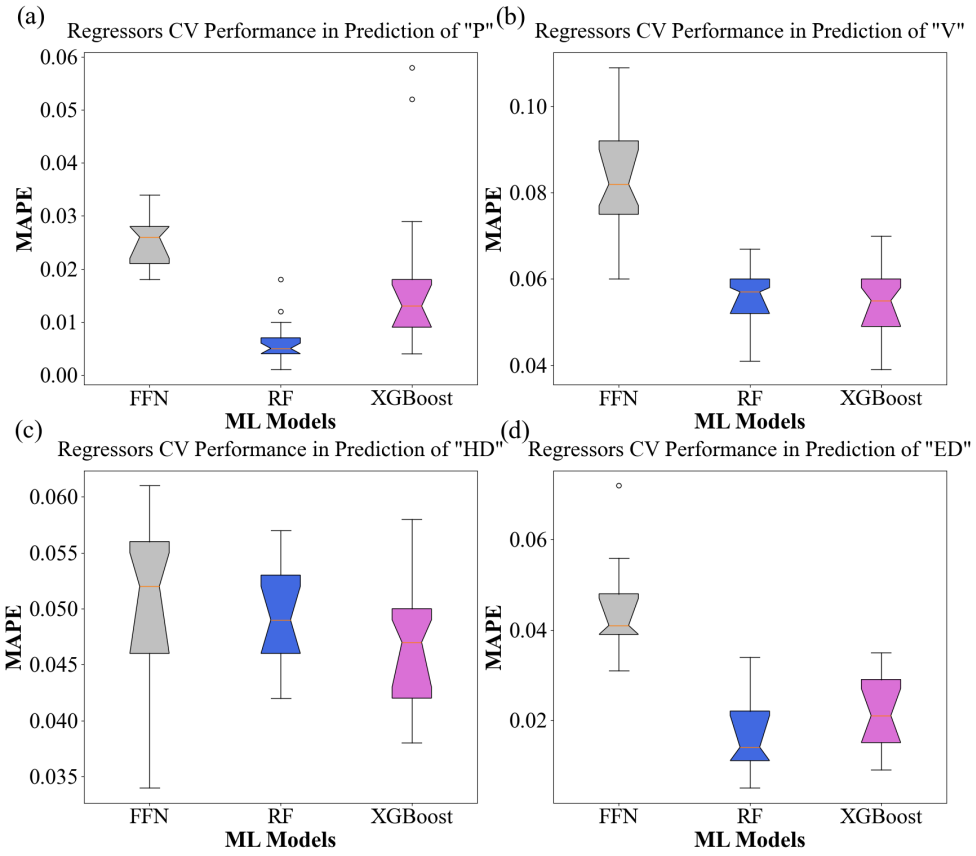


Figure 4-8: Notched box plot of regression errors of three regressor algorithms obtained in CV analysis implemented on the train set. The results are shown for (a) power (P), (b) velocity (V), (c) hatch distance (HD), and (d) energy density (ED).

However, it should be noted that RF and XGBoost achieved a low R^2 score in hatch distance modeling despite exhibiting good point-wise prediction accuracy. This may suggest a potential issue of overfitting to the training samples. Accordingly, further inspection into model generalization was implemented relying on algorithms evaluation on unseen samples.

4.3.2.4 Train Set Size Impact

While CV allows for assessing the model performance based on variations in sample set combination, the number of training samples remains fixed in each fold. However, the size of the training set is one of the most influential factors in the performance of regressors, necessitating an examination of the performance of the regressors across different training set sizes. Limited training data sizes can result in overfitting and limited generalization performance. Increasing the size of the training set typically

enhances the accuracy of regression models, especially those with strong generalization abilities. The regressors' overall prediction error (average error of all targets) was assessed in various fractions of a train set in the [60%, 70%, 80%, 90%] range taken from the set of 312 samples. Therefore, in each train size fraction, models predicted the samples in the remained fraction of 312 samples. As shown in Figure 4-9, it is observed that an increase in train set samples mostly led to higher accuracy (except for a negligible increase from 70% to 80% in RF and XGBoost models). This implies that there is no significant overfitting/sensitivity to data which confirms the admissible level of generalization of the models.



Figure 4-9: Impact of train set size on the overall performance of three regressors in the estimation of process parameters. The overall error is shown by the $\log_{10}$(MAPE) metric.

Overall, FNN consistently resulted in higher errors in all target estimations, which agrees with the previous analyses. However, it should be considered that none of these error levels are considered major in LPBF applications and do not disqualify any of the models. In conclusion, in a comparative study among the three regressors, FFN seems to have lower accuracy and generalization. RF and XGBoost, in most cases, demonstrated similar levels of performance and better generalization compared to FFN.

4.3.2.5 Regression of Unseen Samples Analysis

Regression Accuracy Evaluation

The final evaluation of the models in terms of generalization performance involves regression on unseen samples that were not included in previous analyses. The models were subsequently reset and retrained on the entire training set (i.e., 312 samples) to make predictions on a pre-selected set of 12 unseen samples. It should be noted that since these 12 samples were the optimum samples of the DoE

in terms of surface roughness and density, they count as extreme points compared to the utilized training set. Consequently, it enables the analysis to evaluate the robustness of the models to outliers/extreme points too.

In addition, an epoch number examination was carried out separately to identify the optimal epoch number that balances the overfitting and bias of the FFN model. In this test, 20% (from 312 samples) of the training set was used to evaluate the FFN performance in various epoch numbers. Figure 4-10 illustrates a significant decrease in error within the initial 20 epochs, followed by a highly gradual decline in error from 20 to 100 epochs. The epoch number of 50 was chosen since it promotes a good level of accuracy in estimations. The model's training had also reached stability in the learning process at this epoch number.

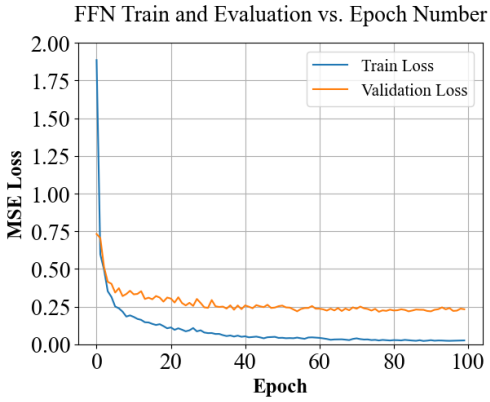


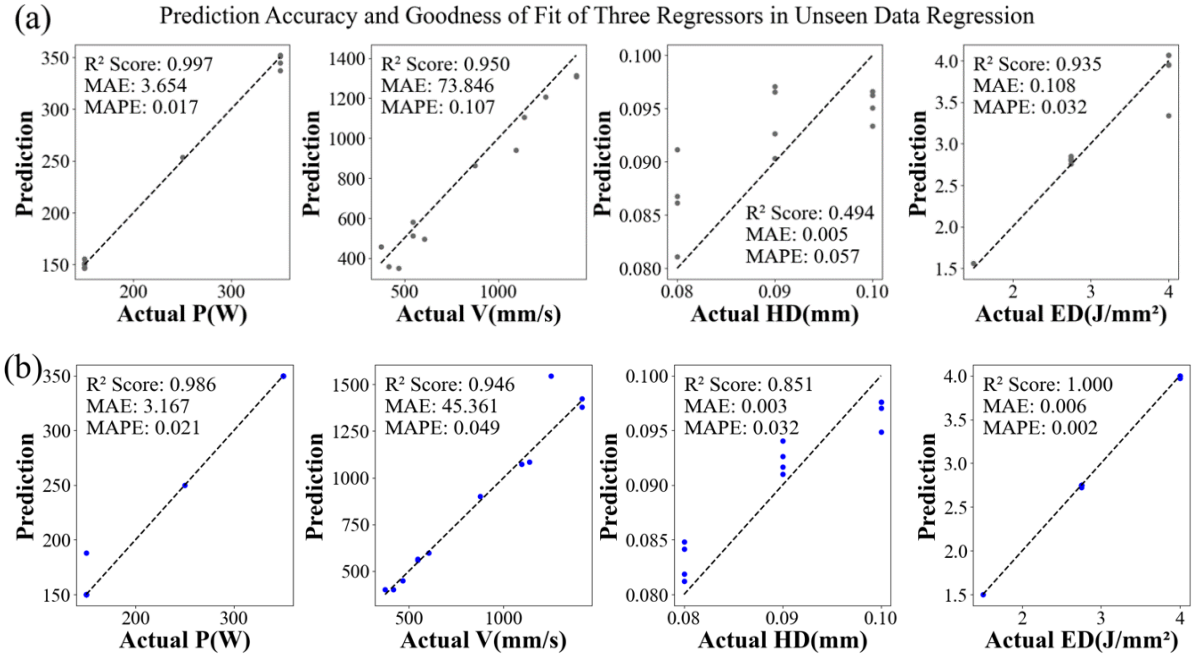
Figure 4-10: The impact of epoch numbers on FFN performance was tested and shown by the MSE metric. The algorithm was tested on 80% of the initial train set as the new train set and 20% of the initial train set as the validation set.

The RF algorithms and XGBoost regressors used in the prediction of the unseen samples are the same tuned models that were used in the previous analysis. In this step, all three algorithms demonstrated good prediction accuracy for the majority of targets. Among four targets, the lowest prediction accuracy was seen in the hatch distance estimations for all three candidate models. This was expected according to the performance of the models in the previously reported analysis. The total R^2 score achieved by FFN in predicting unseen samples was ~0.84. The performance of FFN on unseen samples for each four targets (R^2 and MAPE) is shown in Figure 4-11(a). FFN shows very accurate generalization for power, velocity, and energy density estimations, yet achieved a lower R^2 score in the hatch distance

prediction. However, the observed performance of FFN in the hatch distance estimation for unseen samples is consistent with the results obtained in the CV analysis (training samples).

A great total R^2 score of 0.94 was achieved by the RF regressor in predicting unseen samples. As demonstrated in Figure 4-11(b), RF achieved above 90% accuracy in power, velocity, and energy density estimations, as well as FFN. However, RF showed more reliable results than FFN in hatch distance estimations, considering both R^2 and MAPE measures. This suggests an overall good generalization and robustness in the RF regressor.

XGBoost regressor showed results slightly less accurate than RF and marginally better than FFN on the unseen samples set, with a total R^2 of 0.9. The R^2 and MAPE obtained by XGBoost regressor in unseen sample estimation are shown in Figure 4-11(c). XGBoost yielded results almost as accurate as RF across all targets, except in hatch distance estimations where an R^2 score of ~0.67 was obtained which explains that both FFN and XGBoost could not provide accurate data fitting for this target. Meanwhile, RF resulted in robust hatch distance modeling (R^2 score~0.8) and accurate predictions (lowest MAPE).



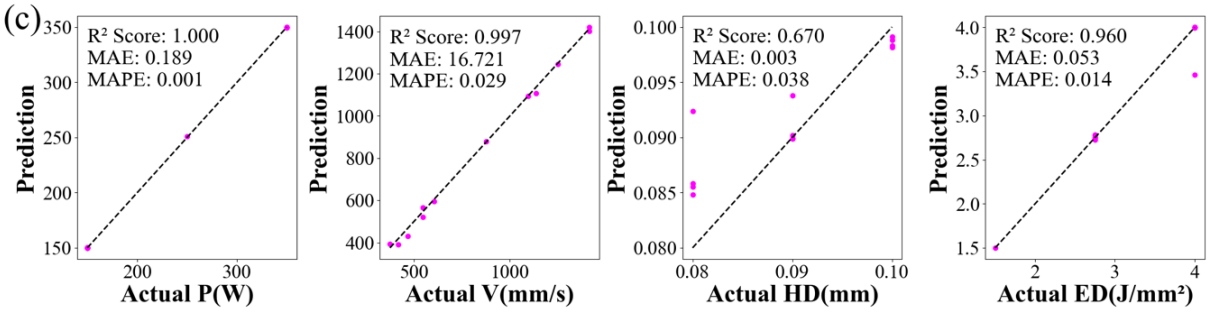


Figure 4-11: Prediction vs. actual target plots obtained from the unseen samples' regression in the four targets. Prediction accuracy of regressions is also shown by MAE and MAPE for (a) FFN regressor (b) RF regressor, and (c) XGBoost regressor. Power, velocity, hatch distance, and energy density are denoted by (P), (V), (HD), and (ED), respectively.

Statistical Evaluation

As previously mentioned, due to the low variance of targets in the dataset, regression performance needs to be assessed in various aspects using different techniques. The R^2 score, error variability, and prediction accuracy were discussed and compared for the three trained models. However, to further ensure the generalization of the models, the results obtained by the three regressors in estimating the unseen samples were also examined from a statistical perspective. This was implemented to examine the predictions in terms of distribution. The KS test, a non-parametric technique that does not assume any distribution for the data, was applied to test the regression results. The test results indicated that the predicted values by all three models are from the same population distribution as the targets, further confirming the credibility of the models' performance.

In addition, the Probability Density Function (PDF) was used to visualize the probability density distribution of the predicted population and the target population in the unseen samples tested by the KS tests. Based on the PDF in Figure 4-12(a), it can be concluded that FFN yields less overlap between predictions and actual target distribution in velocity, hatch distance, and energy density targets. By comparing the PDF plots in Figure 4-12(b) and Figure 4-12(c), obtained for each of the targets, it can be concluded that RF and XGBoost generally exhibit better data learning, but RF achieves slightly more accurate data learning overall. This aligns with the previous analysis and the fact that RF generates a density distribution closer to the actual targets.

The plot of hatch distance density distributions indicates that all three models frequently estimated a hatch distance of 0.09 (instead of 0.08 and 0.1), suggesting that they lacked accurate data variation learning concerning hatch distance changes at different extents. This deviation, however, is very marginal when it comes to the optimization of LPBF. Notably, the hatch distance variations in the DoE, ranging from 0.08 to 0.1, do not have a major impact on the quality of the parts, unlike the other three target changes. In addition, the prediction MAE of models in hatch distance fell within an acceptable small range [0.003–0.005]. Therefore, the hatch distance estimation accuracy obtained by the models could still be considered an admissible result for such a DoE in terms of the quality and functionality of the LPBF-made product.

