INTERPRETABLE METHODS FOR QUANTITATIVE MEASUREMENT AND CLASSIFICATION OF SURFACE TOPOGRAPHY

by

Jesse Redford

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Mechanical Engineering

Charlotte

2024

Approved by:

Dr. Brigid Mullany

Dr. Jason Fox

Dr. Angela Allen

Dr. Edward Morse

Dr. Jeff Kimble

ABSTRACT

JESSE REDFORD. Interpretable Methods for Quantitative Measurement and Classification of Surface Topography. (Under the direction of DR. BRIGID MULLANY)

The functionality of manufactured components is intricately linked to their surface topography, a characteristic profoundly shaped by the fabrication process. Repeatable quantitative characterization of surfaces is essential for detecting variations, defects, and predicting performance. However, the plethora of surface descriptors presents challenges in optimal selection of the correct assessment metric. This work addresses two of these aspects: automatic selection of surface descriptors for classification and an application-specific approach targeting scan path strategies in laser-based powder bed fusion (LPBF) additive manufacturing. A framework, titled Surface Quality and Inspection Descriptors (SQuID^{TM}), was developed and shown to provide an effective systematic approach for identifying surface descriptions capable of classifying textures based on process or user-defined differences. Using a form of univariate analysis rooted in signal detection theory, the predictive capability of a discriminability value, d', is demonstrated in the classification of mutually exclusive surface states. A discrimination matrix that offers a robust feature selection algorithm for multiclass classification challenges is also introduced. The generality of the approach is validated on two datasets. The first is the open-source Northeastern University dataset consisting of intensity images from six different surface classes commonly found in rolled steel strip operations. The application of signal detection theory's measure, d', proved successful in quantifying a texture parameter's ability to discriminate between surfaces, even amidst violations of normality and equal variance assumptions regarding the data. To further validate the approach, SQuID^{TM} is leveraged to classify different grades of surface finish appearances. ISO 25178-2 areal surface metrics extracted from bandpass filtered measurements of a set of ten visual smoothness standards obtained from low magnification coherent scanning interferometry are used to quantify different grades of powder-coated surface finish. The highest classification accuracy is achieved using only five multi-scale descriptions of the surface determined by the SQuID^{TM} selection algorithm. In this case, spatial and hybrid parameters were selected over commonly prescribed height parameters such as Sa, which proved ineffective in characterizing differences between the surface grades.

Expanding surface metrology capabilities into LPBF additive manufacturing, additional studies developed a methodology to comprehend the relationship between scanning strategies, interlayer residual heat effects, and atypical surface topography formation. Using a single process-informed surface measurement, a critical cooling constant is derived to link surface topography signatures directly to process conditions that can be calculated before part fabrication. Twelve samples were manufactured and measured to validate the approach. Results indicate that the methodology enables accurate isolation of areas within the parts known to elicit heterogeneity in microstructure and surface topography due to overheating. This approach provides not only a new surface measurement technique but also a scalable parameterization of LPBF scan strategies to quantify track-to-track process conditions. The methodology demonstrates a powerful application of surface texture metrology to characterize LPBF surface quality and predict process outcomes.

Overall, this dissertation contributes a systematic approach for identifying discriminatory parameters for surface classification and a novel process-informed surface measurement for predicting track-scale overheating during LPBF-AM of a nickel superalloy.

DEDICATION

I dedicate this dissertation to Kent Tarbutton. You are dearly missed. I also dedicate this work to my family and all those who have continued to support me through this journey. Thank you, Katie, for your constant support and patience. Kent, Shelley, Mary, and Tim-I wouldn't be here without you. My sisters, Chelsea, Paige, and Megan; my brothers, Lennon and John; as well as my nieces, Quinn, Emma, Soren, Nikki, and Kelley, have all been pillars of support. I want to thank the University of North Carolina at Charlotte and the Center for Precision Metrology for creating an environment that fosters collaboration and innovation. Also, to the wonderful teachers and mentors who have helped me develop my skills, patience, and persistence to complete this document. Cheers!

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my advisor, Brigid Mullany, who introduced me to the fascinating world of surface metrology. I deeply appreciate her continued support in my professional development, her generous investment of time, and her patience. I am also grateful to Jason Fox and Chris Evans for their mentorship, contributions, and support. Special thanks to Edward Morse for his keen insight and fortitude, and to Angela Allen for her readiness to brainstorm ideas and offer help. I would like to acknowledge Jimmie Miller and Stuart Smith for their mentorship, engaging lectures, and valuable feedback throughout my journey. I am thankful for my office mates and colleagues: Dhanooj, Jacob, Rehab, Sevda, Kumar, Raj, Jorian, and Anand, whose camaraderie and collaboration have enriched my experience. Special thanks to all the affiliate members and the Center for Precision Metrology for their invaluable insights, feedback, and support of the projects that contributed to this work. My gratitude extends to the National Institute of Standards and Technology for providing resources, and to all the mentors, including Jordan Weaver and Brandon Lane, for their guidance and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

d'	dee-prime
CSI	Coherence Scanning Interferometry
FV	Focus Variation
HVI	Human Visual Inspection
SDT	Signal Detection Theory

PREFACE

> *"Measurement is the first step that leads to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it."*

-H. James Harrington

CHAPTER 1: INTRODUCTION

1.1 Motivation

Surface texture metrology is a fundamental tool for improving manufacturing processes. Within any manufacturing undertaking, the resulting surface texture necessitates measurement to ascertain its suitability for the intended application. However, the measurement of surface texture encompasses numerous facets, including isotropy, average deviation of heights, and more. Yet, no single solution can often ensure the requisite performance for an application, whether it involves quantitatively discriminating between defect types, evaluating visual grades of aesthetics, or discerning between flawed and ordinary surface structures emerging during manufacturing processes. Ideally, metrics aligned with functional aspects such as aesthetics, mechanical properties, or process conditions are preferred and given precedence. Nonetheless, the relevance of a surface metric can vary depending on the context, whether it pertains to functional or processing considerations of interest. Navigating the parameter space, selecting appropriate measurement techniques, and establishing thresholds for comparing measurement results to categorize surfaces or their distinctive features into process or user-defined groups can be a perplexing endeavor. This complexity arises from the multitude of approaches available for data analysis, algorithms, and parameter space navigation. This dissertation is motivated by two primary aspects: firstly, the objective determination of relevant descriptors to classify surfaces categorized by qualitative and process-related differences. Secondly, to explore previously uncharted applications of surface texture metrology in the characterization of Laser Powder Bed Fusion (LPBF) additive manufacturing processing.

1.2 Aims and objectives

The first aim of the dissertation is to develop a general method for identifying relevant parameters for surface classification. A metric from signal detection theory, d', will be used to assess a surface parameter's ability to enable the correct classification of mutually exclusive surface states. The objective is to create a framework that minimizes user input, and maximizes transparency. The second aim of this dissertation is to expand surface metrology capabilities into Laser Powder Bed Fusion (LPBF) manufacturing; additional studies aim to develop a methodology to comprehend the relationship between scanning strategies, interlayer residual heat effects, and atypical surface topography formation. Furthermore, the aim will be to assess if the methodology enables accurate isolation of areas within the parts that elicit heterogeneity in microstructure and surface topography due to overheating.

In summary, the aims and objectives of this dissertation include:

- To develop, implement, and assess the capability of an interpretable systematic framework for identifying parametric surface descriptions capable of classifying surfaces based on process or user-defined differences.
- Quantify the ability of signal detection theory's measure, d', to assess texture parameter discrimination abilities across various surfaces.
- Develop, implement, and assess a surface metrology-based framework to predict melt pool-surface topography formation in the laser powder bed fusion process.

1.3 Structure of dissertation

This dissertation is structured around four distinct papers, each contributing to the practical and theoretical advancement of surface texture metrology. In this chapter the motivation of the dissertation is presented alongside a breakdown of the aims and objectives. In Chapter 2, a review of relevant literature is provided to inform the research conducted throughout this dissertation. This chapter also contextualizes the motivation and contributions in the broader context of surface texture characterization. Chapter 3 focuses on the development and validation of an algorithmic methodology known as SQuID^{TM} (surface quality and inspection descriptors) [6]. The methodology aims to assess, select, and utilize the best surface metrics for the classification of surface topography according to user or process-defined groupings. The findings of this work have been published in the Journal of Manufacturing Systems and are included in Chapter 3. The citation for the article is given below:

- J. Redford, and B. Mullany. "Construction of a multi-class discrimination matrix and systematic selection of areal texture parameters for quantitative surface and defect classification." *Journal of Manufacturing Systems* vol. 71, pp. 131-143, 2023. doi: doi.org/10.1016/j.jmsy.2023.08.002

Building upon the foundation established in Chapter 3, Chapter 4 extends the research by addressing the multiscale aspects of surface topography evaluation. Here, SQuID^{TM}, low magnification coherence scanning interferometry, and selective bandpass filtering are employed to downselect multi-scale areal texture parameters, enabling the classification of visual smoothness standards [7]. This work has been published in the Journal of Materials and the full citation is shown below.

- J. Redford, and B. Mullany. "Classification of Visual Smoothness Standards Using Multi-Scale Areal Texture Parameters and Low-Magnification Coherence Scanning Interferometry. " *Journal of Materials - Manufacturing Processes and Systems Special Edition on Surface Inspection and Description in Metrology and Tribology (Volume II)*, vol. 2, p. 1653, 2024. doi: doi.org/10.3390/ma17071653

Chapter 5 ventures beyond statistical surface texture characterization based on the ISO 25178-2 parameters and is focused on the basement of surfaces generated via

LPBF, specifically. This paper demonstrated a method to quantitively measure and analytically predict the formation of enlarged melt pools in a series of single-layer trapezoidal geometries. Localization of the overheated regions in these samples was achieved by determining a critical cooling constant from a single process-informed surface topography measurement [8]. This work was presented at the 2023 joint special interest group meeting between the European Society for Precision Engineering and Nanotechnology (EUSPEN) and the American Society for Precision Engineering (ASPE) on advancing precision in additive manufacturing and is published in the proceedings. The citation is given below.

- J. Redford, J. Fox, C. Evans, B. Mullany, A. Allen, and E. Morse. “Centerline-time functions and critical constants for predicting laser powder bed fusion melt pool distortion using one surface topography measurement." *Proceedings of Joint Special Interest Group meeting between euspen and ASPE Advancing Precision in Additive Manufacturing, Leuven, BE.*, 2023. link: `https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=956293`

Chapter 6 builds upon the methodology in Chapter 5 and provides further validation of the approach. Examples of how the underlying methodology and derived critical cooling constant translate to completely different sample geometries produced with the same machine, alloy, and process parameters are provided. The developed pointwise parameterization of the scan strategy effectively identified areas of the part expected to elicit heterogeneity of microstructure and surface topography. The contents of Chapter 6 have been compiled into a paper that will be submitted to the upcoming 2024 joint special interest group meeting between EUSPEN and ASPE on advancing precision in additive manufacturing.

- J. Redford, J. Fox, C. Evans. “A critical cooling constant for elucidation of systemic overheating in laser powder bed fusion additive manufacturing of IN625."

Joint Special Interest Group meeting between euspen and ASPE Advancing Precision in Additive Manufacturing, Golden, USA., 2024.

Chapter 7 covers the dissertation's overall conclusions, a summary of findings, and contributions to the literature.

CHAPTER 2: PRELIMINARIES AND BACKGROUND

This chapter provides an overview of the relevance of surfaces and texture metrology in manufacturing environments and the challenges in selecting metrics that can enable functional characterizations or classifications of surface properties to improve manufacturing processes and the quality of products. An overview of the approach for classifying surface textures and state-of-the-art characterization of laser powder bed fusion surfaces is also discussed.

2.1 Relevance of surfaces and texture metrology

A material's surface constitutes its outermost layer, extending to include nearby material depending on the context. This outer boundary encompasses the outermost atoms. Surface topography defines the geometry of this boundary. The intricate shape of the interface between a material and its surroundings significantly influences its surface properties relating to various functions such as corrosion resistance, adhesion, cleanliness, aesthetics, optical properties etc. Applications of surface texture metrology have been around for over 110 years; for example, in 1912, Binder et al. determined that the electrical conductivity of rough surfaces was lower than expected. The decrease in conductivity was attributed to a reduction in contact area due to surface roughness [9]. Discoveries such as these have prompted significant efforts to understand topography-dependent properties better [10]. Widely available tactile and optical technologies now provide fast digital reconstructions of surface topography and are being incorporated into many manufacturing environments. Surface properties are pivotal in various scenarios: determining whether paints and coatings adhere effectively to consumer products, influencing the efficiency of energy utilization in engines of automobiles and airplanes, affecting the durability of cutting tools, shaping the biocompatibility of medical devices, and even impacting the safety of flooring tiles to prevent slip-and-fall incidents [10].