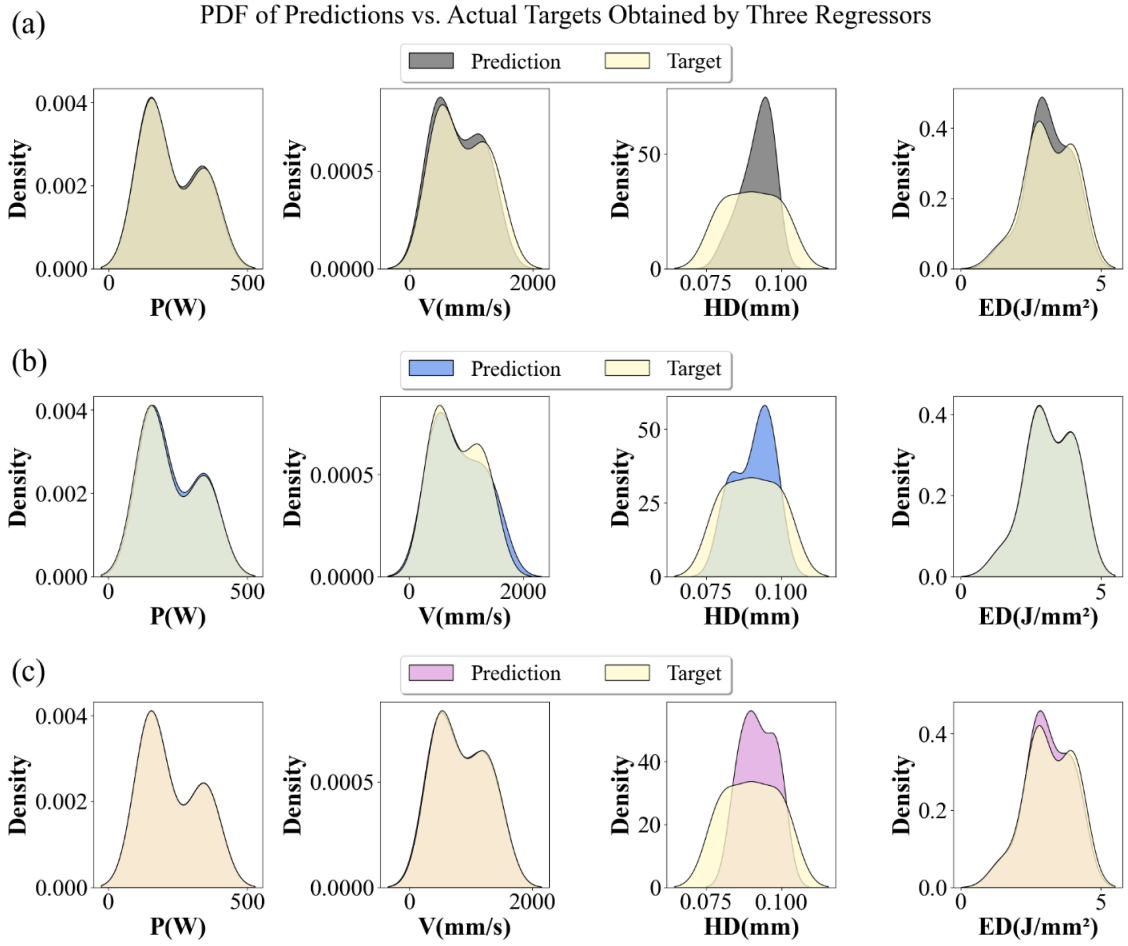


Figure 4-12: Probability density distribution of predicted values and actual target values in unseen samples regression resulted by (a) FFN (b) RF, and (c) XGBoost regressor. Power, velocity, hatch distance, and energy density are denoted by (P), (V), (HD), and (ED), respectively.

4.4 Summary

In this chapter, three distinct regression models, namely Feed-forward Neural Network (FFN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were trained and optimized to forecast process parameters, including power, velocity, hatch distance, and energy density, utilizing light intensity signatures and physical properties of cubical and cylindrical shape samples, made by laser powder bed fusion (LPBF), across standard and off-size Ti6Al4V (Ti64) powder size distributions. The efficacy of these regression models was evaluated through metrics such as the coefficient of determination (R^2 score), mean absolute percentage error (MAPE), and a comprehensive analysis of error distribution within cross-validation (CV) assessments. Additionally, the performance of these algorithms was assessed through the prediction accuracy observed in evaluations of unseen samples. Overall, all three algorithms demonstrated satisfactory performance in estimating process parameters, as evidenced by the magnitude of prediction errors. Among all regressors, RF showed the most reliable results, showing promising generalization in unseen sample evaluation with an overall MAPE error of 2.5% and R^2 score of 0.94 in the prediction of all four process parameters. Following RF, XGBoost emerged as the second-best performer in estimating process parameters for unseen samples, albeit with marginally lower goodness of fit. XGBoost achieved an overall MAPE of 2% and an R^2 score of 0.9. Despite exhibiting less precise model fitting and generalization, the FFN algorithm still produced acceptable results, achieving an R^2 score of 0.84 and a MAPE of 5% in evaluations of unseen samples.

Chapter 5: Complementary Validation of Optimal Powder Rheology Regressors on Unseen Off-size Ti6Al4V Powder Samples

5.1 Abstract

In this chapter, the machine learning (ML) regressors and techniques identified as optimal for rheology prediction in Chapter 3 are further evaluated using three new off-size powders from a coarser particle size range (starting at 45 µm). These evaluations are conducted from two distinct perspectives. First, the generalization capabilities of the regressors and equations selected in Chapter 3 are tested on these new, unseen powder samples. Second, the impact of expanding the training dataset by incorporating these new samples is assessed by performing regression on the same author-defined test set from Chapter 3.

5.2 Methodology and experimental setups

5.2.1 Powder Rheology Estimation

To further evaluate the algorithms developed and examined in this thesis, the regressors were tested on new, distinct datasets. The algorithms that were trained and tested for powder rheology estimation in Chapter 3 were ultimately applied again, using three new off-size PSD ranges: 50-100 µm, 63-112 µm, and 50-63 µm. This evaluation was conducted using two approaches:

Addition of New Samples to the Original Training Set: This approach investigated whether including a limited number of new samples would enhance the performance of the developed regressors. Additionally, the regressors were re-tuned (hyper-parameters) on the new training set to determine if re-tuning could yield more accurate results. Given the minimal nature of the datasets, re-tuning the model is often essential for achieving optimal performance. The results obtained from both the developed and re-tuned models are presented and compared in the following sections.

Generalization of Developed Models: This approach aimed to assess the generalization capability of the developed models on new, unseen powder size ranges of Ti64. The analysis was performed in two scenarios: first, by estimating all three new samples (denoted by "developed model – full test set"), and second, by estimating only the two of the newly added samples while excluding the outlier sample from the analysis (denoted by "developed model – outlier excluded"). The sample that exhibited significantly different rheological measures compared to the other samples, was classified as an outlier. Since outlier

samples can reduce estimation accuracy in various rheological targets, particularly when the dataset is limited, the results for both scenarios—considering and excluding outliers—are reported and compared in the subsequent sections to account for this phenomenon.

5.3 Results and Discussion

5.3.1 Dynamic Rheology

Estimation of BFE and SE targets determining the dynamic rheology of the powders were previously explored extensively in former chapters which ended up in the SVR algorithm, relying on the original author-defined set of inputs, achieving the best results in BFE and SE targets, incorporating the SMOTE augmented dataset and the original dataset, respectively. Therefore, the same developed SVR algorithms were again employed to further evaluate the obtained results using the new powder size ranges.

Basic Flow Energy

The previously developed and tested SVR algorithm (using the author-defined set of inputs) is examined first by adding the new three powder samples to the 7 original training samples and generating a new SMOTE augmented data set from it. This was due to the conclusion of the former analysis that showed the best performance for BFE prediction was attained by SVR trained on an augmented dataset. This new augmented data consisted of 50 samples (7 original train samples + 3 new samples + 40 synthetic samples obtained by 4 rounds of SMOTE on these 10 samples). This set is then used to train the SVR once with the same optimum hyper-parameters achieved in Chapter 3 (Table 3-5), and once re-tuned on the new augmented dataset. Both models are then evaluated on the same original author-defined set of unseen/test samples. The results here would indicate whether the expansion of the dataset, although very limited, would improve the powder rheology regression.

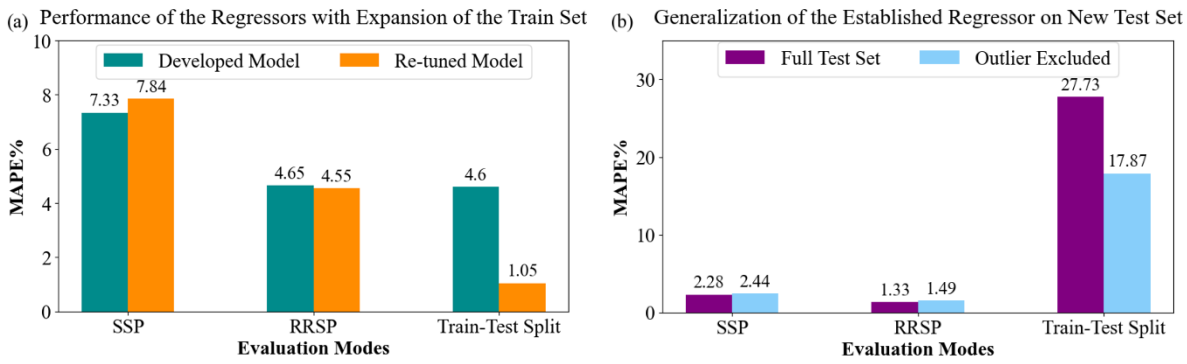


Figure 5-1: Final evaluation of the SVR algorithm in BFE regression using new off-size powder samples. The performance of (a) the developed SVR model compared to the re-tuned SVR model on the original test set, and (b) the developed model with/without the outlier sample (50-63 µm), on the new test set, in three evaluation modes of regression.

As shown in Figure 5-1(a), the established/developed model demonstrated highly satisfactory results in BFE regression, achieving a MAPE of less than 10% across all evaluation modes. Notably, re-tuning the model with the new training set enabled even lower MAPE levels in RRSP and train-test split evaluation modes. When compared to the previously best-achieved result of 3.13% MAPE in the train-test split mode, the improved MAPE of 1.05% indicates that incorporating new limited samples and re-tuning the model enhances the accuracy of BFE regressions on the same unseen samples.

These results suggest that the established model consistently provides highly accurate BFE regression outcomes. However, due to the inherent limitations of the dataset, re-tuning the model significantly aids in achieving even more precise fitting and predictions.

As illustrated in Figure 5-1(b), the second evaluation approach (generalization examination) shows that the developed model yields very good results in SSP and RRSP modes, regardless of the inclusion or exclusion of the outlier sample (50-63 µm). However, in the train-test split mode, the model exhibits reduced accuracy, though excluding the outlier sample decreases the error by approximately 10%.

This suggests that the model can accurately estimate BFE when incorporates a subset of these new samples in random combinations from the entire dataset (RRSP) and single sample predictions (SSP). However, when the model is required to generalize to these new, unseen samples without the inclusion of similar samples and relying solely on the original training samples and their synthetic versions, the

regression accuracy declines. Nonetheless, the generalization to the 50-100 µm and 63-112 µm samples remains acceptable for powder characterization purposes in the LPBF applications.

Specific Flow Energy

The established/developed SVR algorithm (results reported in Chapter 3), which demonstrated optimal performance in SE regression when trained on an author-defined set of input features and the original dataset, was further evaluated by incorporating the new powder samples into the analysis. This evaluation aimed to assess the influence of data expansion and the generalization to new, unseen samples for the SE target.

In the first evaluation, the developed and re-tuned SVR models were trained on a dataset comprising the original seven training samples plus the three new samples. These models were then tested on the same set of test samples (train-test split mode) as well as in other evaluation modes (RRSP and SSP).

In the second evaluation (generalization), the performance of the developed model, trained on the original seven-sample training set, was assessed for its regression accuracy on the new samples. This assessment compared scenarios with and without the inclusion of the outlier sample (50-63 µm).

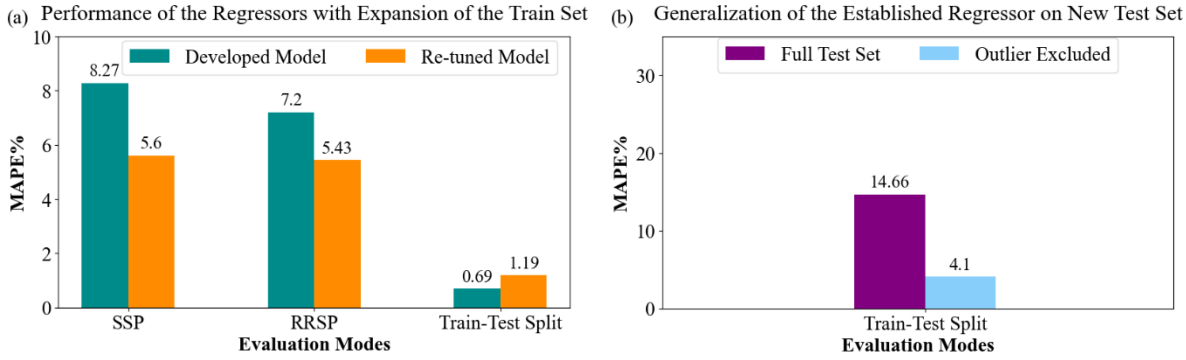


Figure 5-2: Final evaluation of the SVR algorithm in SE regression using new off-size powder samples. The performance of (a) the developed SVR model compared to the re-tuned SVR model on the original test set, and (b) the developed model with/without the outlier sample (50-63µm), on the new train set, in three evaluation modes of regression.

As shown in Figure 5-2(a), similar to the results obtained for BFE, the established SVR algorithm maintained satisfactory regression accuracy, with a MAPE of below 9% across all evaluation modes. However, the re-tuned model provided even higher accuracy in multi-sample regression tests. This

indicates that algorithm tuning after changing the training dataset can enhance the performance, although the unchanged algorithm also yielded good results.

The achieved MAPE of below 1% in train-test split mode suggests that adding three new samples to the training set enabled even more accurate estimation for the same set of test samples than when estimations were made relying only on the original train samples. The accuracy of the regression in random subsets (RRSP) or single-sample (SSP) test sets that included these newly added samples, was also satisfactory since all yielded regression errors below 10%. It should be considered that the results achieved by both the developed and re-tuned models are acceptable for LPBF applications, although the re-tuned model obtained relatively more accurate results in RRSP and SSP modes.

Figure 5-2(b) indicates that when the generalization of the developed SVR model, trained only on the original seven samples, is evaluated on the new powder samples (excluding the outlier sample), an acceptable level of accuracy is achieved for SE regression, with a MAPE of approximately 4%. Additionally, the obtained SE estimation for the outlier sample can still provide an admissible rough approximation of its rheology, as the total regression error, when the outlier is included in the analysis, remains within an acceptable range for the LPBF applications (MAPE ~15%).

5.3.2 Shear Rheology

Based on previously reported shear rheology estimations, it was concluded to use GP-SR-derived equations for UYS3 and UYS9 target estimation and apply the SVR regressor on the SMOTE-augmented data to predict UYS6. The same methodologies were adopted in the final evaluation incorporating new samples. The GP-SR equations for UYS3 and UYS9 were utilized to estimate the new samples, assessing the generalizability of the derived equations.

The previously optimized SVR regressor for UYS6 was applied to a SMOTE-augmented dataset generated from the original seven samples and three new powder samples. The performance of this model, along with the re-tuned model in RRSP mode, is reported in the following section to provide a comparative assessment with previous results.

Furthermore, the generalization of the model, utilizing the same developed algorithm and the SMOTE-augmented data from the original training samples, was examined on the three new samples. Special consideration was given to the influence of outlier samples on the results.

Unconfined Yield Strength at 3kPa

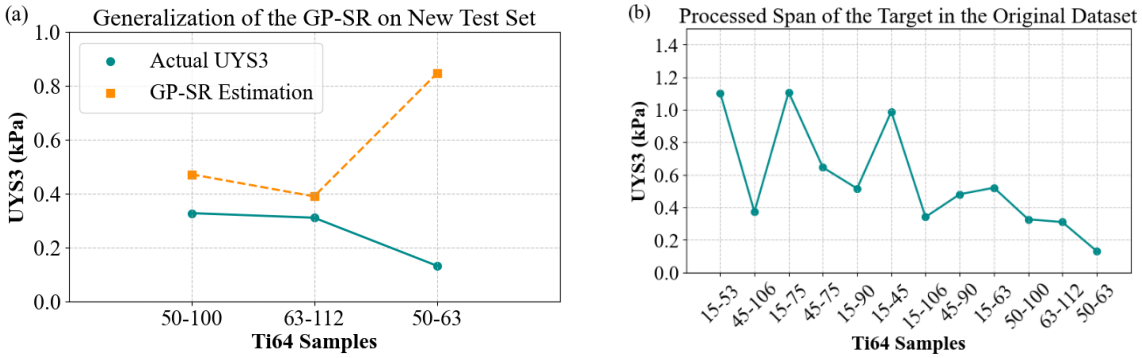


Figure 5-3: (a) Estimated UYS3 values by the developed GP-SR equation in three new powder samples and (b) comparison of UYS3 measures of the new samples relative to the trained range.

The estimated values for both the 50-100 µm and 63-112 µm samples are acceptable, as shown in Figure 5-3(a). However, as illustrated in Figure 5-3(b), the 50-63 µm sample is an outlier in this dataset, resulting in inaccurate estimations. This suggests that while the developed equations can provide reasonable initial estimates of the shear rheology, they lack precision for new samples whose UYS3 measurements fall outside the trained range, necessitating extrapolation.

Unconfined Yield Strength at 6kPa

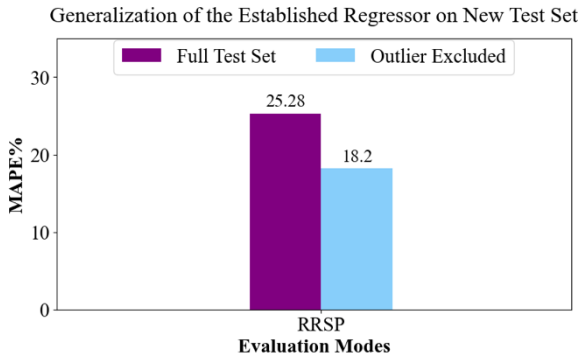


Figure 5-4: The performance of the developed model with/without the outlier sample (50-63 µm), on the new test set, in three evaluation modes of regression.

When the re-tuned SVR algorithm relies solely on the original training samples and their corresponding synthetic samples, as illustrated in Figure 5-4, it achieves an acceptable level of estimation accuracy in RRSP mode, considering the most recent results. For comparison, at the most optimum level achieved in the previously reported analysis (Chapter 3), the MAPE level was reduced to approximately 27.9% but the model here yielded a MAPE of ~18.2% when the outlier sample was excluded. The algorithm also yielded an accuracy of about the same level formerly achieved (on the original dataset) when no outlier exclusion was performed. These findings suggest the SVR algorithm's generalization performance for the UYS6 target, aligns with the previously obtained results but the algorithm was not able to perform well in cases of significant extrapolation (outlier) and data cleaning with outlier removal aids the generalization of the model. This is expected in scenarios with restricted training datasets. Thus, to enable the model for acceptable performance in further broad ranges of PSD, more variability in the original dataset is necessary.

Unconfined Yield Strength at 9kPa

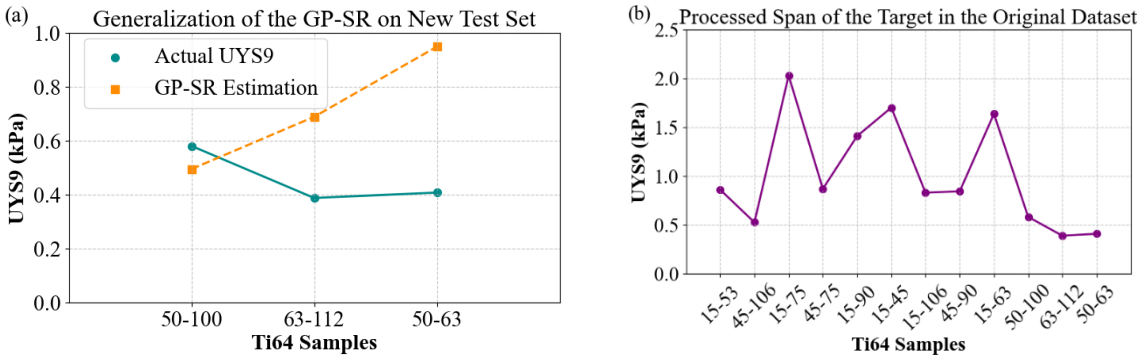


Figure 5-5: (a) Estimated UYS3 values by the developed GP-SR equation in three new powder samples and (b) comparison of UYS6 measures of the new samples relative to the trained range.

Similar to the results discussed for UYS3, as it can be seen in Figure 5-5(a) and Figure 5-5(b), the developed GP-SR equation for UYS9 performs well in estimating the new sample that lies within the trained span of UYS9 data (50-100 µm). However, it estimates new samples that fall outside this range with poor accuracy.

In general, it can be concluded that the GP-SR is highly sensitive to the training set when generating fitting equations. In cases involving new samples that significantly deviate from the training data, the GP-SR is unable to produce acceptable results.

5.3.3 Hall Flowability

The established SVR model in Chapter 3 for hall flowability regression, trained on the SMOTE-augmented dataset with an optimal set of five input features, was utilized here to further evaluate the algorithms. To assess the model's performance, the new samples were used, examining the influence of data expansion and generalization. Similar to the results reported for BFE and SE, both the original and re-tuned algorithms were retrained on the SMOTE-augmented data generated from the seven original training samples plus the three new samples to predict hall flowability in the same set of test samples.

For the generalization analysis, unlike the exclusion of the outlier sample in the dynamic rheology section, the results included all three new samples as the test set. This approach was taken because, from a hall flowability data perspective, the new samples fall within the range of the previously processed data (original data), and none of them is considered an outlier requiring extrapolation for data fitting.

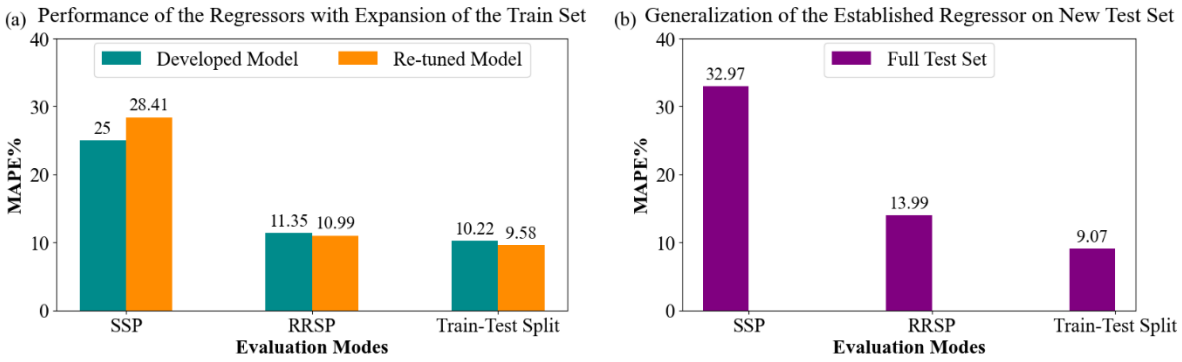


Figure 5-6: Final evaluation of the SVR algorithm in hall flowability regression using new off-size powder samples. The performance of (a) the developed SVR model compared to the re-tuned SVR model on the same test set, and (b) the developed model on the new test set, in three evaluation modes of regression.

According to Figure 5-6(a), the SVR algorithm did not achieve better performance compared to the previous trial, which relied on the original training set to predict the author-defined test set. This was consistent for both the developed and re-tuned models trained on the SMOTE-augmented data generated from the original training samples plus the three new samples.

A similar trend was observed in both RRSP and SSP modes, where expanding the training set did not lead to a significant improvement in accuracy. This decrease in accuracy may be attributed to a potential bias introduced during the regressor's training process. The inclusion of new samples and their corresponding synthetic counterparts from the 50 μm starting range class of powders (coarse powders) might have introduced imbalances in the dataset. As a result, accuracy declined relative to the previously reported results.

Despite this, the observed results still offer a reliable method for estimating powder flowability in LPBF, given their acceptable level of accuracy. However, it is still evident that data limitations continue to impact the analysis of hall flowability targets.

Regarding the generalization of the developed model with unchanged hyper-parameters and training set on the three new samples, Figure 5-6(b) shows that acceptable results were achieved in train-test split mode, with a MAPE of approximately 9%. The highest error was observed in SSP mode, but this was not attributable to the new samples. Instead, the error stemmed from the samples in the 15 μm starting range class of powders (fine and broad PSD powders). Similar to the potential bias previously discussed (due to the inclusion of more coarse powders), this discrepancy may have again led to less accurate hall flowability estimation for fine and broad powder samples.

These observations underscore that further improvements in hall flowability estimation accuracy are contingent upon the availability of a more inclusive and comprehensive dataset encompassing fine, broad, and coarse powder samples.

5.4 Summary

The regression techniques identified as optimal for predicting the rheological and flowability properties of Ti6Al4V powder in standard and off-size particle size distribution (PSD) ranges were further evaluated using a new set of coarse powder samples across three different PSDs. With the exception of the unconfined yield strength at 6kPa consolidation (UYS6) target, where results were comparable to those in Chapter 3 with no significant change in regression accuracy, the techniques demonstrated good to acceptable generalization on the new powder samples. The performance improvement observed in some targets when the entire new set or a subset was included in training highlights the importance of expanding the dataset to cover a broader range of powder PSDs for further performance enhancement. Additionally, the sensitivity of the regression techniques to outlier samples across all targets suggests limitations in their accuracy when extrapolating beyond the existing data.

Chapter 6: Application of the Optimal Melt Pool Monitoring-enhanced Process Parameters Tailoring Algorithms on Complex Artifacts

6.1 Abstract

In this chapter, a generalization assessment for the second objective of the study, that focused on developing predictive algorithms to predict the process parameters based on light intensity signatures, part properties, and powder feedstock size range, was performed. The random forest (RF) and extreme gradient boosting (XGBoost) algorithms, previously selected as the optimal regressors for predicting process parameters using melt pool monitoring data and mechanical properties, are further tested. These tests evaluate their ability to generalize to complex artifacts fabricated with three different Ti6Al4V (Ti64) particle size distribution (PSD) ranges, without altering the algorithms' tuning or input feature sets to account for geometrical features.

6.2 Methodology and experimental setups

Three regressors (RF, XGBoost, and FFN) were trained to predict the process parameters of the LPBF (power, velocity, hatch distance, and energy density). These algorithms were trained and evaluated using a dataset that included photodiode-based light intensity data, physical and mechanical properties of the printed parts (density, hardness, and surface roughness), and the process parameters. A total of 324 cubical and cylindrical parts were printed using three different size ranges of Ti64 powder feedstock. The investigation aimed to input the statistical signatures of the in-situ light intensity signals, extracted from the raw MPM data, along with the part properties, into the regressor algorithms. The goal was for these algorithms to accurately predict the optimum process parameters applied during the printing of the parts.

According to the evaluations conducted and reported in the previous chapters, RF and XGBoost, with minimal difference to RF, achieved the best performance. This performance was consistent in both the cross-validation analysis of the entire 312-part training set and the generalization to unseen samples (12 parts reserved as unseen). Accordingly, RF and XGBoost were further evaluated using a distinct set of prints. Three artifacts were printed across three various powder size ranges of Ti64, featuring more complex geometries. (artifact illustrated in Figure 6-1).

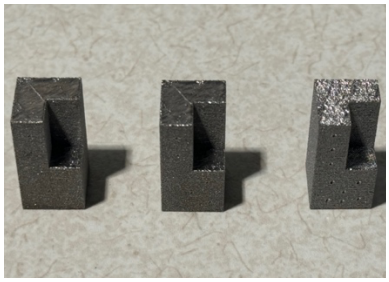


Figure 6-1: Artifacts printed using each of the three batches (15-53 µm, 15-106 µm, and 45-106 µm) of Ti64 powder feedstock.

The in-situ light intensity signals recorded during the printing of the artifacts were processed as described in Chapter 4 and were used, along with the associated part properties, as input features for a new set of unseen samples in the RF algorithm. The performance of the algorithm in regressing the process parameters for these artifacts was then assessed, as reported below. It is important to note that since the artifacts originate from an entirely different set of prints, with significant differences in geometry and a markedly different distribution (especially in the MPM signals), the data for these new test samples were normalized separately using their normalization parameters. While such a procedure is not typically employed in regression, it was applied here due to the substantial differences between the new test samples and the training set.

6.3 Results and discussion

According to the Figure 6-2(a-d), the RF model achieved a satisfactory level of regression accuracy on the new artifacts, despite the absence of new input features to account for the significant geometric differences between the artifacts and the original prints of cubical and cylindrical parts. As shown in Figure 6-2(a), the lowest accuracy was observed in predicting the power process parameters of the artifacts. RF resulted in an overall average MAPE level of 26% error across all four targets. In contrast, the XGBoost regressor was able to predict the power within a more acceptable range for LPBF applications.

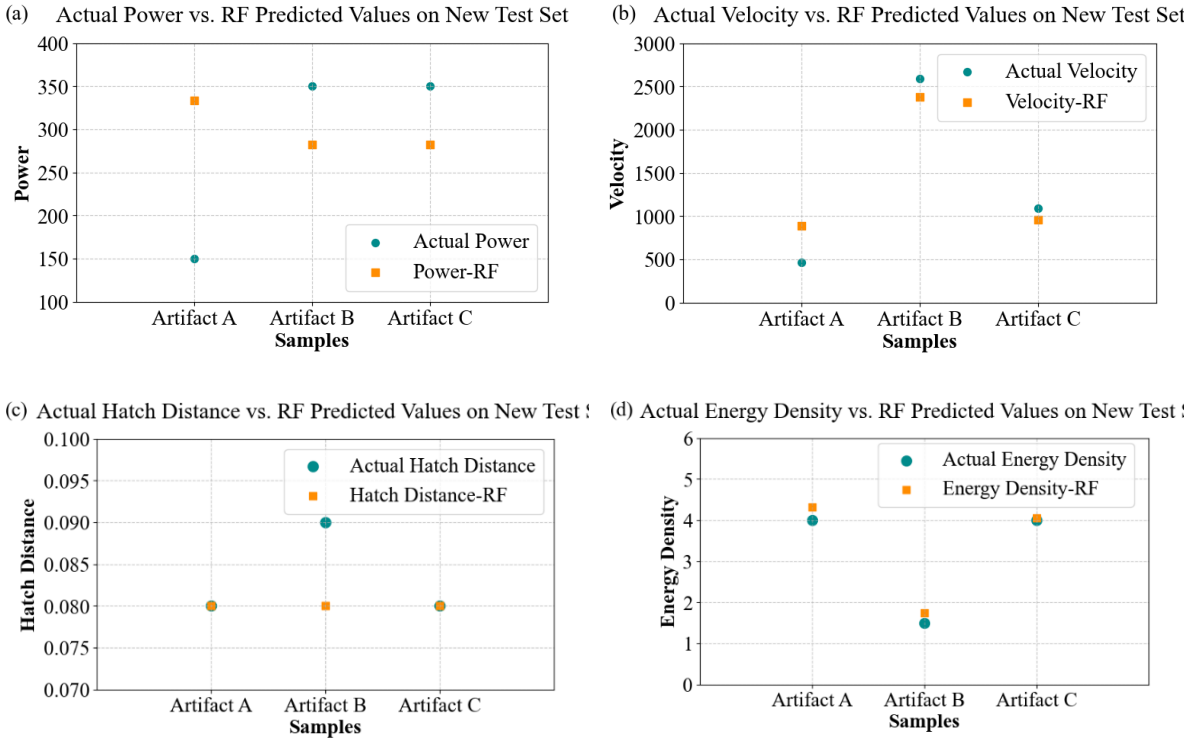


Figure 6-2: the performance of the RF and XGBoost regressors on the artifacts in estimation of the (a) power, (b) velocity, (c) hatch distance, and (d) energy density.