Much has been said about Industry 4.0 and the transition to zero-defect manufacturing. One area that is often overlooked is the role that surface texture metrology can play in this transition. From a quality control perspective, it is essential to distinguish between surfaces that perform and were fabricated, modified, or treated differently [11]. Common net-shape manufacturing processes produce parts that require additional post-processing (i.e., polishing, lapping, honing, grinding, etc) to obtain desired surface finishes. Surface texture parameters facilitate the control of surfaces by assigning a quantitative value calculated via a series of mathematical operations [1]. For instance, the single-valued parameters listed in the ISO 25178-2 [12] standard provide a way to measure surface finish in a holistic way, considering not just the roughness of the surface but also other factors such as spatial periodicity and homogeneity. By measuring these factors at appropriate scales, manufacturers can more accurately assess the quality of their surfaces and make necessary adjustments to their processes. This, in turn, can lead to reduced costs and increased efficiency. However, to be useful, the selected parameter(s) should be function-related [13, 14, 15] manufacture-related, and possess low sensitivity to measurement errors [16, 17]. Figure 2.1 illustrates the link between surface texture and its characterization, specification, processing, and function.

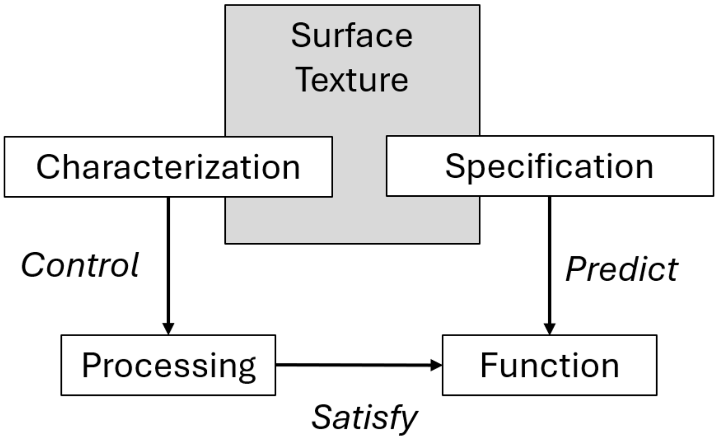


Figure 2.1: Illustration of the link between surface texture, processing, and function. Adopted from [1] [2].

2.2 Quantification of surfaces

The literature on surface quantification encompasses a wide range of analytical techniques and terminologies. The following sections offer a brief overview of different metrics used in surface metrology and image processing communities for characterizing surfaces. For further details and examples, readers are directed to [18, 19, 20, 21, 22] for insights into tribological research, including friction squeal. Comprehensive reviews of standard areal texture parameters for describing wear, friction, and fatigue are available in [16, 17]. Quantitative characteristics are discussed in [23, 24]. Aspects of surface texture's influence on design, substrate performance, and bonded joints are explored in [25, 26, 27]. An overview of different surface description parameters can be found in [28, 29, 30, 31, 32], with feature parameters detailed in [33]. For a broader understanding of natural surfaces and digital image texture representations, refer to [34, 35, 36, 37, 38, 39, 40].

2.2.1 Surface texture and ISO 25178-2 parameters

Surface texture refers to the small-scale features that comprise a surface but not those that contribute to the form or overall shape of the part [41]. The ISO 25178-2 standard [12] specifies terms, definitions, and surface texture parameters for characterizing 2.5D areal representations of surface topography. This standard provides areal parameters that play a crucial role in characterizing surface topography and are used in various scientific, industrial, and measurement applications. Areal parameters range from amplitude to spatial, functional, and hybrid parameters, such as Sa, Str, Vvv, and Sdr, respectively. Each parameter is designed to capture different characteristics of the surface. Height parameters include the arithmetic mean of the surface deviations Sa and second-order statistics Sq, equivalent to the root-mean-square (RMS) of the surface, in addition to Skewness, Ssk, and kurtosis Sku. The former provides insight into the sharpness of the roughness profile, while the latter presents the degree of bias of the roughness shape with $Sku > 3$, indicating the height distribution are spiked. An $Ssk = 0$ indicates that the height distribution of (peaks and pits) are symmetrical about the mean plane. Hybrid parameters consider the local slopes of the topography and comprise two parameters. The developed interfacial area ratio Sdr describes the relative increase in additional surface area due to texture compared to the planar definition area of the surface. Hence, for a perfectly flat surface, Sdr would equal zero. The RMS gradient of the surface Sdq considers the averages of the local slopes along the x and y sample directions at all points within the definition area. Again, Sdq would equal zero for a perfectly flat surface, and becomes larger when surface slopes increase. The auto-correlation length Sal represents the horizontal distance in the direction where the auto-correlation function (ACF) decays to a fixed value (0.2 by default) the fastest. The texture aspect ratio Str is a measure of surface texture uniformity and is calculated by dividing Sal by the horizontal distance in the direction of the slowest decay of ACF values [42]. Str is a great example of a parameter that can be intuitively linked to the structure of a surface. Str describes the degree of anisotropy on a scale of 0 to 1, with one indicating a spatially isotropic texture and zero indicating a dominant lay pattern on the surface. The parameter can often confirm what may appear qualitatively to look like structures or patterns on a surface that can be difficult to describe with words. For additional details of ISO parameters, readers are referred to the standard itself [12] or works by [43, 16, 17, 44], and an industrial survey by [45].

2.2.2 Image processing and intensity metrics

It is not practical to provide an exhaustive survey of all texture features used in the image processing community; this section concentrates on techniques that are widely used in texture analysis. Readers are referred to texture analysis survey papers [40, 46, 47, 48, 49, 50, 51] for more information. However, a general overview of different methods is provided in this section. It should be noted that because many image processing techniques are based on intensity representations of the surface, the physical interpretation and meaning of the metrics concerning the actual surface can be challenging.

Image processing metrics provide quantitative measures for analyzing digital images and discerning their characteristics. Unlike height maps, where surface texture parameters are utilized, image processing metrics focus on intensity-based representations of surfaces. For example, contrast measures brightness discrepancies between adjacent pixels, while entropy captures the unpredictability of pixel intensity distribution, both revealing aspects of image complexity. Edge sharpness evaluates the clarity of transitions between intensity regions, emphasizing defined edges within the image. Texture features typically highlight spatial patterns of pixel intensities, offering insights into surface texture and structure. Histogram-based metrics, including mean, standard deviation, skewness, and kurtosis, derive statistical information from pixel intensity histograms, with some being akin to standard height parameters in ISO 25178-2 standard. Spatial graylevel co-occurrence matrices (GLCMs) [52] are one of the most well-known and widely used texture features. GLCMs produce a matrix that captures the frequency of occurrence of different combinations of gray levels at specified spatial relationships. However, the results of GLCM analysis can be sensitive to the parameters chosen during computation; there is generally no accepted solution for choosing the optimal displacement vector [40]. Local binary patterns (LBPs) transform each pixel to a zero or one based on comparisons with neighboring pixels to create binary patterns. These patterns are then used to describe the texture of the image. However, because LBP focus on local texture patterns within a neighborhood of pixels, they fail to capture long-range spatial relationships in the image [40]. Filter-based approaches typically consist of applying filter banks on an image and characterizing some aspect of the filter response(s). The methods can be divided into spatial, frequency, and joint spatial-frequency domain techniques. Filter methods are generally used to extract edges, lines, isolated dots, etc. In the spatial domain Sobel, Canny, and Laplacian are common, while frequency methods typically employ Fourier transforms. Measuring edge strength and edge frequency is mentioned as being one of the earliest attempts to discriminate different textures [40]. Fractals were originally proposed by Mandelbrot [53] as geometric primitives that are self-similar and irregular in nature [53]. A fractal dimension serves as a measure of complexity and irregularity, and lacunarity measures the structural variation or inhomogeneity [40].

2.2.3 Parameter versus feature-based quantification

Feature-based characterization involves extracting higher-level descriptors from surface data to capture complex surface patterns comprehensively. These features may include texture histograms, fractal dimensions, and wavelet transforms. By considering a broader range of surface characteristics, feature-based methods can enhance discrimination capabilities and capture intricate surface details that parameter-based approaches may miss. Feature parameters can also imply calculation on one or more unique region(s) or instance(s) found on the surface. Shape descriptions such as volume, area, sphericity, etc, of unique instances are commonly used in the image processing community. Areal ISO-25178-2 parameters can still be computed, but a possible "feature-based" approach would first segment the surface into regions, compute ISO parameters for each region, and provide descriptive statistics of the results (i.e., mean, std, max, etc). However, for feature based approaches the parameter values depend highly on how the surface is segmented and features isolated. Therefore, a greater emphasis on segmenting the features of interest from the surface should typically be prioritized before solving the problem of what parameters should be used to characterize the segmented regions. Segmentation of the surface is critical to applying ISO parameters on a macro scale. Superficial edges or unrelated surfaces within the field of view render the measurement area incapable of being assessed with traditional ISO parameters. Work by [54] highlights an excellent example of a geometric-based approach, using dimensional measurands and functional ISO parameters to characterize critical dimensions and the bearing area of micro-dimpled samples. Demonstrating how a well-tuned segmentation method provides a consistent way to extract and assess the functional features of microstructured surfaces [54]. However, selecting appropriate features requires domain expertise and careful consideration of the specific characteristics of the surface texture data being analyzed [52, 55].

2.2.4 Challenges in identifying appropriate metrics for quantification

Challenges in identifying the right metric for quantifying surface texture are generally amplified by the fact that there are simply too many metrics and length scales that could be applied. Although the characterization of surfaces using texture parameters aims to provide a parametric description that enables the control of processing or the prediction of surface performance, the relevance of the parameters used in this description is paramount [1, 2, 15, 56]. A common challenge in applying surface texture metrology to new or existing applications lies in selecting the most appropriate parameters for specific applications and determining threshold values that signify meaningful changes in surface conditions or process states. Researchers have grappled with this issue. As Brown et al. (2018) noted, "The lack of a system of characterizations, analyses, and methods has handicapped research on topographically related phenomena" [57]. Ideally, parameters should enable the ability to systematically design, impart, and measure surface topography to facilitate rational control surface properties [10]. The ability to do so would enable manufacturers to overcome many challenges such as defining thresholds for texture sensation and tolerance in relation to customer expectations and satisfaction [58]. This example application is exceptionally challenging because aesthetic preferences can be subjective, and therefore, there may be no sure way of determining which metrics will be most relevant for translating qualitative groups of surfaces to quantitative ranges. "Relevant" parameters can vary depending on the context, encompassing discriminatory power between surfaces processed differently or precision in measuring specific surface properties such as oil retention [1]. One texture parameter may be more sensitive to the textural differences between two cylinder bores but does not characterize the property of interest (i.e., the amount of oil retained at the surface). Numerous case studies and reviews demonstrate successful applications of topography measurements in discriminating surfaces, controlling processes, or predicting functionality [59, 16, 17, 60]. However, a systematic and comprehensive approach to parameter selection is often lacking [61, 62]. The following section provides an overview of feature selection and classification algorithms commonly used in texture analysis.

2.3 Parameter selection for classification approaches

In binary classification, there are four possible prediction outcomes; the goal of feature selection methods is generally to minimize or maximize some metric that considers the relationship between these four outcomes: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). True positives (TP) represent instances correctly predicted as positive, true negatives (TN) are instances correctly predicted as negative, false positives (FP) are instances incorrectly predicted as positive, and false negatives (FN) are instances incorrectly predicted as negative. Feature selection for classification aims to identify the most relevant features that optimize these metrics, ultimately improving the model's performance in correctly classifying instances into their respective categories.

2.3.1 Feature selection methods

Feature selection for classification approaches involves choosing the most relevant features (i.e., metric, parameter, etc.) from a dataset to improve model performance and interpretability. Selection algorithms are crucial in identifying relevant surface parameters for effective classification. These algorithms aim to balance reducing the dimensionality of the feature space and preserving discriminatory power. Different approaches offer varying trade-offs in terms of computational complexity, robustness, and interoperability. Feature selection algorithms for classification can be divided into three groups: filter, wrapper, and embedded methods illustrated in Figure 2.2. This section will provide a brief overview of each.

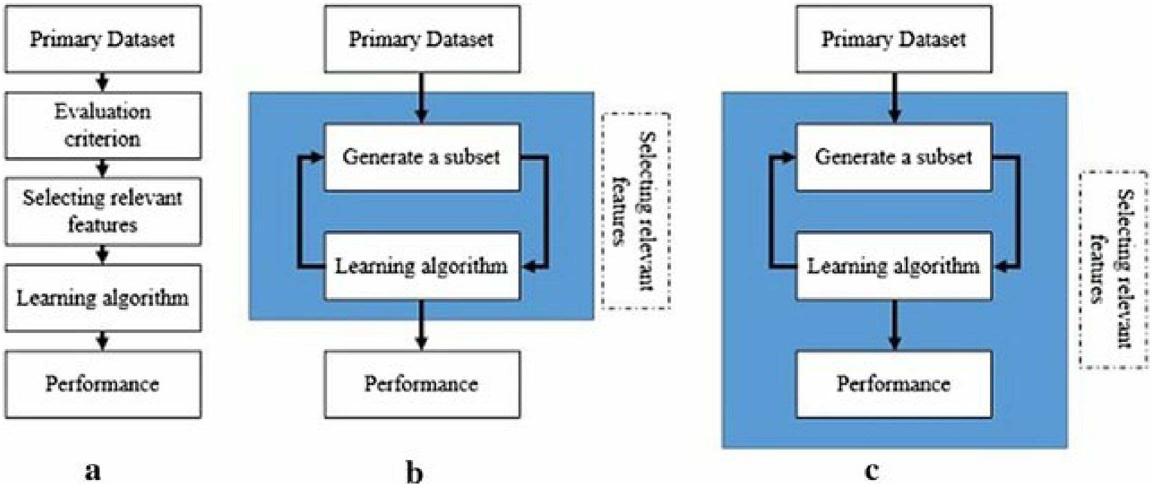


Figure 2.2: Illustration of (a) filter methods, (b) wrapper methods, and (c) embedded methods adopted from [3].