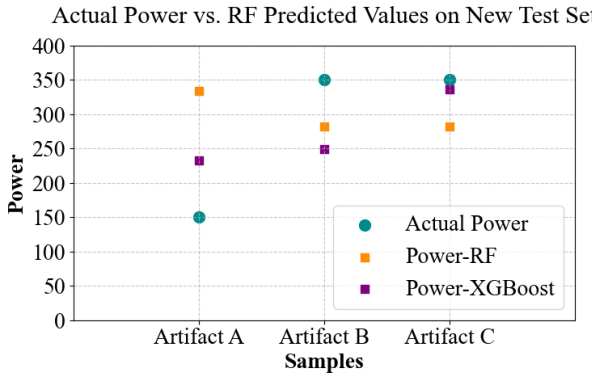


Figure 6-3: Performance of the XGBoost in the estimation of the power of the artifacts.

As depicted in Figure 6-3, the XGBoost regressor effectively predicted the power for artifact A and artifact C (corresponding to 15-53 µm and 45-106 µm powder batches, respectively). Given the distinct design conditions and significant geometric differences of the artifacts compared to the original cubical and cylindrical parts, it can be concluded that the algorithms demonstrated an acceptable level of

generalization to more complex part samples. From the detailed comparative performance of the RF and XGBoost models for each process parameter illustrated in Figure 6-4, it is evident that both models demonstrate relatively accurate regression depending on the specific process parameters. Therefore, either model could be selected as the regressor for future applications. However, since XGBoost achieved approximately half the error in the power prediction compared to RF, and their performance difference in the velocity estimation is negligible, it can be concluded that XGBoost achieved a relatively better generalization in this case.

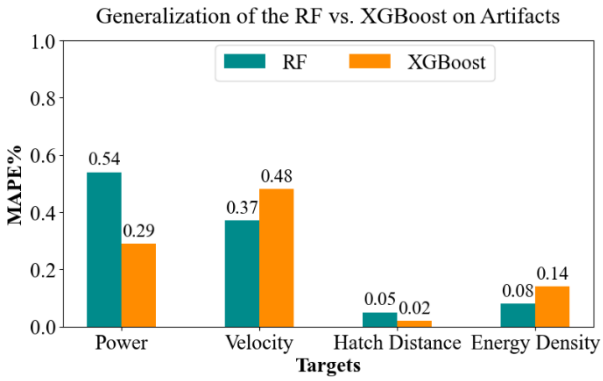


Figure 6-4: Comparative performance between the established RF and XGBoost regressors (trained on cubical and cylindrical parts dataset) in the estimation of process parameters of the artifacts.

However, to further advance the generalization capabilities of this platform, a more comprehensive dataset that represents a broader range of Ti64 parts—both in terms of geometric features and printing designs—is necessary. Additionally, a deeper investigation into feature engineering is crucial for reliably estimating process parameters in diverse prints. For example, incorporating geometry-indicating input features will be essential for achieving more precise regression in artifacts.

6.4 Summary

Tuned random forest (RF) and extreme gradient boosting (XGBoost) algorithms were developed, trained and evaluated using datasets generated from the printing of cubical and cylindrical samples, each categorized across three different particle size distribution (PSD) ranges of Ti64 powder. To test the generalization capabilities of these machine learning models, they were subsequently applied to three artifacts with more complex geometries. Despite the lack of algorithmic modifications or adjustments to the input feature set to explicitly account for variations in geometry, the models performed well. This is particularly noteworthy given the distinct light intensity melt pool monitoring

(MPM) datasets that emerged from the differences in artifact design. Nevertheless, both RF and XGBoost demonstrated a satisfactory level of accuracy in predicting process parameters, indicating that the models retained robustness across diverse geometrical challenges.

Chapter 7: Conclusions and Future Work

7.1 Conclusions of ML-based Powder Rheology Characterization

The primary objective of this thesis was to assess the feasibility of various machine learning (ML) techniques for predicting rheological properties (shear and dynamic) and Hall flowability of Ti6Al4V (Ti64) powder across different particle size distributions (PSD). Considering the importance of PSD in LPBF feedstock, the study aimed to predict rheological characteristics for both standard and off-size PSDs. ML algorithms like Support Vector Regression (SVR), Decision Tree (DT), and K-Nearest Neighbors (KNN), etc. were used, with the best regressor evaluated through multiple testing methods. To overcome data limitations, techniques like genetic programming-based symbolic regression and modified SMOTE were explored. The final models were tested on unseen off-size samples to evaluate their generalization. The conclusions drawn from this study are summarized below:

7.1.1 Dynamic Rheology:

- For dynamic rheology, the SVR algorithm, using the small original dataset, achieved accurate predictions of specific energy (SE) across all performance evaluation metrics. The results were similarly promising when tested on the new set of off-size samples, particularly after excluding one outlier sample that fell outside the SE measurement range of the dataset.
- For the basic flow energy (BFE) target in dynamic flowability, it was concluded that SVR trained on a SMOTE-augmented dataset provided a more accurate estimation, though it did not perform as well as it did for the SE target. The SVR model trained on augmented data was then evaluated using the new unseen powder samples. The results indicated that acceptable generalization was achieved, particularly when the outlier sample was excluded.
- In BFE results, the solid performance observed in evaluation modes that used a subset of new samples as a training set (single sample prediction (SSP) and repeated random split prediction (RRSP) modes) suggests that achieving highly accurate estimations for BFE is feasible, but it requires expanding the original dataset to include a broader range of PSD variations.
- Although the primary trained SVR algorithm (obtained from the analysis in Chapter 3) demonstrated good generalization performance, it was observed that due to the limited dataset, retuning the model with an expanded dataset could further enhance accuracy.

7.1.2 Shear Rheology:

- Unconfined yield strength (UYS) is the most challenging target to model and regress due to the complexity of its measurement procedure and the higher likelihood of measurement uncertainty.
- For shear rheology targets, it was concluded that equation fitting using the GP-SR algorithm on the training set significantly reduced estimation errors for UYS at 3 kPa and 9 kPa targets (UYS3 and UYS9). However, the GP-SR technique was not effective for fitting UYS6 data. As a result, the SVR algorithm trained on SMOTE-augmented data was tested, which, while reducing the error, did not achieve a highly accurate and reliable estimation.
- In evaluations using new, unseen powder samples, it was observed that the fixed SVR regressor, trained on the same augmented data with random subsets of the new samples (RRSP mode), generated predictions with accuracy comparable to previous results. This suggests that further improvement in UYS6 modeling necessitates a more variable dataset. None of the methods examined in this study proved as effective for UYS6 as they did for other targets.
- Additionally, tests involving new samples revealed that both GP-SR and SVR algorithms are highly sensitive to outlier samples. The GP-SR algorithm was able to estimate UYS with good accuracy for in-range samples; however, notable errors were observed with outlier samples (one out of three in UYS3 and two out of three in UYS9). Similarly, for UYS6, the SVR results showed approximately a 7% reduction in error when the outlier sample was excluded from the RRSP test.

7.1.3 Hall Flowability:

- After evaluating all ML algorithms, it was ultimately concluded that, given the available dataset, the SVR algorithm trained on the SMOTE-augmented dataset was the most effective method, achieving good performance in hall flowability regression.
- The same algorithm was subsequently tested on new powder samples. Overall, the results across all evaluation modes remained consistent and comparable with the previous results obtained from the original dataset.
- In terms of generalization to new, unseen samples, using the fixed augmented training data and the tuned algorithm from the previous analysis (Chapter 3), an acceptable level of regression accuracy was achieved, with a mean absolute percentage error (MAPE) of approximately 12%.

- Throughout all evaluations involving new samples, it was observed that the inclusion of new off-size samples, all from the coarse powder size range (>45 µm), led to a decrease in prediction accuracy for samples from the fine size range (15 µm). This imbalance in the dataset in terms of hall flowability characteristics likely contributed to the overall reduction in performance.

7.2 Conclusions of ML-based Process Parameters Tailoring Using Photodiode Melt Pool Monitoring

A photodiode sensor was used to record light intensity signals from three print rounds of cubical and cylindrical parts, each using a different PSD range of Ti64 powder (15-53 µm, 15-106 µm, and 45-106 µm). These signals were clustered using the self-organizing maps (SOM) algorithm. Part density, hardness, and surface roughness, along with light intensity cluster signatures and powder PSD data, were input into three fine-tuned models: feed-forward neural network (FFN), random forest (RF), and XGBoost. The models predicted process parameters, and their performance was assessed using cross-validation, regression accuracy metrics, and testing on unseen parts. Generalization was also evaluated using three complex artifacts, one from each powder PSD range. The following conclusions can be drawn from this study:

- In the cross-validation analysis conducted on the training samples, both RF and XGBoost demonstrated excellent performance in terms of statistical metrics (R^2, error deviation) and regression metrics (error). The FFN algorithm exhibited higher regression errors across all targets and a wider deviation of errors, indicating sensitivity to changes in data. Despite this, FFN still achieved acceptable performance in terms of regression accuracy.
- The CV analysis revealed that all three algorithms struggled to achieve the same level of accuracy for hatch distance as they did for other targets. As discussed in Chapter 3, the lower variation in hatch distance may have contributed to the less accurate data fitting for this target.
- The generalization performance of the model on unseen cubical and cylindrical parts across three PSD powder ranges demonstrated that XGBoost achieved lower overall regression error (with an average MAPE of ~2%) compared to RF (which had a MAPE of ~2.5%). However, RF achieved a higher R^2 score (~0.94) compared to XGBoost (~0.90), indicating

a slightly greater probability of good generalization. But in general, these suggest that both RF and XGBoost can provide accurate predictions of process parameters based on powder PSD, light intensity signatures, and part physical properties.

- Consistent with the cross-validation analysis, the lowest accuracy for all three regressors was observed in the hatch distance target during tests with unseen samples.
- Although FFN demonstrated less accuracy compared to RF and XGBoost on unseen samples, its performance was still considered relatively good, with an R^2 score above 0.8 and an overall average MAPE error of approximately 5% across all targets.
- In the additional evaluation of the RF and XGBoost regressors on a complex geometry artifact, it was concluded that, despite the lack of suitable input features related to geometric characteristics in the study design and dataset, and the fact that the artifacts were from a different set of prints with distinct applications generating different light intensity signals, the results were still deemed acceptable. Overall, RF and XGBoost achieved MAPE errors of approximately 26% and 23%, respectively.

7.3 Future Work

The two primary objectives of this thesis were to explore the feasibility and potential of adopting ML in underexplored aspects of LPBF, particularly when typical ML applications, such as the need for large datasets, may not be applicable. This research focused on assessing ML's viability for unconventional purposes within LPBF, laying a preliminary but promising foundation for future advancements. The findings suggest that with the application of more advanced, innovative, and comprehensive techniques, further enhancements in LPBF could be realized. Recommendations for future work to advance both objectives of this thesis are outlined below.

The algorithms developed and evaluated in this thesis for ML-based powder characterization were specifically designed to address an author-defined application: estimating rheology and flowability based on simpler powder characteristics across various PSD ranges, with a focus on the Ti64 material.

- **Expand the study's scope** by collecting powder characteristic datasets from a broader range of materials.
- **Enhance feature engineering** by incorporating flowability-correlated properties, such as surface energy, to improve the accuracy of rheology and flowability estimations.
- **Improve experimental data collection** by increasing the number of measurement repetitions, particularly for shear rheology (unconfined yield strength) targets.
- **Adopt advanced data augmentation techniques**, such as conditional generative adversarial networks (CGAN), to generate synthetic datasets that closely mimic the original data while introducing new information. However, a preliminary extension and refinement of the raw data is necessary to adequately train these neural networks, ensuring the generation of high-quality, application-relevant synthetic datasets.
- **Utilize powder behavior simulation models** to artificially generate additional data samples or introduce new, effective input features for the algorithms.
- **Integrate image and numerical datasets** and apply image processing and deep learning algorithms to the combined data, potentially addressing data limitations while extracting powder behavior-specific insights from numerical data.
- **Conduct extensive data collection** from available online databases to gather a diverse range of characteristics for multiple metal powder materials.

Similar to the first objective, the second objective of this research was chosen due to the limited studies that have utilized MPM and ML to develop more effective process parameter tailoring systems. This approach integrates critical quality-impacting factors such as powder feedstock size range and melt pool signatures. The study developed an off-situ process parameter tailoring technique, which, in its current form, could serve as a valuable tool in the design of experiments and optimization processes. While traditional methods often involve multiple extensive steps of statistical design optimization—from simple to advanced models—this technique offers a more time- and cost-effective alternative. However, this methodology holds significant potential for further investigation, as outlined below:

- **Extend material-specific applications**: Similar to the previous objective, this study is focused on a specific material and application. Future work could expand the scope by incorporating data from parts fabricated with different materials.
- **Broaden the application scope**: The approach could be applied to other part-quality related features, such as tensile and yield strength.
- **Expand input feature sets**: Including additional quality-impacting factors like part location on the build plate and geometrical features could make the algorithms more physics-informed, leading to more accurate process parameter optimization and applicability to complex artifacts.
- **Optimize input features**: Investigating the use of dimensionality reduction techniques, particularly for light intensity signature datasets, could enhance the performance of ML algorithms, especially in FFN.
- **Develop complementary algorithms**: Creating an additional algorithm to complement the existing one could elevate the current technique to a comprehensive alternative for process parameter design in LPBF. This complementary algorithm could provide rough estimates of light intensity signatures based on user-defined mechanical properties, enabling the generation of a priori process parameters without the need for initial test prints.
- **Exploring the application of recurrent neural networks**: Recurrent neural networks (RNNs) were not utilized in this thesis due to the transformation of photodiode signals into statistical features. However, RNNs could be trained on real-time time-series signals for integration into in-situ closed-loop quality control systems, enabling dynamic monitoring and adjustment in LPBF.

Letter of copyright permissions

Chapter 2:

Figure 2-1:



RightsLink

Laser-matter interactions in additive manufacturing of stainless steel SS316L and 13-93 bioactive glass revealed by in situ X-ray imaging

Author:
Chu Lun Alex Leung,Sebastian Marussi,Michael Towrie,Jesus del Val Garcia,Robert C. Atwood,Andrew J. Bodey,Julian R. Jones,Philip J. Withers,Peter D. Lee

Publication: Additive Manufacturing

Publisher: Elsevier

Date: December 2018

© 2018 The Authors. Published by Elsevier B.V.

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact Elsevier Global Rights Department.

Are you the author of this Elsevier journal article?

© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and ConditionsComments? We would like to hear from you. E-mail us at customercare@copyright.com

Figure 2-2:





Processing parameters in laser powder bed fusion metal additive manufacturing

Author: J.P. Oliveira,A.D. LaLonde,J. Ma
Publication: Materials & Design
Publisher: Elsevier
Date: August 2020

© 2020 The Authors. Published by Elsevier Ltd.

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

To request permission for a type of use not listed, please contact Elsevier Global Rights Department.