Filter methods evaluate the relevance of features independently of the classification model. They rely solely on statistical measures or domain knowledge to assess the importance of features. This independence makes it easier to understand the criteria used for feature selection, as it's not influenced by the complexity of the classification algorithm or the specific dataset. Since filter methods often employ straightforward criteria for feature selection, such as correlation coefficients, mutual information scores, or statistical tests. These criteria are typically easy to interpret and explain to stakeholders. For example, a high mutual information score indicates a strong dependency between a feature and the target variable, making it clear why that feature was selected. This makes the rationale behind feature selection more intuitive and easier to understand. Although filter methods are generally computationally efficient and independent of the classifier of choice, they may overlook interaction effects between features.

Wrapper methods evaluate subsets of features by training a classification model on different combinations of features and selecting the subset that yields the best performance based on a predefined evaluation metric (e.g., accuracy, F1-score, etc) [63]. Wrapper methods can lead to feature combinations that optimize model performance but may not necessarily be easily interpretable. This is because a classification model is involved in the selection process; thus, understanding why certain features are selected also requires understanding the intricacies of the model's decision-making process. The interactions between features and their impact on model performance may not be straightforward or easily explainable. This approach sometimes offers better classification accuracy but at a higher computational cost.

Embedded methods integrate feature selection directly into the learning algorithm, optimizing feature subsets during model training. Like wrapper methods, embedded methods often involve complex classification models such as support vector machines (SVMs), random forests, or neural networks. They modify the model's parameters to penalize or eliminate less relevant features, aiming to improve model performance while simultaneously selecting informative features. The integration of feature selection with model training makes it difficult to separate the impact of the model from the feature selection process. While embedded methods may result in more parsimonious models by removing irrelevant features, the exact criteria for feature selection can be less transparent.

2.3.2 Classification algorithms

Classification algorithms are essential for distinguishing between different surface textures based on selected parameters. Traditional approaches such as decision trees, support vector machines (SVM), and k-nearest neighbors (k-NN) have been widely used in surface texture classification tasks [64]. Each algorithm has its strengths and weaknesses in terms of classification accuracy, scalability, and interoperability. Decision trees are easy to interpret and visualize, making them suitable for understanding the decision-making process [65]. However, they may suffer from overfitting and lack robustness when dealing with noisy data. SVMs excel in handling high-dimensional data and effectively capture complex relationships between features. Still, they may require careful selection of hyperparameters (i.e., kernel type, regularization parameter, etc) and struggle with large datasets. KNN classifiers are simple and intuitive but can be computationally expensive, particularly with increasing dataset size [66]. Neural networks and deep learning architectures, such as convolutional neural networks (CNN), have shown promising results in surface texture classification tasks. CNNs can automatically learn hierarchical representations of surface features from raw data, eliminating the need for handcrafted feature extraction. However, CNNs generally require a large amount of labeled training data to learn effectively and require more computational resources for training [67]. Figure 2.3 provides a visual illustration of different machine/deep learning algorithms used in classification tasks.

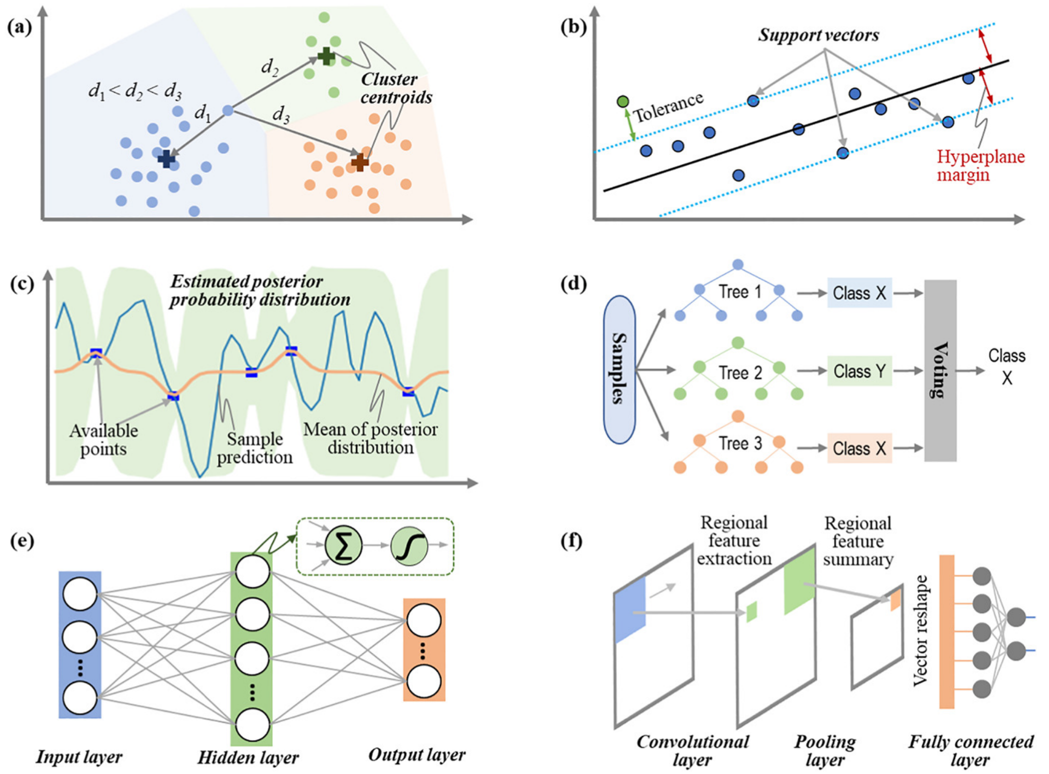


Figure 2.3: Illustration of different classifcation algorithms (a) K-means clustering, (b) support vector machines (c) Gaussian process (d) random forest, (e) artificial neural networks, and (f) convolutional neural network adopted from [4].

2.3.3 Examples of improving or identifying parameters for classification

A variety of works can be found on the topic of the identification of relevant parameters for characterizing surface texture. This section provides a few examples of different approaches for improving and/or selecting parameters for classification.

Kacalak et al. (2021) suggested multiple methods for evaluating the classification ability of parameters considering the geometric and harmonic mean, variance, etc., of normalized parameter increments calculated on one height map of lapped and electrochemical machined bearing steel, titanium alloy, and ceramic surfaces [68]. Several parameters are selected based on one or a combination of these classification ability measures. However, no assessment of classifier performance was provided, nor was there an intuitive explanation of what the classification ability metric(s) values indicate in terms of performance expectations other than relative notions such as higher is better. Works by Lopez et al. (2023) presented a more systematic methodology, including evaluating relevant roughness parameters and surface wettability [69]. The authors use the degree of divergence (DDk) to isolate parameters to quantify the surface differences calculated according to Equation 2.1:

$$DD_k = \frac{\sum \mu_{i,A}(k) - \sum \mu_{i,B}(k)}{\sum \sigma_{i,A}(k) - \sum \sigma_{i,B}(k)} \tag{2.1}$$

where DDk refers to the degree of divergence of each roughness parameter with a mean value $\mu(i)$ (i denotes the area of each image) and a standard deviation $\sigma(i)$ for two types of surface A or B that were processed or perform differently. When DDk is higher than or equal to two, roughness parameters for the two sample types are distinguished with a 95% reliability, meaning that the parameter has high relevance for further description. However, if the DDk is greater than or equal to 1, the roughness parameters are distinguished with 68% reliability. The authors claim that by analyzing the DDk values, the most relevant roughness parameters to describe the influence of the different manufacturing methods can be identified [69]. However, roughness parameters with a DDk higher than or equal to two are selected as relevant parameters. Hence, the user must make an arbitrary setting.

Jordan et al. (2006) suggested the F-test as a quantitative method for comparing the efficacy of texture characterization methods. However, the analysis was limited to only a few of surface types and parameters [62]. Singh et al. (2024) demonstrated surface roughness characterization using representative elementary area (REA) analysis on differently polished quartz and glass surfaces, highlighting the impact of sampling interval on ISO 25178-2 parameter values [70]. The authors identified representative scales by determining a sampling interval threshold beyond which parameter values stabilized. This increases the possible predictive power of parameters for classification but does not inherently address the selection issues. Senge et al. (2022) extended conventional ISO 25178-2 surface texture parameters to shot peen surfaces characterizing coverage; however, the study was limited to only height and hybrid parameters [71]. Caravella et al. (2022) examined CuCrZr-manufactured cubic specimens fabricated from selective laser melting (SLM) using only height parameters. Their findings suggested correlations between laser scanning speed, layer thickness, and surface roughness, but these observations were constrained by the narrow scope of parameters analyzed [72].

Bigerelle et al. (2008) employed multiscale statistical analysis to discriminate roughness in surfaces produced by injection molding. The highest value of a discriminant function based on the Fisher variate is used to compare differences in parameter values calculated at difference scales [61]. This study addresses aspects of increasing the classification ability parameters but not selecting them. This is likely because of the complexity involved with systematically evaluating more than a few parameters. The software application Mesrug^{TM} was introduced by Bigerelle et al. (2011) attempted to solve both of these problems (i.e., relevant scale and parameter) simultaneously. An experimental design is created to generate surfaces, and the inquiry pertains to the influence of process parameters on surface topography. The authors reformulated the question in surface topography formalism: What are the best roughness parameters, evaluated on an appropriate scale, to better discern the effect of the process on surface topography? The Mesrug^{TM} system systematically selects parameters and appropriate length scales using the Fisher variate F to quantify the relevance [73]. Fisher discriminant analysis involves comparing the ratio of between-class variance to within-class variance. However, caution must be exercised when applying these scores to assess parameters across three or more surface types. In the context of analysis of variance (ANOVA) for multiple groups, rejecting the null hypothesis and finding a statistically significant difference between one or more of the groups does not specify which particular group(s) are different. Consequently, without pairwise comparisons, all but one of the groups may appear statistically different. In multiclass classification, metrics ranking based on the F value or p values, typically at a significance level of $\alpha = 0.05$ or $\alpha = 0.01$, may bias the selection of parameters to ones with high or low values that can only distinguish one surface type from others but not each surface group individually. Such parameter ranking may result in top metrics primarily relevant for distinguishing one type of surface from others. Conducting pairwise comparisons between groups is essential to mitigate this.

2.4 Theory of signal detection and applications of d' index to parameter selection

Signal Detection Theory (SDT) is a fundamental framework widely applied in various fields, including industrial quality control, to model observer decision-making processes [74]. Originating in the early 20th century, SDT gained prominence during World War II in the field of radar technology for distinguishing between signal (enemy aircraft) and noise (natural phenomena) [75]. Since then, it has found applications in diverse domains, offering a robust methodology for evaluating performance in tasks involving the detection of signals amidst noise. A graphical representation of a theory of signal detection is provided in Figure 2.4

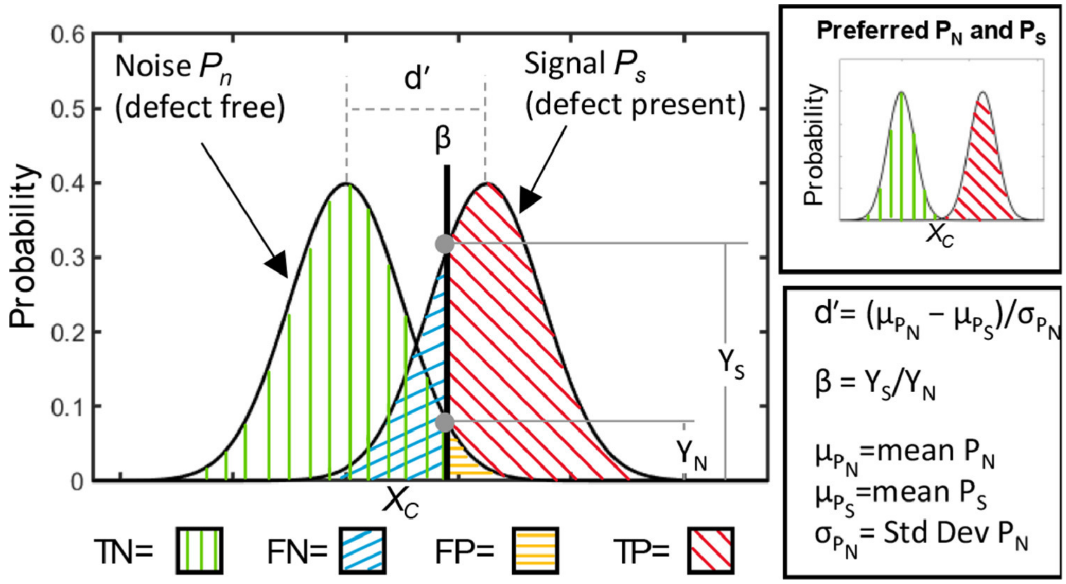


Figure 2.4: A graphical representation of a theory of signal detection adopted from [5].