Are you the author of this Elsevier journal article?

© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and ConditionsComments? We would like to hear from you. E-mail us at customercare@copyright.com

Figure 2-4:

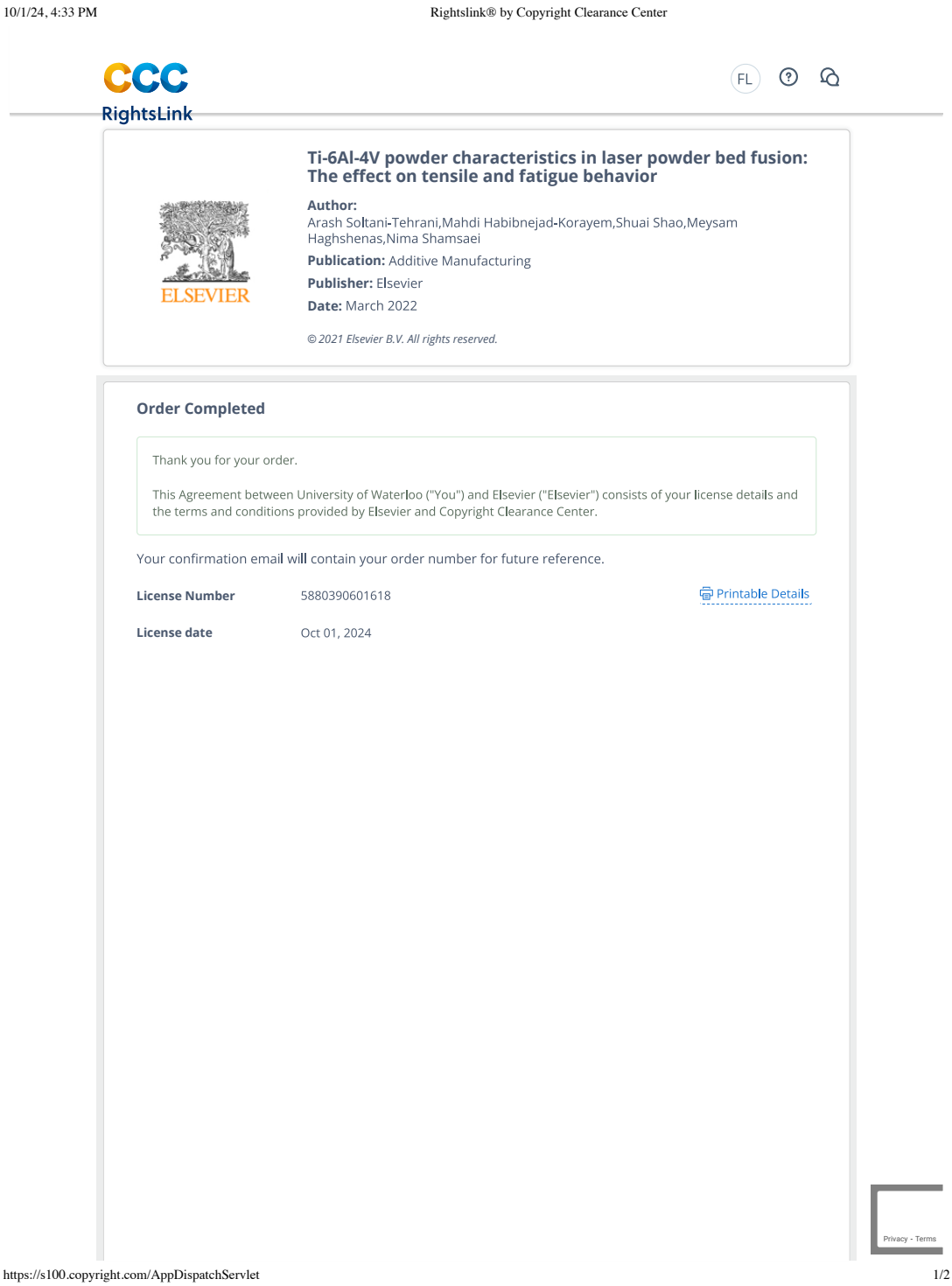
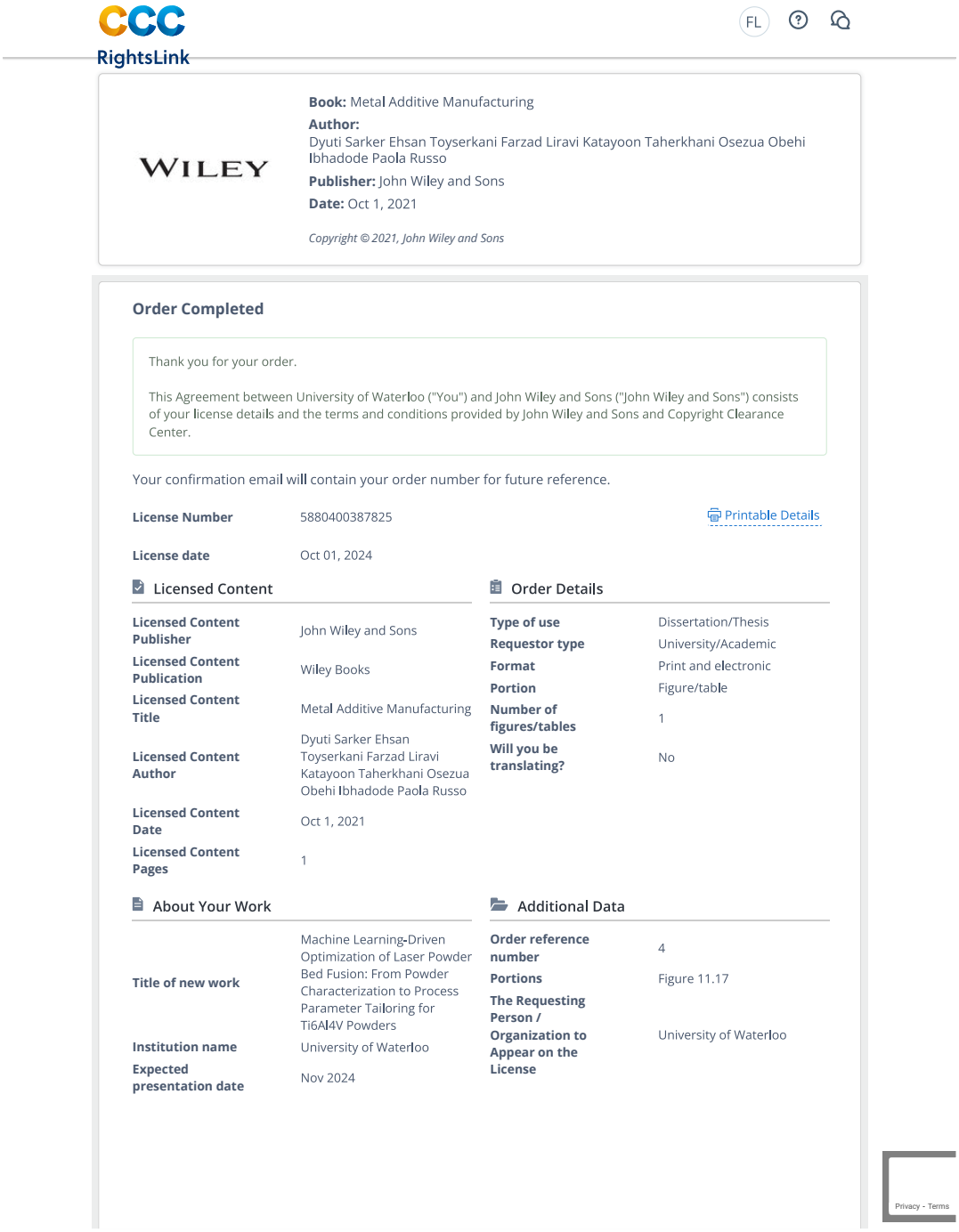

10/1/24, 4:33 PM
Rightslink® by Copyright Clearance Center
RightsLink
FL
Ti-6Al-4V powder characteristics in laser powder bed fusion: The effect on tensile and fatigue behavior
Author:
Arash Soltani-Tehrani,Mahdi Habibnejad-Korayem,Shuai Shao,Meysam Haghshenas,Nima Shamsaei
Publication: Additive Manufacturing
Publisher: Elsevier
Date: March 2022
© 2021 Elsevier B.V. All rights reserved.
ELSEVIER
Order Completed
Thank you for your order.
This Agreement between University of Waterloo ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.
Your confirmation email will contain your order number for future reference.
License Number
5880390601618
Printable Details
License date
Oct 01, 2024
Privacy - Terms
https://s100.copyright.com/AppDispatchServlet
1/2

Figure 2-5:


10/1/24, 4:49 PM
Rightslink® by Copyright Clearance Center
RightsLink
A study on surface morphology and tension in laser powder bed fusion of Ti-6Al-4V
Author: Mahyar Khorasani et al
Publication: The International Journal of Advanced Manufacturing Technology
Publisher: Springer Nature
Date: Nov 4, 2020
Copyright © 2020, The Author(s)
SPRINGER NATURE
Creative Commons
This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
You are not required to obtain permission to reuse this article.
To request permission for a type of use not listed, please contact Springer Nature
© 2024 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Data Security and Privacy | For California Residents | Terms and ConditionsComments? We would like to hear from you. E-mail us at customercare@copyright.com
https://s100.copyright.com/AppDispatchServlet?title=A study on surface morphology and tension in laser powder bed fusion of Ti-6Al-4V&author=Mahyar Khorasan… 1/1

Figure 2-6:



RightsLink

WILEY

Book: Metal Additive Manufacturing
Author: Dyuti Sarker Ehsan Toyserkani Farzad Liravi Katayoon Taherkhani Osezua Obehi Ibhadode Paola Russo
Publisher: John Wiley and Sons
Date: Oct 1, 2021

Copyright © 2021, John Wiley and Sons

Order Completed

Thank you for your order.

This Agreement between University of Waterloo ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Printable Details

License Number	5880400387825
License date	Oct 01, 2024

Licensed Content

Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Wiley Books
Licensed Content Title	Metal Additive Manufacturing
Licensed Content Author	Dyuti Sarker Ehsan Toyserkani Farzad Liravi Katayoon Taherkhani Osezua Obehi Ibhadode Paola Russo
Licensed Content Date	Oct 1, 2021
Licensed Content Pages	1

Order Details

Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No

About Your Work

Title of new work	Machine Learning-Driven Optimization of Laser Powder Bed Fusion: From Powder Characterization to Process Parameter Tailoring for Ti6Al4V Powders
Institution name	University of Waterloo
Expected presentation date	Nov 2024

Additional Data

Order reference number	4
Portions	Figure 11.17
The Requesting Person / Organization to Appear on the License	University of Waterloo

Privacy - Terms

Figure 2-7:

© 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (**https://creativecommons.org/licenses/by/4.0/**).

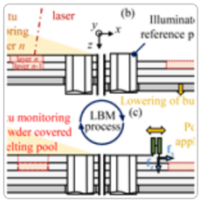
Figure 2-8:

https://doi.org/10.5194/jsss-10-219-2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Review paper | 13 Aug 2021

Methods and procedure of referenced in situ control of lateral contour displacements in additive manufacturing

Martin Lerchen, Jakob Hornung, Yu Zou, and Tino Hausotte



Chapter 3:

RightsLink

On the performance of expert-augmented machine learning with limited experimental data collected from powder particle characteristics used in laser powder bed fusion

Author: Farima Liravi et al
Publication: The International Journal of Advanced Manufacturing Technology
Publisher: Springer Nature
Date: Jul 9, 2024

Copyright © 2024, The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature

Order Completed

Thank you for your order.

This Agreement between University of Waterloo ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number 5880410390207 — Printable Details

License date Oct 01, 2024

Licensed Content

Licensed Content Publisher	Springer Nature
Licensed Content Publication	The International Journal of Advanced Manufacturing Technology
Licensed Content Title	On the performance of expert-augmented machine learning with limited experimental data collected from powder particle characteristics used in laser powder bed fusion
Licensed Content Author	Farima Liravi et al
Licensed Content Date	Jul 9, 2024

Order Details

Type of Use	Thesis/Dissertation
Requestor Type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	50000 or greater
Author of this Springer Nature content	yes

About Your Work

Title of new work	Machine Learning-Driven Optimization of Laser Powder Bed Fusion: From Powder Characterization to Process Parameter Tailoring for Ti6Al4V Powders
Institution name	University of Waterloo
Expected presentation date	Nov 2024

Additional Data

Order reference number	5
The Requesting Person / Organization to Appear on the License	University of Waterloo

Privacy - Terms

Chapter 4:

© 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (**https://creativecommons.org/licenses/by/4.0/**).

Appendix B:

Figure B-1:

RightsLink



Support vector machine applications in the field of hydrology: A review

Author: Sujay Raghavendra. N,Paresh Chandra Deka
Publication: Applied Soft Computing
Publisher: Elsevier
Date: June 2014

Copyright © 2014 Elsevier B.V. All rights reserved.

Order Completed

Thank you for your order.

This Agreement between University of Waterloo ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Printable Details

License Number	5880840919415
License date	Oct 02, 2024

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Applied Soft Computing
Licensed Content Title	Support vector machine applications in the field of hydrology: A review
Licensed Content Author	Sujay Raghavendra. N,Paresh Chandra Deka
Licensed Content Date	Jun 1, 2014
Licensed Content Volume	19
Licensed Content Issue	n/a
Licensed Content Pages	15

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

About Your Work

Title of new work	Machine Learning-Driven Optimization of Laser Powder Bed Fusion: From Powder Characterization to Process Parameter Tailoring for Ti6Al4V Powders
Institution name	University of Waterloo
Expected presentation date	Nov 2024

Additional Data

Order reference number	6
Portions	Fig. 5
The Requesting Person / Organization to Appear on the License	University of Waterloo

Privacy - Terms

References

[1] T. Pereira, J. V Kennedy, and J. Potgieter, "A comparison of traditional manufacturing vs additive manufacturing, the best method for the job," *Procedia Manuf*, vol. 30, pp. 11–18, 2019.

[2] Z. Y. Chua, I. H. Ahn, and S. K. Moon, "Process monitoring and inspection systems in metal additive manufacturing: Status and applications," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 4, no. 2, pp. 235–245, 2017, doi: 10.1007/s40684-017-0029-7.

[3] R. McCann *et al.*, "In-situ sensing, process monitoring and machine control in Laser Powder Bed Fusion_ A review," *Addit Manuf*, vol. 45, p. 102058, 2021.

[4] M. Grasso and B. M. Colosimo, "Process defects and in situ monitoring methods in metal powder bed fusion: A review," Apr. 01, 2017, *Institute of Physics Publishing*. doi: 10.1088/1361-6501/aa5c4f.

[5] F. Liravi, M. Habibnejad-Korayem, and E. Toyserkani, "On the performance of expert-augmented machine learning with limited experimental data collected from powder particle characteristics used in laser powder bed fusion," *The International Journal of Advanced Manufacturing Technology*, vol. 133, no. 11, pp. 5829–5847, 2024, doi: 10.1007/s00170-024-14044-2.

[6] F. Liravi, S. Soo, S. Toorandaz, K. Taherkhani, M. Habibnejad-Korayem, and E. Toyserkani, "Tailoring Laser Powder Bed Fusion Process Parameters for Standard and Off-Size Ti6Al4V Metal Powders: A Machine Learning Approach Enhanced by Photodiode-Based Melt Pool Monitoring," *Inventions*, vol. 9, no. 4, 2024, doi: 10.3390/inventions9040087.