As pointed out by Baker [74] investigators have applied SDT to the inspection of defects in small fabricated parts (Smith and Barany [76]), flat glass (Drury [77], Drury and Addison [78]), hooks (Sheehan and Drury [79]), electronic conductors (Wallack and Adams [80]), and microminature electronic components (Smith [81]). One of the central tenets of SDT is its capability to provide a measure of detection ability independent of variations in the response criterion, a critical factor in industrial quality control settings [74]. The response criterion represents the subjective standard observers use to determine whether items are classified as conforming or defective, highlighted as β in Figure 2.4. Specifically, d' remains constant despite changes in the response criterion affecting the accept or reject decision. However, this is only necessarily true when the two distributions are normal and are of equal variance. The calculation of d' in this case is shown in Equation 2.2.

$$d' = \frac{\mu_{P_S} - \mu_{P_N}}{\sigma} \tag{2.2}$$

Where μ_{PN} and μ_{PS} are the means of noise and signal distribution, respectively with equal variance $\sigma = \sigma_{PN} = \sigma_{PS}$. d' is helpful as a measure that can be translated from humans' performance ability relative to the expected performance of a quantitive surface measurement. When normality and equal variance assumptions are met, the d' index makes it possible to describe the difficulty of a task independent of the decision criterion and, therefore, also the classification algorithm applied to the data.

Historically in manufacturing environments, the SDT framework was used to model the decision-making process involved in human visual inspection (HVI), with d' providing the means to quantify the ease of detection and characterize expected performance variations within the decision process [5]. A 2021 review paper by See et al. [82] and works by [83] suggest that most HVI systems have d' in the 2.5 range, with 3.5 being considered a very easy detection task and d' values less than 1.5 regarded as very difficult. The d' metric provides a distinct advantage over Receiver Operating

Characteristic (ROC) curves in terms of interpretability and handling imbalanced class distributions. ROC curves [84] illustrate the trade-off between true positive and false positive rates at various classification thresholds, however, d' is expressed as a single value, representing the separation between distributions in terms of standard deviations. Although the Area Under the Curve (AUC) can be derived from ROC curves, it is limited to a maximum value of 1. While a high AUC suggests a good separation between distributions, it does not provide information on the full extent of the separation. In contrast, d' can take on infinite values, allowing for a more explicit understanding of the separation between the distributions. Readers are referred to [5] for a graphical representation of the relationship between $d' \in [0, 3]$ and ROC curves [6].

2.5 State of the art surface measurement & characterization for AM

Several methods are available for providing topography measurements of AM surfaces, and each has different technologies and operating principles for how the 2, 2.5, 3D data is acquired. Insitu monitoring enables the characterization of the liquid melt pool during LPBF fabrication. However, due to the long working distance the sampling interval is typically on the order of 8-100 μm per pixel [85, 86]. Also, the bright emission of the laser leads to high contrast, which can over-saturate detectors. Therefore, many details of the as-built surface cannot be resolved by melt-pool monitoring alone. Surface topography measurements can provide a wealth of information about meltpool conditions after the build process. For instance, the ends of the tracks reveal a snapshot of a frozen solidified meltpool [87]. Observing the topography formed by one more multiple tracks made up of solidified melt pool(s) provided a sense of melt pool stability during part fabrication. A common ex-situ surface measurement or image used in LPBF is single laser tracks on smooth, bare metal surfaces. This assessment provided data on melt pool behavior that is free of the variability of the powder layer, substrate roughness, and irregularities. Much of this characterization is often qualitative. For instance, plateau Rayleigh instabilities, balling, distortion, and other effects can be interpreted without numerical quantification. Ricker et al. (2019) analyzed the chevrons of single tracks. The authors provide a characterization of the steady-state melt pool morphology that could be represented by a normal distribution with parameters that were dependent on the build conditions used to create the tracks. The authors also mention that examining single or overlapping autogenous laser tracks may offer a relatively simple, rapid, and inexpensive way of evaluating the amenability of an alloy to the high thermal gradients, solidification rates, and stresses inherent in additive manufacturing [88]. Similar studies that focus on analysis of single scan or multi-scan track analysis are included in [89, 90, 91, 92, 93, 94, 95].

Parameter-based characterization methods for AM surfaces focus on quantifying the statistical arrangement of topographical data. According to a review paper by Townsend et al. (2016) [96], most AM surface characterization up until 2016 was predominantly achieved by stylus measurements and computing ISO 4287 texture parameters on profiles, with the Ra parameter being the most widely adopted. The review also suggested that texture characterization is primarily performed to improve understanding of the capabilities of the AM technology being studied. Surface texture metrology tends to lend itself to early-stage development of manufacturing technologies. However, at that point in time, there has been limited research into the correlation between component functional performance and surface texture [96]. It has been pointed out Fox et al. (2016) and many others that Ra provides little information on the specific features that make up the surface texture [97, 98]. Still, measuring Ra and Sa is common. This is because the legacy parameters have an intimate history rooted in tolerance for different applications, such as mating or bearing surfaces. The average surface roughness of parts produced through Laser Powder Bed Fusion (LPBF) varies depending on the material, process parameters, part geometry, etc. Generally, LPBF parts exhibit smoother top surfaces with average roughness (Ra) ranging from 5 to 30 μm [99, 100, 101]. In contrast, side walls and overhangs tend to have higher roughness, typically ranging from 5 to 50 μm or higher, due to layering effects, laser angles, and dross formation [102, 103, 104, 105]. Achieving desired surface finishes often involves post-processing techniques like machining, shot blasting, or chemical treatments. However, this is one application of surface texture metrology, and Ra is not always the most relevant parameter for characterizing AM surfaces. The correlation between surface texture and internal defects or density of LPBF parts has been an area of research aiming to elucidate the effect of process parameters on surface roughness and the influence of internal defects and inconsistencies in surface roughness during melting and solidification [106]. Gockel et al. looked at the relationship between fatigue, process parameters, and surface roughness. The authors used structured light scanning to measure fabricated tensile samples. They show that Sa and Sv decrease with increasing power; at high speeds, there is a large variation in the measured Sa values for the replicate samples. This indicated that the process might be unstable and create additional variation in the surface results at higher laser velocity speeds. The authors also highlight challenges with using the ISO 25178-2 parameters in isolation. This is because the powder particles stick to the surface, and other protrusions superficially make it difficult to observe changes in the melted surface [107]. Work by Thompson (2019) and others provides examples of common texture parameter values when particles are segmented from the surface before their computation [108].

To provide quality assurance to additive manufacturing processes, there is a growing desire to measure and certify each process layer using different optical systems. Insitu fringe projection and line scanner systems can provide topographic information of the powder and subsequently fused layers. However, there are several tradeoffs involved in terms of the accuracy and resolution of the measurements that can be acquired.

Meanwhile, the use of texture parameters to characterize spread powder layer quality is an increasing research trend. Notably, Berez et al. (2022) used a Keyence LJ-X8080 laser line profiler, providing a 12.5 μm sampling interval. The preliminary works indicated that the profiler and use of areal surface texture parameters were viable for characterizing defect content, but an algorithmic approach was currently lacking [85]. Sien et al. (2018) provide good examples of how DigitalSurf MountainsMap software can be used to isolate spatter, geometric features of weld track features, and enhance chevron patterns using different data processing techniques [109]. However, the use of these measures to inform scan strategies is not well established. Other state-of-the-art methods focus on predicting what surface parameters would be based on variations in processing parameters [110].

2.6 Discussion

The literature indicates that many industries and research would benefit from clear examples of how to relate surface topography to processing conditions. Furthermore, the literature review also suggests that a continued concerted effort towards standardization and methodological transparency could enhance the transferability of knowledge in this dynamic field. Researchers and manufacturers alike face a variety of challenges when trying to select appropriate surface texture parameters to characterize their process. To make matters worse, the so-called "parameter rash" and an existing plethora of surface descriptors for areal intensity maps and height maps make the selection of optimal surface descriptors Edisonian in nature and dependent on experiential knowledge. Additionally, choosing an appropriate surface scale and set of filtering/form removal operations further complicates the selection processes, as changes to both the latter and former can significantly change parameter values. In the absence of domain expertise, anything less than a brute-force evaluation of all possible parameter combinations, surface scales, and filtering sequences leaves the question of whether a suggested characterization is truly best for an open-ended application.

The literature also suggests quantifying a part's surface using only one or a few metrics in a way that can be easily understood and correlated directly with undesirable process variations and performance outcomes is far from trivial, with additive manufactured surfaces being a prime example. The variety of high and low-frequency spatial content of LPBF surfaces has the potential to make statistical characterizations of the surface topography highly variable in comparison to more conventional processes such as milling or turning. General characterizations of the surface, such as Sa, can provide a general sense of surface roughness but may not be a reliable process monitoring index for statistical process control and qualification. The question remains as to what surface measurements are most helpful and can be leveraged to support the qualification of LPBF parts and processes. Surface characterization could aid in understanding the relationship between processing parameters and part quality; however, there remains a significant gap in understanding how these parameters actually affect part quality and defect generation mechanisms [111]. Addressing this gap is essential for advancing additive manufacturing technologies and improving the quality and reliability of AM-produced components. Additionally, although flawed surface topography will be described qualitatively, the literature does not provide well-established measurands for determining the relative shape and size of atypical solidified melt pool geometries report in [87, 112, 113, 59].

Overall, the literature underscores the complexity of surface characterization and the challenges involved in selecting appropriate parameters and thresholds for classification. Developing systematic methodologies for isolating relevant descriptions and gaining a deeper understanding of the relationship between surface parameters, processing conditions, and part quality are crucial steps toward advancing surface characterization techniques and optimizing surface properties.

2.7 Summary

Through the review of state of the art, key findings in relation to the aims and objectives of this dissertation can be found:

- Parameters are selected ideally based on their relevance to the functional requirements of the surfaces being analyzed. For example, in automotive or aerospace applications, parameters related to sealing, friction, or wear resistance may be prioritized.
- The large variety of applications and the multitude of available parameters for specifying and characterizing surface texture present challenges in selecting the most relevant parameters for a given application.
- No reported use of d' index for selecting the best surface texture parameters to quantify the effect of a process on the surface topography. However, the Fisher score, ANOVA, and other measures are reported.
- Laser powder bed fusion additive manufacturing produces complex surface structures, creating challenges in isolating features of interest and quantifying topography with conventional texture parameters.

In this complex landscape of developing surface characterization solutions to application specific problems, algorithmic and intuitive measures of the surface area are preferable. Given the number of standardized texture parameters that can be used to describe a surface and its features, it was decided that the developed methodology to meet the aims and objectives of this dissertation should automatically down-select parameters for the user without requiring an arbitrary threshold. However, for LPBF AM surface characterization, a separate algorithmic methodology will be developed to isolate and characterize topographical regions of interest. The intent will be to establish new quantitative measures of flawed topography that can be used to better understand the relationship between local process conditions during part fabrication and meltpool-surface formation.

CHAPTER 3: (Paper 1) CONSTRUCTION OF A MULTI-CLASS DISCRIMINATION MATRIX AND SYSTEMATIC SELECTION OF AREAL TEXTURE PARAMETERS FOR QUANTITATIVE SURFACE AND DEFECT CLASSIFICATION

Overview of paper 1

The first paper presents a comprehensive evaluation of the framework developed to achieve the aims and objectives of this dissertation. A systematic and interpretable methodology, SQuID^{TM}, is introduced to objectively select the best subset of ISO 25178-2 parameters for classifying various surface states. An algorithm for constructing a discrimination matrix and selecting parameters for multiclass classification is proposed. The benefit of this approach lies in its ability to determine the effectiveness of the chosen parameters in distinguishing between different surface types, independent of the classifier used. By minimizing the number of parameters required for accurate classification and eliminating subjective threshold-setting, the methodology ensures objectivity in the selection process. Additionally, the paper provides an assessment of the d' metric to quantify the ability of surface parameters to facilitate successful classification. The methodology is evaluated using the open-source NEU dataset to establish a common benchmark between the developed approach and existing techniques in image processing, machine learning, and deep learning. This contribution is significant as previous reports lack a holistic assessment enabled by the d' matrix to assess classification capabilities. Furthermore, the paper presents a graphical relationship between the accuracy of unbiased classifiers and the d' value for the respective metrics employed by those classifiers, illustrating the complexity of achieving different levels of classification accuracy.

Abstract

The surface of every manufactured component has a topography resulting from its fabrication route, which influences its final functionality (optical performance, wear resistance, aesthetic quality, etc.). While the quantitative characterization of the surface is essential for detecting process-induced variations or defects and predicting functional performance, the plethora of surface descriptors developed by the image processing and surface metrology communities makes the selection of optimal surface descriptors Edisonian in nature and highly dependent on experiential knowledge. This work proposes a systematic approach for selecting surface parameters that best distinguish between different surfaces as grouped by visual or process-related differences. Using a form of univariate analysis rooted in signal detection theory, the predictive capability of a discriminability value, d', is demonstrated in the classification of mutually exclusive surface states. A 'discrimination matrix' that offers a robust feature selection algorithm for multiclassification challenges is also introduced. The generality of the approach is demonstrated on the Northeastern University dataset consisting of intensity images from six different surface classes commonly found in hot-rolled steel strip operations. Using the outlined approach, it was found that only four surface descriptors used in conjunction with a simple decision tree classifier achieved a 95% classification accuracy. Surface descriptors used in the study were limited to those described within the ISO 25178-2 standard, while machine learning approaches were limited to a decision tree classifier. The reasoning for both is to maintain as much algorithm output interpretability as possible; the advantages of such is discussed from the perspective of the larger goal of linking surface texture to manufacturing processes and surface functionality fundamental mechanisms.