[7] I. Gibson, D. Rosen, B. Stucker, and M. Khorasani, *Additive Manufacturing Technologies*, vol. 17. Springer International Publishing, 2021. doi: 10.1007/978-3-030-56127-7.

[8] D. Dev Singh, T. Mahender, and A. Raji Reddy, "Powder bed fusion process: A brief review," in *Materials Today: Proceedings*, Elsevier Ltd, 2021, pp. 350–355. doi: 10.1016/j.matpr.2020.08.415.

[9] H. Shipley *et al.*, "Optimisation of process parameters to address fundamental challenges during selective laser melting of Ti-6Al-4V: A review," May 01, 2018, *Elsevier Ltd*. doi: 10.1016/j.ijmachtools.2018.01.003.

[10] S. Sun, M. Brandt, and M. Easton, “2 - Powder bed fusion processes: An overview,” in *Laser Additive Manufacturing*, M. Brandt, Ed., Woodhead Publishing, 2017, pp. 55–77. doi: https://doi.org/10.1016/B978-0-08-100433-3.00002-6.

[11] J. J. Beaman, D. L. Bourell, C. C. Seepersad, and D. Kovar, “Additive Manufacturing Review: Early Past to Current Practice,” *J Manuf Sci Eng*, vol. 142, no. 11, Sep. 2020, doi: 10.1115/1.4048193.

[12] C. Haase *et al.*, “Exploiting process-related advantages of selective laser melting for the production of high-manganese steel,” *Materials*, vol. 10, no. 1, p. 56, 2017, doi: 10.3390/ma10010056.

[13] S. Chen, F. Tao, C. Jia, and J. Yang, “Status and Progress of Selective Laser Melting Forming Technology,” in *3rd International Conference on Machinery, Materials and Information Technology Applications*, 2015, pp. 1295–1298.

[14] C. L. A. Leung *et al.*, “Laser-matter interactions in additive manufacturing of stainless steel SS316L and 13-93 bioactive glass revealed by in situ X-ray imaging,” *Addit Manuf*, vol. 24, pp. 647–657, Aug. 2018, doi: 10.1016/j.addma.2018.08.025.

[15] J. P. Oliveira, A. D. LaLonde, and J. Ma, “Processing parameters in laser powder bed fusion metal additive manufacturing,” *Mater Des*, vol. 193, p. 108762, Apr. 2020, doi: 10.1016/j.matdes.2020.108762.

[16] T. DebRoy *et al.*, “Additive manufacturing of metallic components – Process, structure and properties,” *Prog Mater Sci*, vol. 92, pp. 112–224, Mar. 2018, doi: 10.1016/j.pmatsci.2017.10.001.

[17] A. Cooke and J. Slotwinski, “Properties of metal powders for additive manufacturing: A review of the state of the art of metal powder property testing,” 2015, pp. 21–48.

[18] “Particle Size and Shape Analyzer CAMSIZER® X2-PARTICLE CHARACTERIZATION OF METAL POWDERS.” Accessed: Oct. 01, 2024. [Online]. Available: https://www.microtrac.com/files/99760/particle-characterization-of-metal-powders.pdf

[19] S. R. Narasimharaju *et al.*, “A comprehensive review on laser powder bed fusion of steels: Processing, microstructure, defects and control methods, mechanical properties, current

challenges and future trends," *J Manuf Process*, vol. 75, pp. 375–414, Mar. 2022, doi: 10.1016/j.jmapro.2021.12.033.

[20] L. Haferkamp *et al.*, "The Influence of Particle Shape, Powder Flowability, and Powder Layer Density on Part Density in Laser Powder Bed Fusion," *Metals (Basel)*, vol. 11, no. 3, 2021, doi: 10.3390/met11030418.

[21] K. Riener *et al.*, "Influence of particle size distribution and morphology on the properties of the powder feedstock as well as of AlSi10Mg parts produced by laser powder bed fusion (LPBF)," *Addit Manuf*, vol. 34, Aug. 2020, doi: 10.1016/j.addma.2020.101286.

[22] A. Mussatto, R. Groarke, A. O'Neill, M. A. Obeidi, Y. Delaure, and D. Brabazon, "Influences of powder morphology and spreading parameters on the powder bed topography uniformity in powder bed fusion metal additive manufacturing," *Addit Manuf*, vol. 38, p. 101807, 2021.

[23] S. E. Brika, M. Letenneur, C. A. Dion, and V. Brailovski, "Influence of particle morphology and size distribution on the powder flowability and laser powder bed fusion manufacturability of Ti-6Al-4V alloy," *Addit Manuf*, vol. 31, p. 100929, 2020.

[24] A. Soltani-Tehrani, M. Habibnejad-Korayem, S. Shao, M. Haghshenas, and N. Shamsaei, "Ti-6Al-4V powder characteristics in laser powder bed fusion_ The effect on tensile and fatigue behavior," *Additive Manufacturing* , vol. 51, p. 102584, 2022.

[25] P. Kiani, U. Scipioni Bertoli, A. D. Dupuy, K. Ma, and J. M. Schoenung, "A Statistical Analysis of Powder Flowability in Metal Additive Manufacturing," *Adv Eng Mater*, vol. 22, no. 10, Oct. 2020, doi: 10.1002/adem.202000022.

[26] Y. Ma, T. M. Evans, N. Philips, and N. Cunningham, "Numerical simulation of the effect of fine fraction on the flowability of powders in additive manufacturing," *Powder Technol*, vol. 360, pp. 608–621, Jan. 2020, doi: 10.1016/j.powtec.2019.10.041.

[27] Y. Zhao *et al.*, "Controlling factors determining flowability of powders for additive manufacturing: A combined experimental and simulation study," *Powder Technol*, vol. 393, pp. 482–493, Nov. 2021, doi: 10.1016/j.powtec.2021.08.006.

[28] Z. Young *et al.*, "Effects of Particle Size Distribution with Efficient Packing on Powder Flowability and Selective Laser Melting Process," *Materials*, vol. 15, no. 3, Feb. 2022, doi: 10.3390/ma15030705.

[29] Y. Tan, J. Zhang, X. Li, Y. Xu, and C. Y. Wu, "Comprehensive evaluation of powder flowability for additive manufacturing using principal component analysis," *Powder Technol*, vol. 393, pp. 154–164, Nov. 2021, doi: 10.1016/j.powtec.2021.07.069.

[30] A. T. Sutton, C. S. Kriewall, M. C. Leu, and J. W. Newkirk, "Powder characterisation techniques and effects of powder characteristics on part properties in powder-bed fusion processes," Jan. 02, 2017, *Taylor and Francis Ltd.* doi: 10.1080/17452759.2016.1250605.

[31] M. A. Spurek, L. Haferkamp, C. Weiss, A. B. Spierings, J. H. Schleifenbaum, and K. Wegener, "Influence of the particle size distribution of monomodal 316L powder on its flowability and processability in powder bed fusion," *Progress in Additive Manufacturing*, vol. 7, no. 4, pp. 533–542, Aug. 2022, doi: 10.1007/s40964-021-00240-z.

[32] S. Ziri, A. Hor, and C. Mabru, "Combined effect of powder properties and process parameters on the density of 316L stainless steel obtained by laser powder bed fusion," *International Journal of Advanced Manufacturing Technology*, vol. 120, no. 9–10, pp. 6187–6204, Jun. 2022, doi: 10.1007/s00170-022-09160-w.

[33] M. H. Sehhat, A. T. Sutton, C. H. Hung, J. W. Newkirk, and M. C. Leu, "Investigation of Mechanical Properties of Parts Fabricated with Gas- and Water-Atomized 304L Stainless Steel Powder in the Laser Powder Bed Fusion Process," *JOM*, vol. 74, no. 3, pp. 1088–1095, Mar. 2022, doi: 10.1007/s11837-021-05029-7.

[34] F. Chu *et al.*, "Influence of satellite and agglomeration of powder on the processability of AlSi10Mg powder in Laser Powder Bed Fusion," *Journal of Materials Research and Technology*, vol. 11, pp. 2059–2073, Mar. 2021, doi: 10.1016/j.jmrt.2021.02.015.

[35] J. S. Weaver *et al.*, "The effects of particle size distribution on the rheological properties of the powder and the mechanical properties of additively manufactured 17-4 PH stainless steel," *Addit Manuf*, vol. 39, Mar. 2021, doi: 10.1016/j.addma.2021.101851.

[36] M. Xue, X. Chen, X. Ji, X. Xie, Q. Chao, and G. Fan, "Effect of Particle Size Distribution on the Printing Quality and Tensile Properties of Ti-6Al-4V Alloy Produced by LPBF Process," *Metals (Basel)*, vol. 13, no. 3, 2023, doi: 10.3390/met13030604.

[37] S. Hu *et al.*, "Influence of the properties of Ti6Al4V powder materials on the build quality of laser powder bed fusion–manufactured components," *The International Journal of Advanced*

Manufacturing Technology, vol. 129, no. 1, pp. 969–982, 2023, doi: 10.1007/s00170-023-12244-w.

[38] N. Koju, J. Hermes, S. E. Saghaian, and L. Yang, "Laser powder bed fusion additively manufactured thin lightweight Ti6Al4V features: an experimental investigation on the influence of powder feedstock, geometry, and process parameters on property/quality," *The International Journal of Advanced Manufacturing Technology*, vol. 130, no. 3, pp. 1541–1561, 2024, doi: 10.1007/s00170-023-12712-3.

[39] M. Habibnejad-Korayem, M. Lalh, C. Schunk, and Y. Zou, "Offsize particle size utilization for laser powder bed fusion processing of plasma atomized Ti-6Al-4V powders: Impacts on part properties and powder safety," *J Manuf Process*, vol. 107, pp. 559–573, 2023, doi: https://doi.org/10.1016/j.jmapro.2023.10.036.

[40] I. Ludwig and M. Kluge, "Investigation of an Increased Particle Size Distribution of Ti-6Al-4V Powders Used for Laser-Based Powder Bed Fusion of Metals," *Materials*, vol. 17, no. 12, 2024, doi: 10.3390/ma17122942.

[41] P. Bourot, L. Weiss, D. Boisselier, and P. Laheurte, "Effect of bimodal size distribution on the behavioural properties of powders used in the LPBF process," *Powder Technol*, vol. 439, p. 119670, 2024, doi: https://doi.org/10.1016/j.powtec.2024.119670.

[42] Y. Wang, X. Chen, S. Jayalakshmi, R. A. Singh, K. Sergey, and M. Gupta, "Process Parameters, Product Quality Monitoring, and Control of Powder Bed Fusion," in *Transactions on Intelligent Welding Manufacturing*, 2020, pp. 89–108. doi: 10.1007/978-981-13-8192-8_4.

[43] A. M. Beese and B. E. Carroll, "Review of Mechanical Properties of Ti-6Al-4V Made by Laser-Based Additive Manufacturing Using Powder Feedstock," *JOM*, vol. 68, no. 3, pp. 724–734, Mar. 2016, doi: 10.1007/s11837-015-1759-z.

[44] E. Malekipour and H. El-Mounayri, "Common defects and contributing parameters in powder bed fusion AM process and their classification for online monitoring and control: a review," *International Journal of Advanced Manufacturing Technology*, vol. 95, no. 1–4, pp. 527–550, Mar. 2018, doi: 10.1007/s00170-017-1172-6.

[45] H. Gong, K. Rafi, H. Gu, G. J. Ram, T. Starr, and B. Stucker, “Influence of defects on mechanical properties of Ti–6Al–4V components produced by selective laser melting and electron beam melting,” *Mater Des*, vol. 86, pp. 545–554, 2015.

[46] A. M. Khorasani, I. Gibson, A. Ghaderi, and M. I. Mohammed, “Investigation on the effect of heat treatment and process parameters on the tensile behaviour of SLM Ti-6Al-4V parts,” *International Journal of Advanced Manufacturing Technology*, vol. 101, no. 9–12, pp. 3183–3197, Apr. 2019, doi: 10.1007/s00170-018-3162-8.

[47] A. Khorasani, I. Gibson, U. S. Awan, and A. Ghaderi, “The effect of SLM process parameters on density, hardness, tensile strength and surface quality of Ti-6Al-4V,” *Addit Manuf*, vol. 25, pp. 176–186, 2019.

[48] A. M. Khorasani, I. Gibson, A. Ghasemi, and A. Ghaderi, “A comprehensive study on variability of relative density in selective laser melting of Ti-6Al-4V,” *Virtual Phys Prototyp*, vol. 14, no. 4, pp. 349–359, Oct. 2019, doi: 10.1080/17452759.2019.1614198.

[49] J. W. Pegues *et al.*, “Fatigue of additive manufactured Ti-6Al-4V, Part I_ The effects of powder feedstock, manufacturing, and post-process conditions on the resulting microstructure and defects,” *Int J Fatigue*, vol. 132, p. 105358, 2020.

[50] M. Khorasani *et al.*, “A study on surface morphology and tension in laser powder bed fusion of Ti-6Al-4V,” *The International Journal of Advanced Manufacturing Technology*, vol. 111, pp. 2891–2909, 2020, doi: 10.1007/s00170-020-06221-w/Published.

[51] E. Toyserkani, D. Sarker, F. Liravi, P. Russo, and K. Taherkhani, *Metal Additive Manufacturing*. 2021. [Online]. Available: www.wiley.com

[52] Z. Yan *et al.*, “Review on thermal analysis in laser-based additive manufacturing,” *Opt Laser Technol*, vol. 106, pp. 427–441, 2018.

[53] P. Yadav, O. Rigo, C. Arvieu, E. Le Guen, and E. Lacoste, “In situ monitoring systems of the SLM process: On the need to develop machine learning models for data processing,” Jun. 01, 2020, *MDPI AG*. doi: 10.3390/cryst10060524.

[54] C. Schwerz and L. Nyborg, “Linking in situ melt pool monitoring to melt pool size distributions and internal flaws in laser powder bed fusion,” *Metals (Basel)*, vol. 11, no. 11, Nov. 2021, doi: 10.3390/met11111856.

[55] M. Lerchen, J. Hornung, Y. Zou, and T. Hausotte, “Methods and procedure of referenced in situ control of lateral contour displacements in additive manufacturing,” *Journal of Sensors and Sensor Systems*, vol. 10, pp. 219–232, Aug. 2021, doi: 10.5194/jsss-10-219-2021.

[56] “InfiniAM Spectral-Energy input and melt pool emissions monitoring for AM systems.” [Online]. Available: https://www.renishaw.com/en/infiniam-spectral--42310

[57] T. Toeppel *et al.*, “3D Analysis in Laser Beam Melting Based on Real-time Process Monitoring,” in *Mater Sci Technol Conf*, 2016, pp. 123–132.

[58] D. Alberts, D. Schwarze, and G. Witt, “High speed melt pool & laser power monitoring for selective laser melting (SLM®),” in *9th International Conference on Photonic Technologies LANE*, 2016.

[59] G. Seo, Md. R. U. Ahsan, Y. Lee, J.-H. Shin, H. Park, and D. B. Kim, “A functional modeling approach for quality assurance in metal additive manufacturing,” *Rapid Prototyp J*, vol. 27, no. 2, pp. 288–303, Jan. 2021, doi: 10.1108/RPJ-12-2018-0312.

[60] R. Valente *et al.*, “Classifying Powder Flowability for Cold Spray Additive Manufacturing Using Machine Learning,” in *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 2919–2928. doi: 10.1109/BigData50022.2020.9377948.

[61] R. Tkachenko, Z. Duriagina, I. Lemishka, I. Izonin, and A. Trostianchyn, “Development of machine learning method of titanium alloy properties identification in additive technologies,” *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 12–93, pp. 23–31, 2018, doi: 10.15587/1729-4061.2018.134319.

[62] B. L. DeCost, H. Jain, A. D. Rollett, and E. A. Holm, “Computer Vision and Machine Learning for Autonomous Characterization of AM Powder Feedstocks,” *JOM*, vol. 69, no. 3, pp. 456–465, Mar. 2017, doi: 10.1007/s11837-016-2226-1.

[63] J. Zhang, M. Habibnejad-korayem, Z. Liu, T. Lyu, Q. Sun, and Y. Zou, “A Computer Vision Approach to Evaluate Powder Flowability for Metal Additive Manufacturing,” *Integr Mater Manuf Innov*, vol. 10, no. 3, pp. 429–443, Sep. 2021, doi: 10.1007/s40192-021-00226-3.

[64] Z. Liang, Z. Nie, A. An, J. Gong, and X. Wang, “A particle shape extraction and evaluation method using a deep convolutional neural network and digital image processing,” *Powder Technol*, vol. 353, pp. 156–170, 2019.

[65] X. Zhou, N. Dai, X. Cheng, A. Thompson, and R. Leach, “Intelligent classification for three-dimensional metal powder particles,” *Powder Technology*, vol. 397, p. 117018, 2022.

[66] G. Bakas *et al.*, “A Tool for Rapid Analysis Using Image Processing and Artificial Intelligence: Automated Interoperable Characterization Data of Metal Powder for Additive Manufacturing with SEM Case,” *Metals (Basel)*, vol. 12, no. 11, 2022, doi: 10.3390/met12111816.

[67] N. Sandler and D. Wilson, “Prediction of granule packing and flow behavior based on particle size and shape analysis,” *J Pharm Sci*, vol. 99, no. 2, pp. 958–968, 2010, doi: 10.1002/jps.21884.

[68] A. Schuliger, S. Price, B. C. Sousa, D. L. Cote, and R. Neamtu, “Multiple-Instance Regression for Metal Powder Hall Flow Rate Prediction Using Augmented Particle Size and Shape Data,” *Powders*, vol. 2, no. 1, pp. 189–204, 2023, doi: 10.3390/powders2010013.

[69] D. Mutel, S. Gélinas, and C. Blais, “Rheological characterisation of water atomised tool steel powders developed for laser powder bed fusion by supervised and unsupervised machine learning,” *Powder Metallurgy*, vol. 66, no. 3, pp. 195–207, Jul. 2023, doi: 10.1080/00325899.2023.2191236.

[70] P. Yadav, O. Rigo, C. Arvieu, E. Le Guen, and E. Lacoste, “Drift Detection in Selective Laser Melting (SLM) Using a Machine Learning Approach,” in *Industrializing Additive Manufacturing*, Springer International Publishing, 2021, pp. 177–191. doi: 10.1007/978-3-030-54334-1_13.

[71] A. J. Dunbar and A. R. Nassar, “Assessment of optical emission analysis for in-process monitoring of powder bed fusion additive manufacturing,” *Virtual Phys Prototyp*, vol. 13, no. 1, pp. 14–19, Jan. 2018, doi: 10.1080/17452759.2017.1392683.

[72] M. Montazeri, A. R. Nassar, A. J. Dunbar, and P. Rao, “In-process monitoring of porosity in additive manufacturing using optical emission spectroscopy,” *IISE Trans*, vol. 52, no. 5, pp. 500–515, May 2020, doi: 10.1080/24725854.2019.1659525.

[73] J. Harbig *et al.*, “Methodology to Determine Melt Pool Anomalies in Powder Bed Fusion of Metals Using a Laser Beam by Means of Process Monitoring and Sensor Data Fusion,” *Materials*, vol. 15, no. 3, Feb. 2022, doi: 10.3390/ma15031265.

[74] T. Zhang *et al.*, “Hardness Prediction of Laser Powder Bed Fusion Product Based on Melt Pool Radiation Intensity,” *Materials*, vol. 15, no. 13, p. 4674, Jul. 2022, doi: 10.3390/ma15134674.

[75] S. Jayasinghe, P. Paoletti, C. Sutcliffe, J. Dardis, N. Jones, and P. L. Green, “Automatic quality assessments of laser powder bed fusion builds from photodiode sensor measurements,” *Progress in Additive Manufacturing*, vol. 7, no. 2, pp. 143–160, Apr. 2022, doi: 10.1007/s40964-021-00219-w.

[76] L. Cao, L. Yu, J. Xu, S. Luo, X. Huang, and J. Li, “Deep learning-based in-situ monitoring of the selective laser melting process with the photodiode signal.” [Online]. Available: https://ssrn.com/abstract=4134029

[77] K. Taherkhani, C. Eischer, and E. Toyserkani, “An unsupervised machine learning algorithm for in-situ defect-detection in laser powder-bed fusion,” *J Manuf Process*, vol. 81, pp. 476–489, Sep. 2022, doi: 10.1016/j.jmapro.2022.06.074.

[78] M. Montazeri and P. Rao, “Sensor-Based Build Condition Monitoring in Laser Powder Bed Fusion Additive Manufacturing Process Using a Spectral Graph Theoretic Approach,” *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 140, no. 9, Sep. 2018, doi: 10.1115/1.4040264.

[79] I. A. Okaro, S. Jayasinghe, C. Sutcliffe, K. Black, P. Paoletti, and P. L. Green, “Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning,” *Addit Manuf*, vol. 27, pp. 42–53, May 2019, doi: 10.1016/j.addma.2019.01.006.

[80] V. Pandiyan *et al.*, “Deep Learning-based Monitoring of Laser Powder Bed Fusion Process on Variable Time-Scales Using Heterogeneous Sensing and operando X-ray Radiography Guidance,” *Addit Manuf*, vol. 58, p. 103007, Jun. 2022, doi: 10.1016/j.addma.2022.103007.

[81] N. Hasan, A. Kumar Saha, A. Wessman, and M. Shafae, “Machine Learning-based Layer-wise Detection of Overheating Anomaly in LPBF using Photodiode Data,” Mar. 2024. doi: https://doi.org/10.48550/arXiv.2403.13861.

[82] Q. Wu, F. Yang, C. Lv, C. Liu, W. Tang, and J. Yang, “In-Situ Quality Intelligent Classification of Additively Manufactured Parts Using a Multi-Sensor Fusion Based Melt Pool Monitoring System,” *Additive Manufacturing Frontiers*, vol. 3, no. 3, p. 200153, 2024, doi: https://doi.org/10.1016/j.amf.2024.200153.

[83] Z. Mao *et al.*, “Continuous online flaws detection with photodiode signal and melt pool temperature based on deep learning in laser powder bed fusion,” *Opt Laser Technol*, vol. 158, p. 108877, 2023, doi: https://doi.org/10.1016/j.optlastec.2022.108877.

[84] Z. Zou *et al.*, “In situ monitoring with melt pool data based on multi-signal fusion method in laser powder bed fusion,” *Measurement*, vol. 234, p. 114877, 2024, doi: https://doi.org/10.1016/j.measurement.2024.114877.

[85] H. Liu *et al.*, “Inference of highly time-resolved melt pool visual characteristics and spatially-dependent lack-of-fusion defects in laser powder bed fusion using acoustic and thermal emission data,” *Addit Manuf*, vol. 83, p. 104057, 2024, doi: https://doi.org/10.1016/j.addma.2024.104057.

[86] C. Yao, T. Zhang, Z. An, and M. Dang, “Process identification of laser powder bed fusion anomalies based on process-monitored melt pool radiation intensity data,” *J Phys Conf Ser*, vol. 2746, no. 1, p. 12047, May 2024, doi: 10.1088/1742-6596/2746/1/012047.

[87] G. O. Barrionuevo, M. Walczak, J. Ramos-Grez, and X. Sánchez-Sánchez, “Microhardness and wear resistance in materials manufactured by laser powder bed fusion: Machine learning approach for property prediction,” *CIRP J Manuf Sci Technol*, vol. 43, pp. 106–114, 2023, doi: https://doi.org/10.1016/j.cirpj.2023.03.002.

[88] M. Kuehne, K. Bartsch, B. Bossen, and C. Emmelmann, “Predicting melt track geometry and part density in laser powder bed fusion of metals using machine learning,” *Progress in Additive Manufacturing*, vol. 8, no. 1, pp. 47–54, 2023, doi: 10.1007/s40964-022-00387-3.

[89] Z. Gu *et al.*, “A universal predictor-based machine learning model for optimal process maps in laser powder bed fusion process,” *J Intell Manuf*, vol. 34, no. 8, pp. 3341–3363, 2023, doi: 10.1007/s10845-022-02004-0.

[90] S. Dharmadhikari, N. Menon, and A. Basak, “A reinforcement learning approach for process parameter optimization in additive manufacturing,” *Addit Manuf*, vol. 71, p. 103556, 2023, doi: https://doi.org/10.1016/j.addma.2023.103556.

[91] M. Zhao, H. Wei, Y. Mao, C. Zhang, T. Liu, and W. Liao, “Predictions of Additive Manufacturing Process Parameters and Molten Pool Dimensions with a Physics-Informed Deep Learning Model,” *Engineering*, vol. 23, pp. 181–195, 2023, doi: https://doi.org/10.1016/j.eng.2022.09.015.

[92] S. Liu, A. P. Stebner, B. B. Kappes, and X. Zhang, “Machine learning for knowledge transfer across multiple metals additive manufacturing printers,” *Addit Manuf*, vol. 39, p. 101877, 2021, doi: https://doi.org/10.1016/j.addma.2021.101877.

[93] S. Eshkabilov, I. Ara, and F. Azarmi, “A comprehensive investigation on application of machine learning for optimization of process parameters of laser powder bed fusion-processed 316L stainless steel,” *The International Journal of Advanced Manufacturing Technology*, vol. 123, no. 7, pp. 2733–2756, 2022, doi: 10.1007/s00170-022-10331-y.

[94] J. Wang, S. G. Jeong, E. S. Kim, H. S. Kim, and B.-J. Lee, “Material-agnostic machine learning approach enables high relative density in powder bed fusion products,” *Nat Commun*, vol. 14, no. 1, p. 6557, 2023, doi: 10.1038/s41467-023-42319-x.

[95] M. S. Rahman, N. S. Sattar, R. U. Ahmed, J. Ciaccio, and U. K. Chakravarty, “A Machine Learning Framework for Melt-Pool Geometry Prediction and Process Parameter Optimization in the Laser Powder-Bed Fusion Process,” *J Eng Mater Technol*, vol. 146, no. 4, Aug. 2024, doi: 10.1115/1.4065687.

[96] S. Mondal, D. Gwynn, A. Ray, and A. Basak, “Investigation of Melt Pool Geometry Control in Additive Manufacturing Using Hybrid Modeling,” *Metals (Basel)*, vol. 10, no. 5, 2020, doi: 10.3390/met10050683.

[97] Q. Zhong, X. Tian, X. Huang, C. Huo, and D. Li, “Using feedback control of thermal history to improve quality consistency of parts fabricated via large-scale powder bed fusion,” *Addit Manuf*, vol. 42, p. 101986, 2021, doi: https://doi.org/10.1016/j.addma.2021.101986.

[98] F. Ogoke and A. B. Farimani, “Thermal control of laser powder bed fusion using deep reinforcement learning,” *Addit Manuf*, vol. 46, p. 102033, 2021, doi: https://doi.org/10.1016/j.addma.2021.102033.

[99] S. Lapointe, G. Guss, Z. Reese, M. Strantza, M. J. Matthews, and C. L. Druzgalski, “Photodiode-based machine learning for optimization of laser powder bed fusion parameters in complex geometries,” *Addit Manuf*, vol. 53, May 2022, doi: 10.1016/j.addma.2022.102687.

[100] “PARTICLE SIZE AND SHAPE ANALYZER CAMSIZER® X2.” Accessed: Oct. 01, 2024. [Online]. Available: https://www.microtrac.com/products/particle-size-shape-analysis/dynamic-image-analysis/camsizer-x2/

[101] “FT4 Powder Rheometer.” Accessed: Oct. 01, 2024. [Online]. Available: https://www.micromeritics.com/ft4-powder-rheometer/

[102] “AS-300 Hall Flowmeter｜Carney Flow Meter Funnel｜Metal Powder Flow Rate｜Apparent Density｜Manufacturer.” Accessed: Oct. 01, 2024. [Online]. Available: https://www.hmk-test.com/as-300-hall-flowmeter/

[103] “GranuPack: High resolution tapped density analyzer.” Accessed: Oct. 01, 2024. [Online]. Available: https://www.granutools.com/en/granupack

[104] “SO 13322-2:2021 Particle size analysis — Image analysis methods Particle size analysis — Image analysis methods — Part 2: Dynamic image analysis methods,” 2021. Accessed: Oct. 01, 2024. [Online]. Available: https://www.iso.org/standard/72566.html

[105] L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, “SMOTE for regression,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, pp. 378–389. doi: 10.1007/978-3-642-40669-0_33.

[106] K. Taherkhani, E. Sheydaeian, C. Eischer, M. Otto, and E. Toyserkani, “Development of a defect-detection platform using photodiode signals collected from the melt pool of laser powder-bed fusion,” *Addit Manuf*, vol. 46, p. 102152, 2021, doi: https://doi.org/10.1016/j.addma.2021.102152.

[107] K. Taherkhani, C. Eischer, and E. Toyserkani, “An unsupervised machine learning algorithm for in-situ defect-detection in laser powder-bed fusion,” *J Manuf Process*, vol. 81, pp. 476–489, 2022, doi: https://doi.org/10.1016/j.jmapro.2022.06.074.

[108] N. Ketkar and J. Moolayil, *Deep Learning with Python Learn Best Practices of Deep Learning Models with PyTorch*. Apress Berkeley, CA, 2021.

[109] L. Breiman, “Random Forests,” *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[110] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*. 2016. doi: 10.1145/2939672.2939785.

[111] T. Hastie, J. Friedman, and R. Tibshirani, “Model Assessment and Selection,” in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, T. Hastie, J. Friedman, and R. Tibshirani, Eds., New York, NY: Springer New York, 2001, pp. 193–224. doi: 10.1007/978-0-387-21606-5_7.

[112] A. J. Smola, B. Sch¨olkopf, and S. Sch¨olkopf, “A tutorial on support vector regression *,” *Stat Comput*, vol. 14, pp. 199–222, 2004.

[113] S. Raghavendra. N and P. C. Deka, “Support vector machine applications in the field of hydrology: A review,” *Appl Soft Comput*, vol. 19, pp. 372–386, 2014, doi: https://doi.org/10.1016/j.asoc.2014.02.002.

[114] D. A. Augusto and H. J. C. Barbosa, “Symbolic regression via genetic programming,” in *Proceedings - Brazilian Symposium on Neural Networks, SBRN*, IEEE Computer Society, 2000, pp. 173–178. doi: 10.1109/SBRN.2000.889734.

[115] “Self-organizing map.” Accessed: Sep. 26, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Self-organizing_map

Appendix A

The tuned hyper-parameters of the examined algorithms in 3.3.1.2 section.