3.1 Introduction

An Automated Visual Inspection (AVI) system has two main functions in manufacturing environments: surface defect detection and classification [1]. The detection of surface defects is critical to the closed-loop quality problem, while defect classification allows manufacturers to better understand the state of the process, enabling them to make more informed decisions about the appropriate actions to take after the defect has been detected. Although the concept of detecting and classifying a defect is simple, quantitatively defining exactly what constitutes a defect, and the logic which is employed to assign the defect to a particular category is far from trivial [2]. The lack of a standardized framework to perform systematic surface analysis impedes efforts to identify appropriate surface texture metrics to operate as process monitoring indices and support statistical process control [3], and to understand how the surface topography of the part relates to its functional performance [4]. This also reduces opportunities for manufacturers to isolate metrics that can, at a fundamental level, help understand how process physics govern the formation of the resulting surface topography and hence its functionality. While a solution (appropriate metrics and processing algorithm) can typically be found for a specific application, the trial-and-error approach used in its isolation can come at the expense of overfitting, reduced interpretability, and a limited understanding of the final solution's suitability for the application or other classification challenges. In short, selecting optimal image processing routines and metrics to achieve high defect detection and classification accuracy remains Edisonian in nature. For example, many different approaches have been taken to quantify a variety of surface types: [5–11] use gray-level co-occurrence matrices (GLCMs), and [5,12] use local binary patterns (LBP) and its variants for rolled steel covered extensively in the survey by [13]. However, little discussion was provided as to how the metrics were chosen, or indeed as to why they were selected as being the best for the application at hand. The chosen metrics when combined with machine learning approaches achieve classification accuracies ranging from 94% to 98%. While impressive, the complex feature vector resulting from the aforementioned image transforms is generally difficult to interpret and correlate back to the physical nature of the surface topography. This limits a reader's ability to assess whether such approaches would translate to their specific application and expected surface texture variations [5–7,14]. The use of convolutional neural networks (CNNs) to automatically classify common defect types has also found application in manufacturing environments, i.e. [15,16]. Although this approach can be very successful in emulating human visual inspection, it does not inherently provide standardized output metric(s) that designers and manufacturers can use to intuitively link the state of the surface topography to process conditions and/or functionality. Two key elements, currently missing, that would be highly desirable in any interpretable AVI system for defect and texture classification are; (1) quantitative, intuitive metrics, and (2) a systematic method to identify and assess effective metrics for the classification challenge at hand. To address the first, the surface metrics proposed are not the conventional image processing metrics, but rather those found in the ISO 25178-2 standard defining areal surface texture parameters. The reasoning for their use is that they focus on the geometric state of a surface's topography, and by virtue of their lineage and deep roots in manufacturing environments, can provide intuitive characterizations of a surface's physical topography; in turn, this can provide rich insight into correlations and causations arising from processing physics, surface formation, and final functionalities. For instance, works summarized by Jansons et al. [17] and Das et al. [18] infer functional characterizations by correlating ISO 25178-2 metrics with fluid retention abilities and tribological properties [19], in addition to fatigue, friction, adhesion, bonding, and corrosion resistance [20]. To assess the second element, i.e. the poten-tial effectiveness of a metric in any specified classification challenge, an approach grounded in signal detection theory (SDT) utilizes a single, easily interpretable d'value. The ability of the proposed framework to automatically isolate metrics from a predefined list and correctly classify surfaces in binary or multiclass classification inspection sce-narios via a simple decision tree will be compared to those found in the literature using more complex image analysis features combined with machine learning (ML), and deep learning (DL) approaches. The robustness of the d'metric and the overall framework is demonstrated on the publicly available Northeastern University (NEU-CLS) dataset containing six different hot-rolled steel surface classes.

The organization of this paper is as follows: Section 3.2 includes details of the NEU-CLS dataset and a summary of the ISO 25178-2 areal parameters used in this study. Section 3.3 outlines the SDT-based method-ology for predicting the ability of areal parameters to correctly classify the challenging NEU-CLS dataset. Section 3.4 details the data processing framework that formalizes the methods described in Section 3.3 into a set of systematic steps and automatic identification of surface classification metrics. Section 3.5 presents experimental results and validation of this framework when applied to the NEU-CLS dataset. The performance of the approach is also compared to other classification approaches applied to the NEU-CLS dataset as found in the literature. Section 3.6 summarizes the proposed method's advantages, future work directions, and extended use cases for other surface metrology applications.

3.2 Dataset and metrics used for analysis

The following sections provide details on the image dataset and metrics included in the analysis.

3.2.1 Dataset details

The NEU-CLS dataset, see Fig. 3.1, is a popular and publicly available dataset used in surface classification research; the original paper by [5] has more than 400 citations. The dataset consists of images of six different surface classes resulting from steel rolling operations; hot-rolled steel samples containing inclusions (In), scratches (Sc), patches (Pa), crazing (Cr), pitted surfaces (PS), and rolled-in-scale (RS). Each class contains 300, 200 × 200, 8-bit bitmap images, a total of 1800 images. Four examples of each defect class are shown in Fig. 3.1. A link to the original dataset can be accessed at [21]. The NEU-CLS dataset, while only addressing one type of manufactured surface, is chosen as it provides good examples of commonly encountered classification challenges that can be difficult to replicate in a synthetic database. The initial classification of the six types of defects was done via a qualitative, visual assessment; this is a process that inherently lends itself to subjectivity and high degrees of variability [2,22,23]. The images contain similar, yet differently labeled surface textures, i.e. the three rightmost columns in Fig. 3.1. The pitted surface (PS), crazing (Cr), and rolled-in-scale (RS) defect classes all contain small discrete, isolated darker regions on quasi-homogeneous backgrounds. Identification of metrics, never mind their associated threshold values, that will differentiate one class of defect from another is far from obvious. The dataset also contains within-image and image-to-image variations in illumination conditions. For example, in the topmost left image, there is a general increase in background intensity going from left to right, which is not as strong in the other images within the same class. While in column four there is a large difference in the level of background illumination between the top- and bottom most images. Both these types of issues could require some level of preprocessing or normalization to overcome illumination variations; the selection of an optimal algorithm to achieve this is also nontrivial

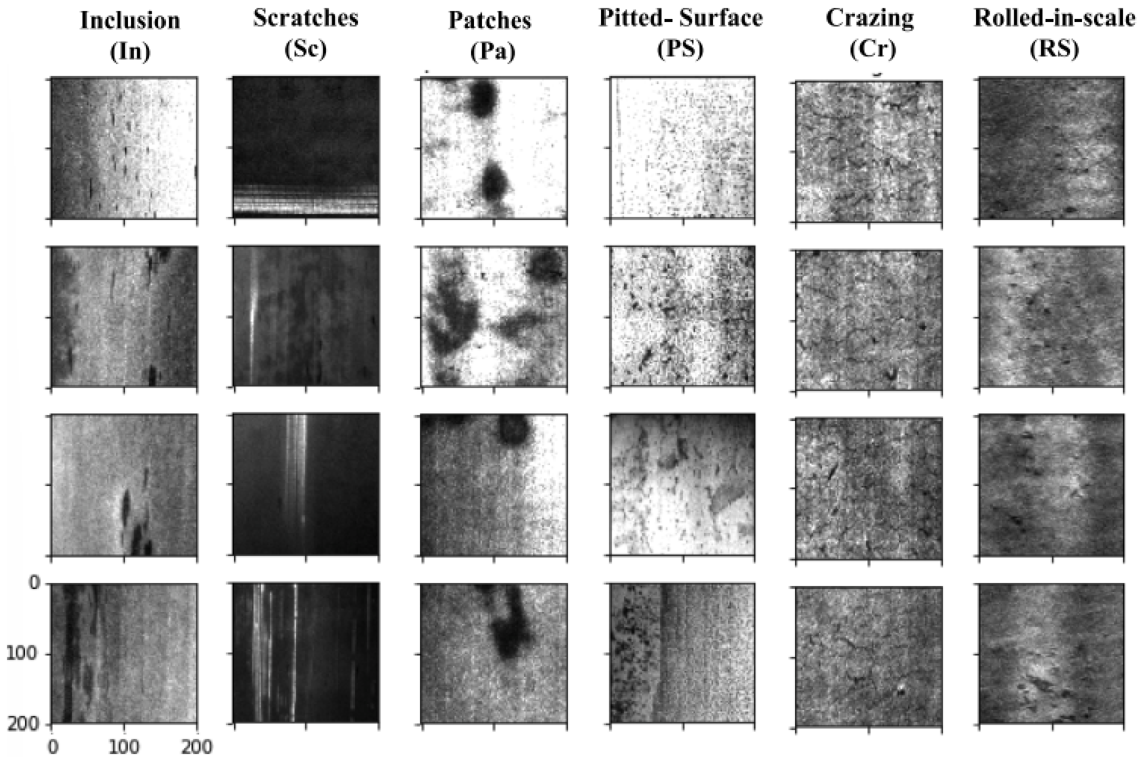


Figure 3.1: Examples of the six surface classes included in the NEU-CLS dataset. Surface classes exhibiting high interclass similarities include Cr and RS, and surface classes exhibiting low interclass similarities include Cr and Sc.

3.2.2 The ISO 25178-2 areal texture parameters

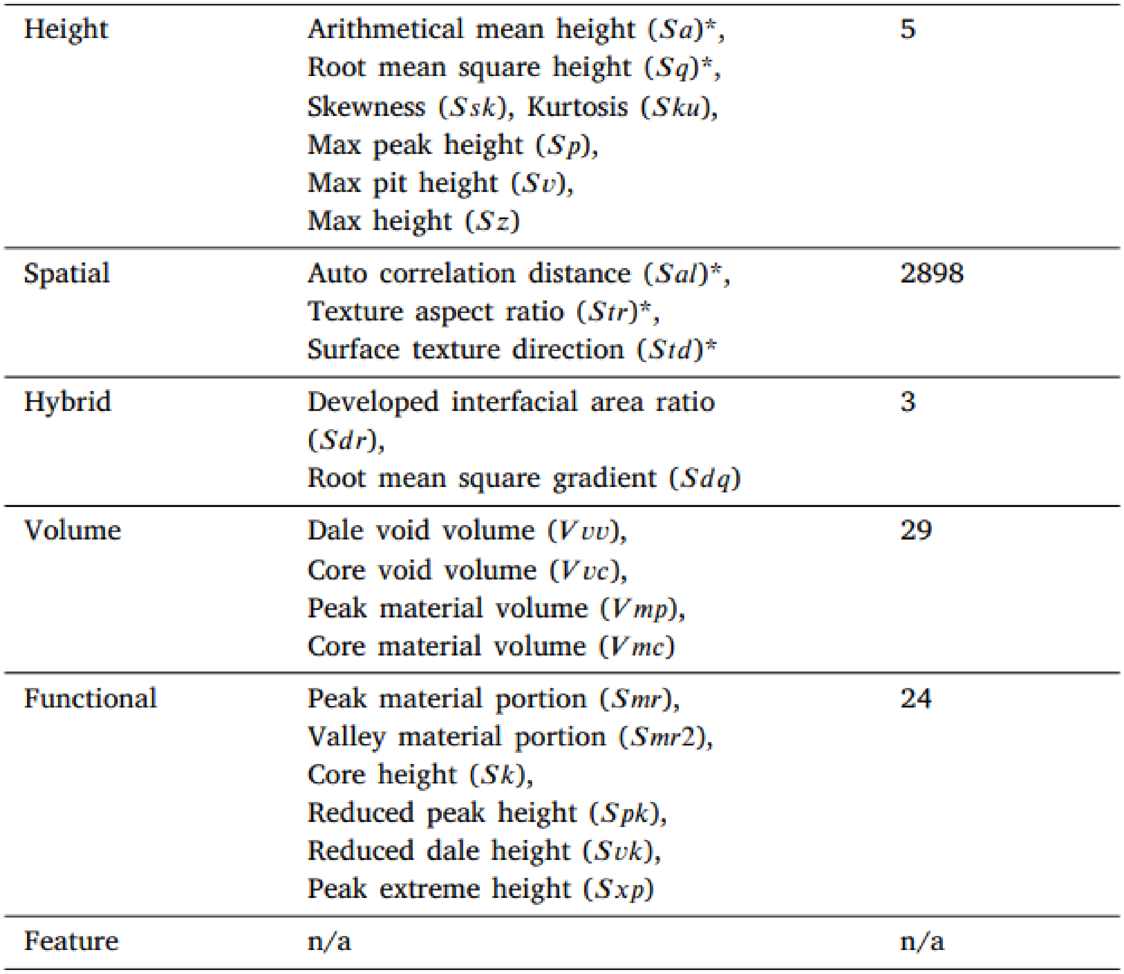
The ISO metrics extend beyond amplitude parameters and include those that within their calculation incorporate the spatial relationship between the surface height variations and have demonstrated their ability to track surface functionality [24,25]. The ISO 25178-2 metrics are typically applied to areal height maps acquired via contact stylus profilometry, coherence scanning interferometry, focus variation microscopy, etc. Traditionally, these metrics have not been applied to intensity images (i.e. images represented by a single channel of pixel values, with each pixel indicating brightness or intensity) captured from conventional CCD and CMOS cameras. While it is acknowledged there is no guarantee that intensity correlates directly with height values, the former does capture the spatial content of a surface. The aim of this work is not to assign equivalency between intensity images and height maps but to offer

ISO 25178-2 parameters as an alternative to conventional image processing metrics for characterizing surface textures described by non-physical units. The initial range of features is limited to ISO 25178-2 metrics to both limit the scope of the work and highlight the potential of these interpretable metrics for tough multiclassification challenges. The ISO 25178-2 standard details over forty metrics to describe surface topographies and textures. A subset of these metrics which are used in this work are listed in Table 1. While this paper will not provide full descriptions of all, a few key ones will be described. For full descriptions refer to the standard itself or other published literature [24-27]. The height metrics are perhaps the most used and understood by industry, with the average surface height, Sa, and the root mean square roughness, Sq, terms being ubiquitous. Skewness, Ssk, and kurtosis, Sku, are the third and fourth central moments of the surface height distributions. The Ssk term is useful in assessing the existence of peaks on the surfaces, while the Sku term is useful for describing the sharpness of peaks on surfaces [25]. The max height of the surface Sz is defined as the distance between the max peak height Sp and max valley depth Sv; representing the absolute highest and lowest points found on the surface. Hybrid parameters combine information on both surface amplitude and wavelength and consist of two metrics, namely, the interfacial area ratio, Sdr, and the root mean square gradient, Sdq. The former expresses the increase in surface area due to texture relative to the surface's planar definition. Whereas the latter is calculated as the root mean square of the surface slopes. Sdr and specifically Sdq demonstrate utility for assessing surfaces in sealing applications and controlling cosmetic appearance [24]. Functional and volume categories are metrics derived from the material ratio curve, i.e. cumulative probability density function of surface heights. These metrics characterize different aspects of the surface by dividing into it three parts peak, valley, and core. These categories are commonly used in tribological and automotive applications for characterizing oil retention and wear. Readers are referred to [25] for detailed descrip-tions. All the metrics are reported after applying specific processing operations outlined in the ISO 25178 standard, i.e. the surface form and appropriate high and low-frequency spatial content should be filtered from the surface prior to calculating the metrics. In this work, only a best-fit plane is removed from any image prior to computing the ISO metrics.

The metrics listed in Table 3.1, key to the computational framework detailed in section Section 4, are calculated using Python 3.7 and a NumPy library. Metrics denoted by '*' have been verified against the NIST surface metrology algorithm testing system [28]. The same table details the time required to compute a particular category of metrics. The computational time is based on an image size of 200 × 200 using an Intel(R) Core (TM) i7-7700 CPU with a 3.60 GHz processor. The reported times include the contributions of pre-processing operations such as form removal and construction of the autocorrelation function (ACF) and the material ratio curve, with the former contributing to most of the computational overhead. The ACF, a prerequisite for the spatial parameters auto-correlation length, Sal, and texture aspect ratio, Str, posed the only potential bottleneck for real-time implementation, accounting for over 97% of the total computation time of the twenty-two metrics. Apart from these two spatial metrics, which take nearly three seconds to compute, the remaining twenty metrics can be calculated in less than 30 ms. Undoubtedly, the reported times could be significantly reduced with optimized implementations of the algorithms and better hardware.

By mathematical definition, some of the metrics listed in Table 3.1 are correlated, meaning that they measure similar aspects of the surface. For example, Sa and Sq are positively correlated, with Sq being more sensitive to abrupt changes in surface height. In addition, as demonstrated by [25] the Sdq and Sdr parameters are highly correlated with Czifra et al. [29] obtaining a determination coefficient of 0.998 between them.

Table 3.1: ISO 25178-2 areal metrics used in the study and associated category computational times



Height	Arithmetical mean height (Sa)*, Root mean square height (Sq)*, Skewness (Ssk), Kurtosis (Sku), Max peak height (Sp), Max pit height (Sv), Max height (Sz)	5
Spatial	Auto correlation distance (Sal)*, Texture aspect ratio (Str)*, Surface texture direction (Std)*	2898
Hybrid	Developed interfacial area ratio (Sdr), Root mean square gradient (Sdq)	3
Volume	Dale void volume (Vvv), Core void volume (Vvc), Peak material volume (Vmp), Core material volume (Vmc)	29
Functional	Peak material portion (Smr), Valley material portion ($Smr2$), Core height (Sk), Reduced peak height (Spk), Reduced dale height (Svk), Peak extreme height (Sxp)	24
Feature	n/a	n/a

3.3 Optimal metric selection for classification

3.3.1 Signal detection theory, d', and evaluation criteria

Signal detection theory (SDT) provides a framework for quantifying the ability of an observer to differentiate between (or classify) two different stimuli. This section outlines the implementation of SDT to select the optimal areal texture metrics to differentiate between two surface states, i.e. between different surface classes within the NEU-CLS dataset. A generalized example of the SDT premise is given in Fig. 3.2.

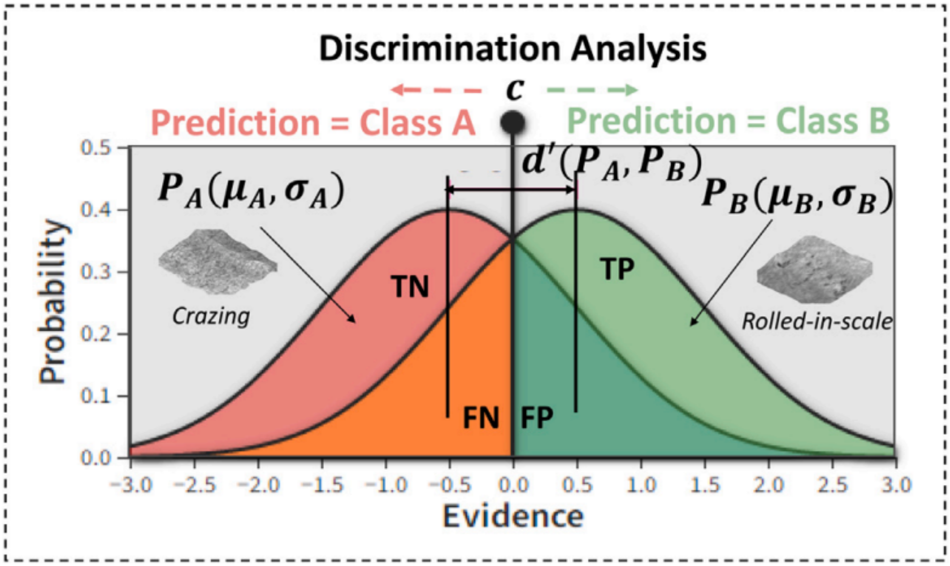


Figure 3.2: SDT-based approach for predicting a surface class according to a single surface metric.

The SDT model requires two probabilistic distributions, P_A and P_B, of a calculated surface metric, see Fig. 3.2. For example, P_A could be the distribution based on a metric listed in Table 3.1 for the Cr (crazing) surfaces, while P_B could represent the distribution for the same metric but calculated on RS (rolled-in-scale) surfaces. The 'evidence' shown along the x-axis in Fig. 3.2 represents the probabilistic range of values for the metric. The distributions P_A and P_B are independently constructed by the probability density function given in Eq. (3.1).

$$P(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2) \tag{3.1}$$

where σ and μ represent the mean and standard deviation of the metric values associated with a surface class, i.e. the mean and standard deviation of the numerical *Ssk* values associated with class ‘A’ (Cr) or ‘B’ (RS), and where x_i is the random variable corresponding to an *Ssk* value obtained exclusively from surface measurements of surface class ‘A’ or ‘B’. The decision criterion, c, shown in Fig. 3.2, indicates the threshold between the two classes. Any newly observed surface with metric value $x_i \geq c$ will be classified as surface class ‘B’ and any x_i is classified as the surface class ‘A’. A prediction resulting in a correct surface classification can be arbitrarily defined as either a true positive (TP) or true negative (TN), for example, the former referring to correctly classifying a type ‘B’ surface as ‘B’, and the latter referring to correctly classifying a type ‘A’ surface as ‘A’. It should also be clear from Fig. 3.2 that incorrect classifications may be made. This is because the probability of observing an x_i computed on a type ‘A’ surface to the right of the decision criteria c, or observing a random variable x_i computed on a type ‘B’ surface to the left of the decision criteria c, is not zero. Hence, a combination of predictions resulting in a false positive (FP) (i.e. classifying a type ‘A’ surface as ‘B’) and/or a false negative (FN) (i.e. classifying a type ‘B’ surface as ‘A’) is to be expected. Depending on the application, the implications of mistaking a surface type ‘B’ for ‘A’, may be greater than that of misclassifying‘A’ for ‘B’. One of these outcomes may allow defective parts to escape a manufacturing plant, while the other could result in scrapping good parts. It’s intuitive to see how shifting the decision criteria from 0 to -2.5 in Fig. 3.2 would result in an inspection system that always correctly classifies a type ‘B’ surface as ‘B’, at the risk of incurring a lot of false positives. For generality, in this work, no preference is given to minimizing the number of false positives over false negatives or vice versa. For this reason, the performance of an observer or classifier is determined by dividing the total number of correct predictions by the total number of predictions made. The assessment given by Eq. (3.2) is referred to as accuracy.

$$Accuracy = \frac{\#\, correct\, predictions}{\#\, predictions\, made} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.2}$$

Although the use of accuracy as a performance indicator is common, accuracy varies according to the placement of the decision criteria as well as the number of samples that are used in the evaluation. For example, for a binary task, with an even number of examples in 'A' and 'B', random guessing would provide 50% accuracy. However, if there are two examples of 'A' and four of 'B', random guessing would increase the accuracy of classifying 'B' correctly to 66%. Hence, balanced accuracy should be used in these cases. The focus of this work is not to determine the optimal placement of decision criteria in order to optimize accuracy but to determine if ISO metrics can sufficiently resolve the differences between any arbitrary surface groups 'A' and 'B', thus providing an optimal classifier. To do this the d' shown in Fig. 3.2 is used to approximate the number of standard deviations between the means of the two distributions being compared. The calculation of d' is given by Eq. (3.3).

$$d'(P_A, P_B) = \begin{cases} \frac{|\mu_B - \mu_A|}{\sigma} \text{ if } \sigma_A = \sigma_B \\ \frac{|\mu_B - \mu_A|}{\sqrt{\frac{(\sigma_A^2 + \sigma_B^2)}{2}}} \text{ if } \sigma_A \neq \sigma_B \end{cases} \tag{3.3}$$

The first and second cases shown in Eq. (3.3) apply when the two distributions are of equal and unequal variance. For rigorous mathematical details of the d' equation, its various forms, and assumptions regarding the data distributions, readers are referred to [30]. In any of its forms, the d' value provides an intuitive measure of the ability of a chosen single-valued metric to differentiate between the two surface types; greater d' magnitudes imply less overlap between the two distributions thus meaning the chosen metric is more likely to assist in correct classification of new unseen surfaces independent of where the decision criterion is placed. The following sections investigate the ability of this simple d' metric on datasets containing non-Gaussian datasets combined with unequal variances. Historically in manufacturing environments, the SDT framework was used to model the decision-making process involved in human visual inspection (HVI), with d' providing the means to quantify the ease of detection and characterize expected performance variations within the decision process [2]. A 2021 review paper by See et al. [31,32] suggest that most HVI systems have d' in the 2.5 range, with 3.5 being considered a very easy detection task and d' values less than 1.5 regarded as very difficult. The d' metric provides a distinct advantage over Receiver Operating Characteristic (ROC) curves in terms of interpretability and handling imbalanced class distributions. ROC curves [33] illustrate the trade-off between true positive and false positive rates at various classification thresholds (c_i), however, d' is expressed as a single value, representing the separation between distributions in terms of standard deviations. Although the Area Under the Curve (AUC) can be derived from ROC curves, it is limited to a maximum value of 1. While a high AUC suggests a good separation between distributions, it does not provide information on the full extent of the separation. In contrast, d' can take on infinite values, allowing for a more explicit understanding of the separation between the distributions. Readers are referred to [2] for a graphical representation of the relationship between $d' \in [0, 3]$ and ROC curves.

3.3.2 Application to binary classification cases

In binary classification tasks, the goal is to simply differentiate between two different classes, i.e. Cr and RS. To achieve this, the d' term is calculated for each metric listed in Table 1, and the calculated d' values are used to rank the metricâs potential to make a correct classification. Metrics with the greatest potential to discern between surfaces ‘A’ and ‘B’ will have larger associated d' values. For the chosen metrics, the threshold level, c, is simply set mid-way between the two distributionâs mean values, such that $c = \frac{\mu_a+\mu_b}{2}$.

3.3.3 Application to multiclass classification cases

This section outlines how the proposed method can be extended to multiclass problems, like classifying all six classes within the NEU-CLS dataset. In such cases, the metric with the maximum d' value is identified for each unique binary classification task that can be constructed from the multi-class dataset, i.e. determine the best metric for each of the following classification tasks. In vs. Sc, In vs. Pa, In vs. Ps, etc. These individual tasks are denoted as T_i. Eq. (3.4) can be used to determine the number of unique binary classification tasks which can be constructed from a multi-class dataset containing a total of L class labels (i.e. types of surfaces).

$$k = L(L-1)/2 \tag{3.4}$$

In Eq. (3.4) k represents the number of unique classification tasks. For instance, the NEU-CLS dataset comprises six surface types, (i.e., $L = 6$), hence a total of 15 binary classification tasks can be constructed. For easy interpretation of the results, a discrimination matrix, or d' matrix for short, is constructed whereby the d' values associated with each candidate metric f_i (rows) and classification task T_i (columns) are displayed in a matrix-like format with a superimposed heat map where matrix elements are colored according to the magnitude of the d'values. More formally this matrix can be expressed as a lookup table represented by the function $d'(f_i, T_i)$.