K-nearest neighbors algorithm:

Table A- 1: Tuned hyper-parameters of k-nearest neighbors algorithm

Target	Neighbors	Weights	Algorithm
SE	2	Distance	Auto
BFE	1	Distance	Auto
UYS3	3	Uniform	Auto
UYS6	4	Distance	Auto
UYS9	2	Uniform	Auto
Hall Flow	6	Distance	

Decision tree algorithm:

Table A- 2: Tuned hyper-parameters of decision tree algorithm

Target	Max Depth	Max Features	Max Leaf Nodes	Min Leaf Samples	Splitter
SE	7	Auto	10	2	Random
BFE	12	Log_2	40	1	Random
UYS3	5	_	50	1	Random
UYS6	1	Log_2	20	2	Random
UYS9	9	Sqrt	70	2	Random
Hall Flow	12	Sqrt	30	1	Random

Random forest algorithm:

Table A- 3: Tuned hyper-parameters of random forest algorithm

Target	Max Depth	Max Leaf Nodes	Estimators	Max Features	Min Samples Leaf
SE	7	20	10	_	1
BFE	11	20	10	Sqrt	1
UYS3	11	None	10	Sqrt	1
UYS6	11	20	10	_	1
UYS9	7	None	10	_	5
Hall Flow	5	None	50	_	1

Gradient boosting algorithm:

Table A- 4: Tuned hyper-parameters of gradient boosting algorithm

Target	Learning Rate	Max Depth	Estimators	Max Features
SE	0.01	1	150	Auto
BFE	0.1	3	150	2
UYS3	0.1	1	150	Auto
UYS6	0.01	1	150	Auto
UYS9	0.1	3	5	2
Hall Flow	0.01	4	100	Sqrt

Appendix B: Machine Learning Algorithms

Support Vector Regression

The SVR foundation is similar to the SVM classification algorithm but modified for regression purposes. With SVR, the best-fitting hyperplane is found using optimum support vectors (SV) while satisfying two constraints: the minimization of prediction error and the minimization of complexity of fitting hyperplane. Another additional parameter in SVR is denoted as ε which demonstrates the distance of hyperplane to SV points located on both sides of the plane (ε tube). Distance of data points outside the margin from ε tube, denoting the slack error (ξ), is the parameter that should be minimized during the loss function minimization, in addition to W (fitting function weights) [112]. A simplified graph of SVR parameters and learning procedure could be seen in Figure B- 1.

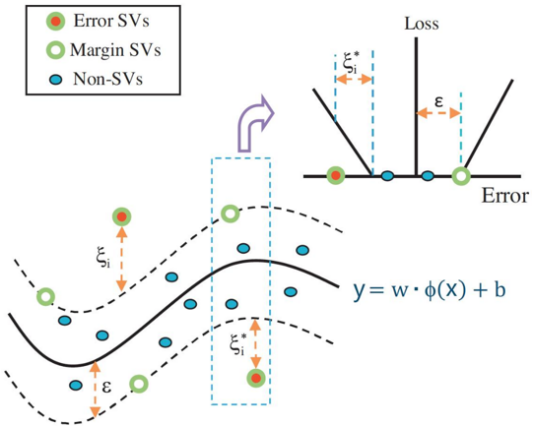


Figure B- 1: Schematics of support vector regressor [113].

A regression margin area depends on ε quantity and should be optimized so that maximum number of datapoints are located in that margin (error-insensitive area), as well as optimizing overfitting and generalization. The trade-off between the prediction error minimization and the regression generalization (sufficient complexity) is tuned by regulation hyperparameter denoted by "C". The mathematical foundation of SVR optimization in linear form is expressed in equation (B- 1) [112].

$$\boldsymbol{minimize}\ \frac{1}{2}\|\boldsymbol{w}\|^2 + \boldsymbol{C}\sum_{i=1}^{n}(\xi_i + \xi_i^*) \qquad \text{(B- 1)}$$

$$\text{Subject to:} \begin{cases} y_i - < w, x > - b \leq \varepsilon + \xi_i \\ < w, x > + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

The SVR application for non-linear problems is enabled by the kernel trick transformation (equation (B- 2)) of data points to a latent space with higher dimension. Radial basis function (RBF) is one of the most practical kernel functions suiting most of the SVR optimization forms [112].

$$K(x, y) = e^{-\gamma \|x-y\|^2} \qquad \gamma > 0 \tag{B- 2}$$

The optimization functions could be computed and solved more conveniently in nonlinear SVR using dual optimization employing Karush-Kuhn-Tucker (KKT) conditions and Lagrange multiplier, which will transform equation (B- 2) to equation (B- 3).

with the inclusion of kernel trick (α_i, α_i^* are Lagrange multipliers) [112]:

$$\begin{aligned} &maximize \left\{ -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*) \right. \\ &\text{Subject to:} \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0 \quad and \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \tag{B- 3}$$

SVR model has multiple hyper-parameters that highly influence the performance of model in fitting an accurate function on train set, and consequently affects the prediction of target for new unseen data. Commonly, hyper-parameters such as the regularization parameter, and SVR gamma/sigma; present in RBF kernel defining the influence radius each support vector has or in other words, the spread and curvature of the decision boundary; are the hyper-parameters needed to be optimized [112].

Genetic Programming's Symbolic Regression

Genetic programming is an AI subcategory which its model training procedure is founded on evolution operations common in genetic science such as mutation. Symbolic regression (GP-SR) is the regressor algorithm working based on genetic programming evolution and training process. GP-SR model start with generation of initial random programs (fitting functions/mathematical expressions) using user-defined set of mathematical operations, constants, mathematical equations, and data variables (inputs and outputs) [114].

These initial programs are represented by tree-like structures. The root node is divided into two branches which will be one of the following two forms:

1. Terminal node (final node of branches in the tree): these nodes are either the variables/predictors (in regression models) of the predefined dataset, or the user-defined constant values of the search space.
2. Operation node: same as the root node, these nodes also are mathematical operations that are user-defined mathematical equations.

Each tree is a 'program' which represents a mathematical expression created by the GP-SR model in the search space (set of infinite number of mathematical expressions that could be generated using a user-defined set of mathematical operations, mathematical equations, and dataset variables). A generated tree-structured program and a graphical review of the GP-SR process is demonstrated in Figure B- 2. The final generated expression will be retrieved from the tree starting from root node moving to the child nodes in the left side and right side, respectively [114].

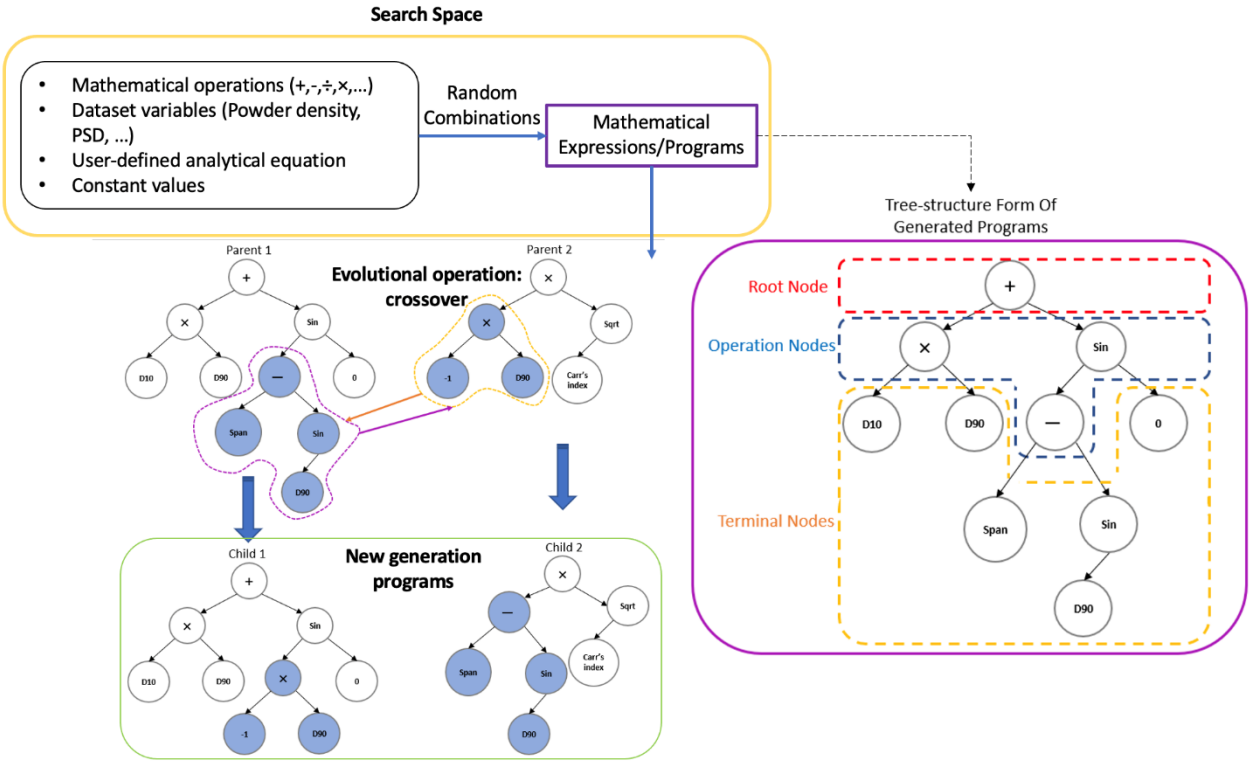


Figure B- 2: Schematic of tree structure of a randomly generated program in genetic programming.

During the GP-SR model training, infinite rounds of 'generations' could be created. Each generation contains *n* (*n* should be defined by user along with number of rounds of generation) number of programs that are represented in tree structures, as stated before. Each of the programs in a generation are referred

to as ‘parent’ programs. a subset of programs in each generation are then chosen randomly to be tested against each other in ‘tournament’. The program with highest fitness is chosen as winner to be passed to the next generation. Genetic evolution operations such as mutation and cross-over are then applied on one (mutation) or two (cross-over) tournament winner programs (parents) to create the superior programs of the next generation (child). Cross-over replaces a random subtree of two winner programs (parent 1 and parent 2) with each other to generate two new programs of the next generation (child 1 and child 2) [114], as depicted in Figure B- 2.

During the mutation, a random subtree of a parent program is substituted by a randomly generated subtree (not a tournament winner) to create a new child program. The evolution operations make the fitting functions of each generation more accurate than previous generations until convergence to the fittest mathematical expression (program) is achieved [114].

Regression using the GP-SR technique seeks to identify and derive the underlying structure of a dataset without making any prior assumptions regarding the mathematical expression of the data fitting model. It retrieves the complex underlaying pattern of the data with traversing various mathematical fitting functions instead of optimization of coefficients of an initial function generated based on model priori-assumption. The fitness of each generated population/program is evaluated by defined error metrics such as mean squared error, mean absolute error etc. Both high estimation accuracy and equation simplicity in the resulting mathematical fitting equation are required to be fulfilled in the optimization process. In the GP-SR training, the search for the best fitting program is only restricted by the termination/optimization condition provided by the user [114].

Self-Organizing Maps

SOM neural network is an unsupervised clustering artificial neural network (ANN) comprised of several nodes as in the other conventional neural networks. The difference between other ANN algorithms and SOM is in training and cost/error function optimization. SOM training operates based on competition while other algorithms’ optimization is mainly based on back-propagating stochastic gradient descent. Schematics of SOM training is depicted in Figure B- 3.

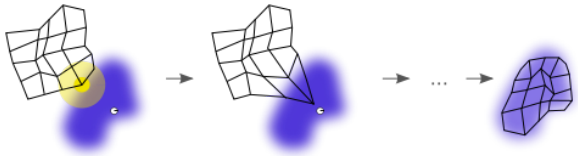


Figure B- 3: Self-Organizing Maps Training Process [115].

The SOM nodes have weight vectors; with the same dimension as the input data; which are randomly initialized. The clustering of dataset happens by modification of nodes weight values. In general, the process could be summarized in three steps of competition, cooperation, and adaptation. In competition step, the winner node for each of the input data points is identified based on the distance between the node weight values and the input data points (usually calculated by Euclidean distance metric). The node with lowest distance to each data point is known as the data point's winner node. In cooperation, the neighboring nodes to the winner node which have a high similarity with it (low distance) are identified using the neighborhood kernel function ($h_{ik}(t)$) shown in equation (B- 4). The winning node and neighborhood will be chosen based on distance (d_{ik}) and time (t), but the size of neighborhood (calculated by the neighborhood size decay parameter ($\sigma(t)$)) gradually decreases after each iteration of model optimization. This process is for narrowing down to the accurate clusters of data points with authentic similarities [107].

$$h_{ik}(t) = exp^{(-\frac{d_{ik}^2}{2\sigma^2(t)})} \qquad \sigma(t) = \sigma_0 exp(-\frac{t}{T_2}) \qquad \text{(B- 4)}$$

The last step of adaptation consists of modification of node weights (W_k) using equation (B- 5). The winner nodes and the winning neighboring nodes are updated more significantly than the nodes in far distances from the samples (X^n), but the scale of modification in the winning neighbor nodes is also dependent on their distance to their winner node. The modification of the weights is regulated with $\eta(t)$, denoted as learning rate decay parameter shown in equation (B- 5) [107].

$$W_k = W_k + \eta(t).h_{ik}(t).(X^n - W_k) \qquad \eta(t) = \eta_0 \exp(-\frac{t}{T_1}) \qquad \text{(B- 5)}$$

The weight optimization continues until no further changes are seen in winning nodes and then the final map of clustered data set could be generated using the final weight values of the nodes [107].

Appendix C

Two information graphs for 3D plots ((a–d)) are provided which map the information of 3D plots in (1) cubical parts and (2) cylindrical parts. The average intensity level per layer is plotted for cubical and cylindrical parts with the highest and lowest surface roughness/density in three powder batches of 15–53 µm (A), 15–106 µm (B), and 45–106 µm (C).

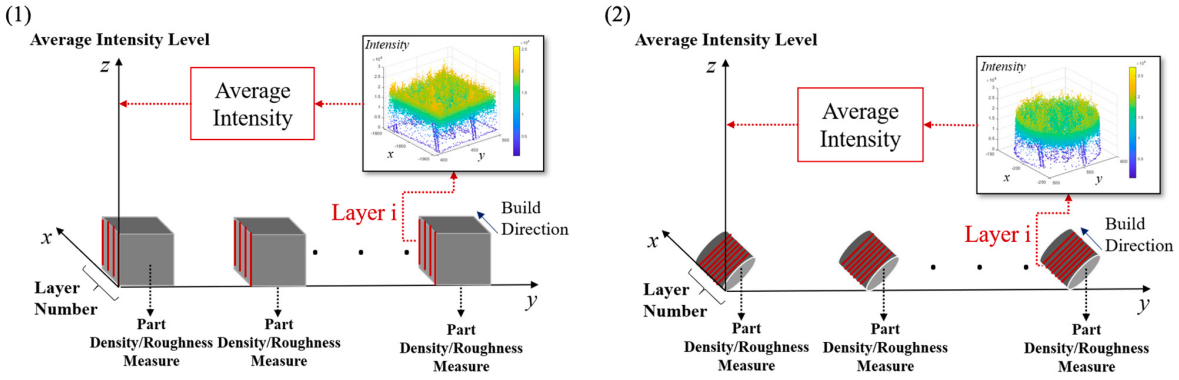


Figure C- 1: Information Map of Figure 4-5.