Figure 3.3 illustrates the concept for a multiclass dataset including four surface types, yielding six classification tasks with seven candidate metrics. Metrics with $d' \geq 4$ are favorable (highlighted in green), $2 \leq d' \leq 4$ are considered less favorable (highlighted in orange), and $d' \leq 2$ are in gray. Highlighting the d' matrix by color can serve as a quantitative visual reference for quickly determining the suitability of ISO metrics for specific classification tasks, with the added value of making it easier for interested observers to understand what type of physical differences exists between classes. Of course, to automate the selection process, the top-ranked metric for each classification challenge is all that is required.

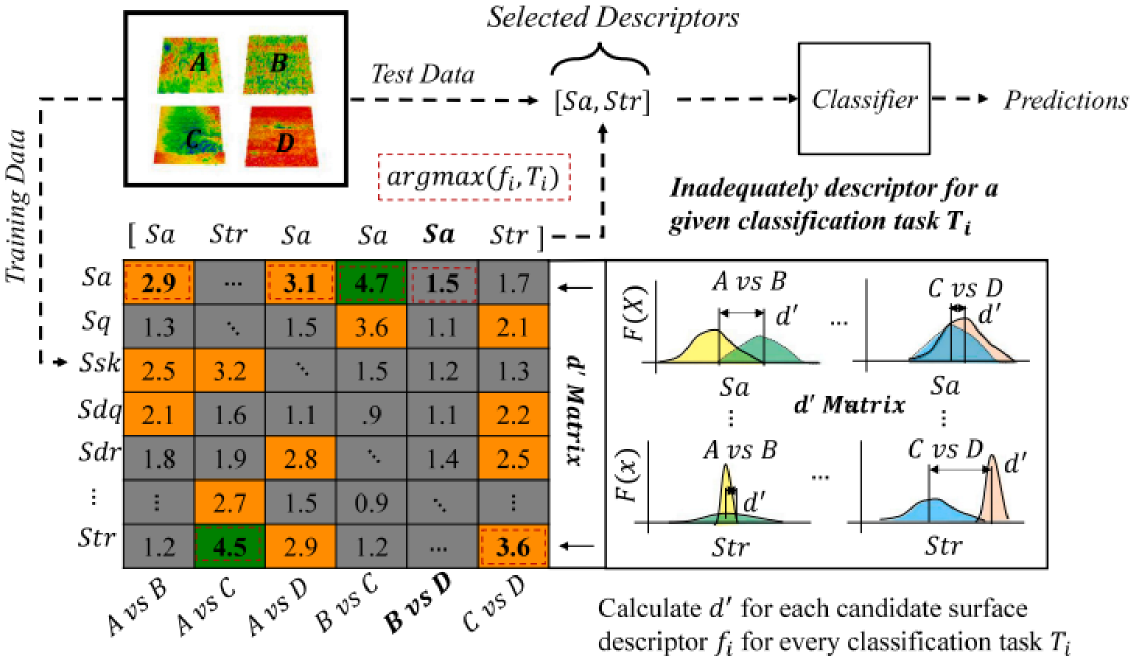


Figure 3.3: Example of d' matrix for a multiclass dataset consisting of 4 categories A, B, C, D. Selected descriptors represent the highest-ranking metrics (rows) for each classification task (columns) according to the calculate d' value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The algorithm for choosing the best surface metrics for the multi-class problem is carried out by selecting the metric with the maximum d' value for each column of the matrix, aggregating the results and removing redundant descriptions. In the ideal case, there would be a single row of highlighted green entries, indicating that a single metric alone can differentiate all the surface types in the dataset with little to no overlap. That, however, is not the case for this dataset. For example, in the d' matrix illustrated in Fig. 3.3, *Sa* is selected three times (‘A’ vs. ‘B’, ‘A’ vs. ‘D’, and ‘B’ vs. ‘D’) while *Str* is selected twice (‘A’ vs. ‘C’, and ‘C’ vs.‘D’). Applying this procedure to the example d' matrix shown in Fig. 3.3, these two metrics are selected for distinguishing between surfaces classes ‘A’, ‘B’, ‘C’, ‘D’. Notice the Sa value bolded for task ‘B’ vs. ‘D’ serves as a visual indication that despite Sa having the greatest d' among other candidate metrics, a value of 1.5 indicates the distributions are not well separated. Therefore based on the discussion in Section 3.3, it is unlikely that Sa, or any of the candidate ISO metrics, will be effective for differentiating between classes ‘B’ vs. ‘D’, i.e. the entire column is in gray. This indication intuitively infers that another candidate metric must be added to the analysis or developed specifically for the ‘B’ vs. ‘D’ task. Without intervention, it can be expected that a classifier which uses the suggested two metrics to make predictions will result in many misclassifications between surface types ‘B’ and ‘D’. On the other hand, one can also infer from this matrix that because Sa has large d' for tasks ‘A’ vs. ‘B’, ‘A’ vs. ‘D’,‘B’ vs. ‘C’, these surface types exhibit a unique roughness signature with respect to one another. In addition, the relatively large d' values for tasks ‘A’ vs. ‘B’ and ‘C’ vs. ‘D’ based on Str indicate that these surface types have different degrees of isotropy and/or anisotropy, which could be intuitively linked to process conditions. The proposed feature selection method inherently tries to minimize the number of suggested metrics to use for classification, and to reduce the ‘parameter rash’ phenomenon [34] of over-reporting and/or use of non-relevant metrics; at most k metrics (Eq. (3.4)) will be returned irrespective of the number of candidate metrics (rows of the d' matrix). Given a dataset (X) that consists of target class labels (L) and input metrics (F), construction of the d' matrix and subsequent down-selection of metrics is implemented according to algorithm 3.1.

Algorithm 3.1: Construction of d' matrix and down selection of metrics

```
Require: Dataset $X$, Target Class Labels $L$, Input Metrics $F$
Ensure: $d'$ Matrix, Selected Metrics $B$

  $d' \leftarrow \text{Matrix}_{|F|\times|k|}$
  for $f$ in $F$ do;
      for $i$ in range(0, $|L|-1$) do
          for $j$ in range(i+1, $|L|$) do
              $d'(f, l_i \text{ vs } l_j) \leftarrow \frac{|\mu(X_{l_i,f})-\mu(X_{l_j,f})|}{\sqrt{\frac{\sigma(X_{l_i,f})^2+\sigma(X_{l_j,f})^2}{2}}}$

  $B \leftarrow$ unique([$F$[argmax($d'$[:, $j$])] for $j$ in columns($d'$)])

return $d'$, $B$
```

3.3.4 Utilization of machine learning for multiclass problems

Although the suggested selection algorithm helps eliminate unnecessary metrics, resolving the differences between multiple surface types (i.e. multiclass classification tasks) often requires more than a single metric. In these cases, determining a set of thresholds/decision rules can be a difficult task. For this reason, machine learning models are often employed. Decision trees, a form of machine learning, offer an interpretable solution for automating both binary and multiclass classification problems. Because they are a type of non-parametric supervised learning method, unlike CNNs, they do not require the specification of model parameters such as the number of convolutional, pooling, and fully connected layers, prior to training. Instead, decision trees predict the class/type of surface by learning a series of decision rules directly from the training data. The deeper the tree the more complex the decision rules become [35]. However, one has the option to balance model complexity and performance by specifying a maximum tree depth prior to training (i.e. pre-pruning) or by removing parts of it afterward (i.e. post-pruning) [36,37]. Figure 3.4 depicts a decision tree model resulting from the training dataset in Section 3.3 where four different surface types were represented by two metrics, surface roughness Sa and surface isotropy Str. The resulting tree provides a visualization of nonlinear data patterns and removes any ambiguity from how an automated decision will be made regarding the classification of new surfaces. For example, given measurement values of Str and Sa of a surface the tree represents the logic that is followed for predicting which class the surface belongs to. This decision process starts at the root node where Str is compared to decision criteria C_1 learned from the training data. If the Str value of the surface is greater than c_1 then the next comparison would be to compare the Sa of the surface to decision criteria c_2. If the Sa value is less than c_2 the surface in question would be classified as 'C' and otherwise as a type 'A' surface. Another advantage of decision tree classifiers is that they do not require normalization (i.e. a preprocessing step that involves scaling all of the metric values to a common range, typically between 0 and 1 or -1 and 1) of the data, further reducing implementation complexity

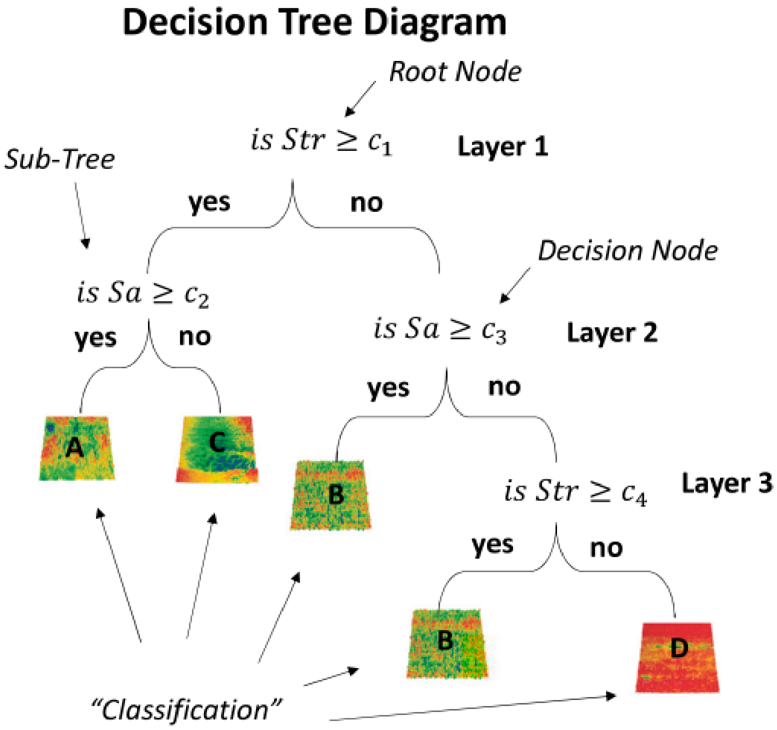


Figure 3.4: Example of a decision tree model for multiclass classification.

3.4 Data processing pipeline and analysis tools

3.4.1 Surface quality and inspection descriptors (SQuID^{TM})

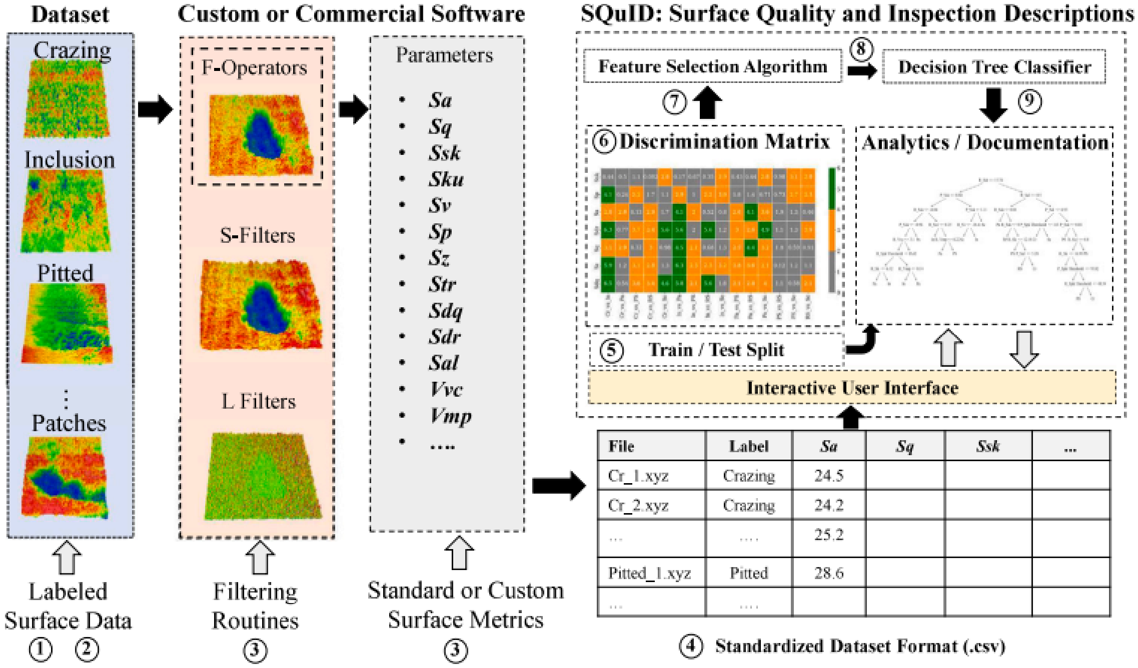
The developed approaches described in Sections 3.1-3.4 are formalized as a software framework called Surface Quality and Inspection Descriptors (SQuID). The package integrates the elements discussed in previous sections within a single software package/application to characterize the differences between surfaces according to categorical visual appearances, processes, or other functionally related differences using single-valued surface metrics. The flow diagram illustrating how the data is processed is detailed in Fig. 3.5. The software utilized to develop SQuID is based on Python programming language, uses NumPy and Pandas packages for data processing and manipulation, Scikit-learn for machine learning modules, and Streamlit for the graphical user interface. Users can access a web-based version of the tool [38] and source code through GitHub link [39]. To classify the six surface types in the

NEU-CLS dataset, twenty-two surface descriptors taken from ISO 25178-2 [26] are utilized, see Table 3.1. Prior to calculating the twenty-two metrics for each surface, a best-fit plane is removed from each image; this is the only pre-processing operation undertaken. The d' matrix is calculated for the

fifteen classification tasks and the metrics with the highest d' values identified. A decision tree classifier is implemented with an opensource package [40]. All default input arguments are used except for the criterion which is set to ââentropyââ. Complete details regarding the decision tree package used in this work can be found at [35,40,41]. For generality, the workflow of the SQuID pipeline consists of the following steps, with steps 4-9 being completely automated:

Collect images or areal measurements of the types of surfaces that need to be characterized and differentiated from one another. (For this example, the NEU-CLS dataset consists of 1800 grayscale images.) 2. Organize the data files into separate folders with the name of each folder indicating the classification. (The NEU-CLS dataset consists of six folders, one for each surface type.) 3. Determine data pre-processing sequence which should be applied to the surface image and subsequent metrics to be computed. (In this case study, best-fit plane removal and the calculation of the twenty-two ISO-based metrics) 4. Apply step 3 to each image and store the resulting data in tabular format; one row for each sample. 5. Split the dataset into training and testing sets using an equal number of samples for each class. (By default, a 50/50 train/test split is used) 6. Using the training dataset only, construct the d' matrix described in Section 3.4. 7. Apply the selection algorithm detailed in Section 3.4 and/or manually down-select metrics based on data analysis. 8. Use optimal metrics from the previous step to generate a decision tree. 9. A report is generated detailing the d' matrix, decision tree diagram, and confusion matrix indicating classification performance on the testing dataset.

In this approach, the greatest time investment is during data collection rather than during the analysis. The amount of time spent computing the ISO metrics to create the tabulated dataset for analysis will scale according to the speed at which the metrics and digital filtering can be applied. Once constructed, generating the discrimination matrix and evaluation of the decision tree occurs in less than a couple of seconds or minutes depending on the number of classes and examples used.



File	Label	*Sa*	*Sq*	*Ssk*	...
Cr_1.xyz	Crazing	24.5			
Cr_2.xyz	Crazing	24.2			
...	....	25.2			
Pitted_1.xyz	Pitted	28.6			
...	....				

Figure 3.5: Surface Quality and Inspection Descriptions (SQuID) frame work

In addition to decision trees, other classification models such as Support Vector Machine (SVM) [42], K-Nearest Neighbors (KNN) [43], and Naive Bayes (Bayes) [44], can also be employed. However, these off-the-shelf machine learning models, accessible at [45-47] that are used in the results with default parameters, have their limitations in terms of interpretability when compared to decision trees. For instance, SVM may exhibit biased predictions towards features with inherently larger values due to the distance metric used in the fitting process. To mitigate this issue, data normalization is required, which can complicate the interpretability of the model. The KNN classifier operates on an intuitive principle of predicting the class based on proximity to other classes in the dataset, however, visualizing or interpreting the decision space becomes challenging when more than three features are used for classification. Moreover, although both decision trees and KNN are non-parametric classifiers the latter relies on the dataset itself for making predictions, making scalability an issue with large datasets and complicating implementation and transferability across diverse manufacturing environments. The implementation of Gaussian Naive Bayes algorithm [47] for classification may also be considered less interpretable as the model does not provide the explicit rules that define the decision boundaries that are used as the basis for future predictions like decision trees.

3.5 Results

3.5.1 Binary classification

Figure 3.6 plots d' versus accuracy (Eq. (3.2)) for each ISO parameter associated with each of the fifteen binary classification datasets within the multiclass NEU dataset. This consists of a total of 330 data points; one per ISO metric for each of the fifteen classification tasks. In general, it was found that d' greater than 2 results in prediction accuracies of about 0.82, and that d' greater than 4 indicates an accuracy of 0.95, or higher, is feasible. It is not surprising that some ISO 25178-2 metric categories perform better at certain classification tasks than others. The height metrics (represented by $\triangle$) are seen to cover the entire d' spectrum, while the hybrid metrics (diamonds $\diamond$) achieve some of the highest d' values. This is likely because metrics such as *Sdr* and *Sdq* consider both spatial and height content, while the height metrics do not capture spatial relationships. That said, hybrid metrics are not infallible; Instances of low d' values for certain tasks can still be found in the bottom left of the plot.

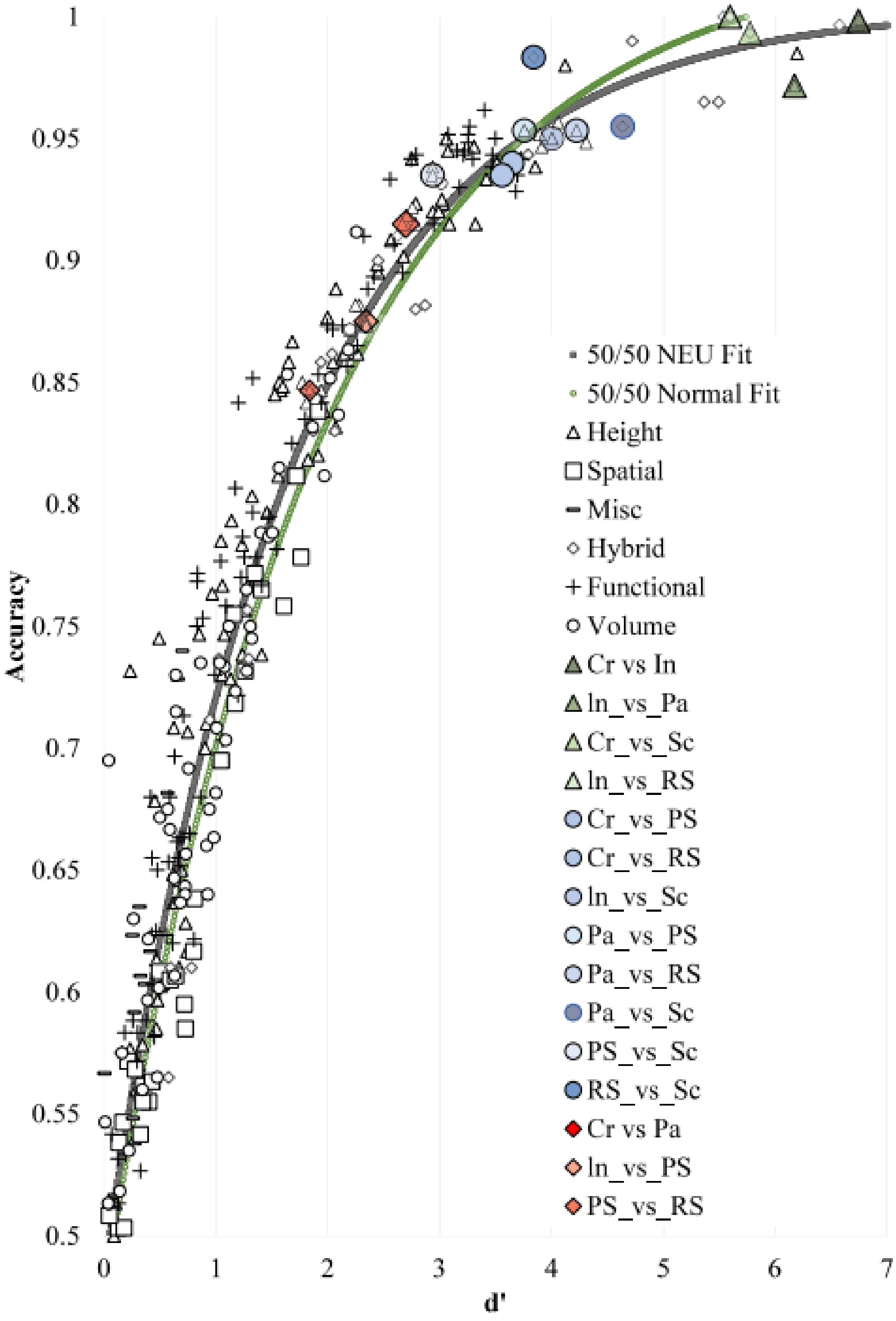


Figure 3.6: Classification accuracy versus d' value for each ISO 25718-2 parameter on every unique binary classification task in the NEU-CLS dataset. references to color in this figure legend, the reader is referred to the web version of this article.)

The variation in reported accuracy values for any d' value can range by ±5%. This variation can be caused by a multitude of factors such as; the random selection and ratio of samples in the training and testing sets, the suboptimal placement of the decision criterion, the presence of erroneous outliers in the dataset, and the degree to which the distribution of the metric values representing the surface types for a particular classification task deviate from a normal distribution. Figure 3.7, which illustrates the distributions associated with two selected ISO metrics, clearly indicates that the calculated metrics are not uniform or standard normal distributions. The light green line in Fig. 3.6 represents the best-fit line obtained for a synthetically generated Gaussian dataset which satisfies the normality assumptions. As little difference exists between this line and the best-fit line through the NEU-CLS dataset (dark-gray heavy line in Fig. 3.6) it suggests that the d' value, even calculated for a range of data distributions, is a robust indicator of a metric's ability to differentiate between two different classes. This is supported by Fig. 3.7 depicting the non-Gaussian distributions formed by the training (shaded) and test (dashed hatted lines) sets for Sq and Sdr on two different classification tasks. The bottom left plot shows Sdr for Sc vs. Cr having a $d' > 6$ and achieving 99% test accuracy.

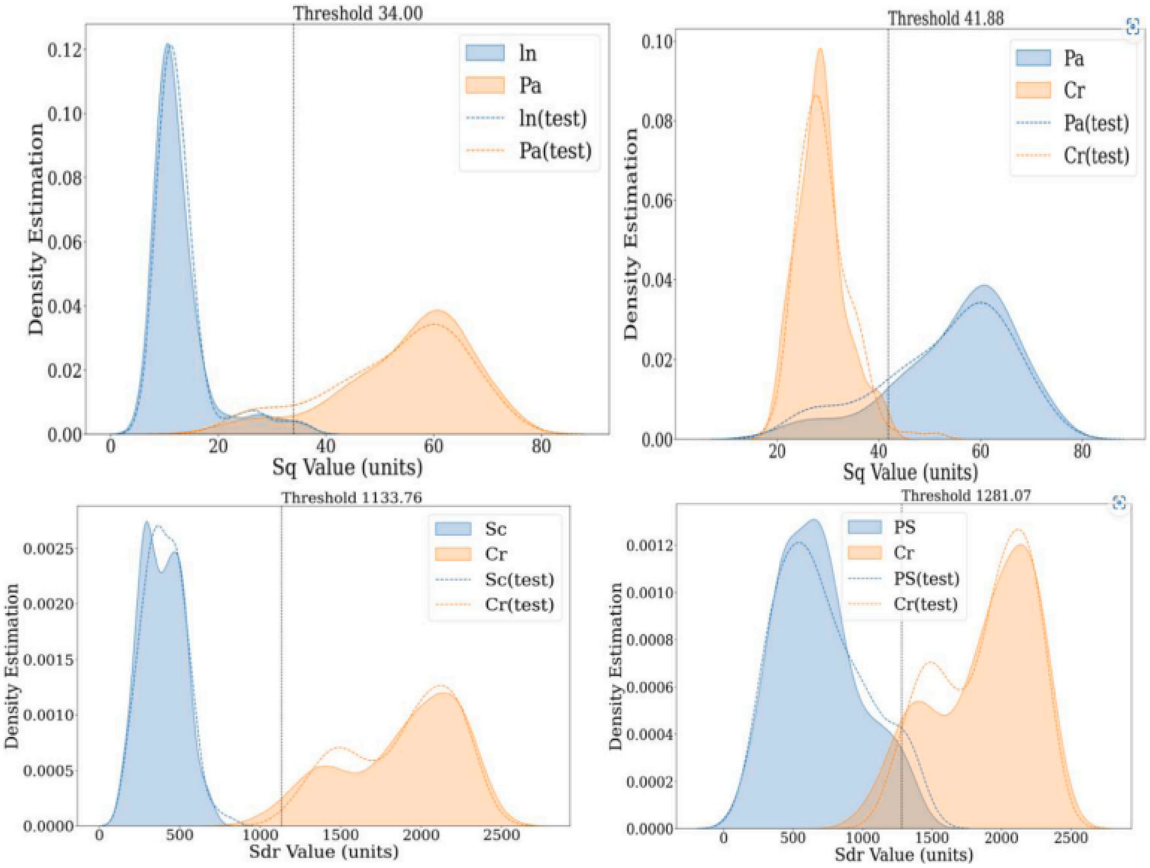


Figure 3.7: Examples of decision variable distributions from select NEU-CLS binary classification datasets Top left (In vs. Pa d' of 4.5, accuracy = 0.94) Top right (Pa vs. Cr d' 2.9, accuracy = 0.89) Bottom left (Sc vs. Cr d' 6.08, accuracy = .99) Bottom right (PS vs. Cr d' 3.76, accuracy = 0.94).

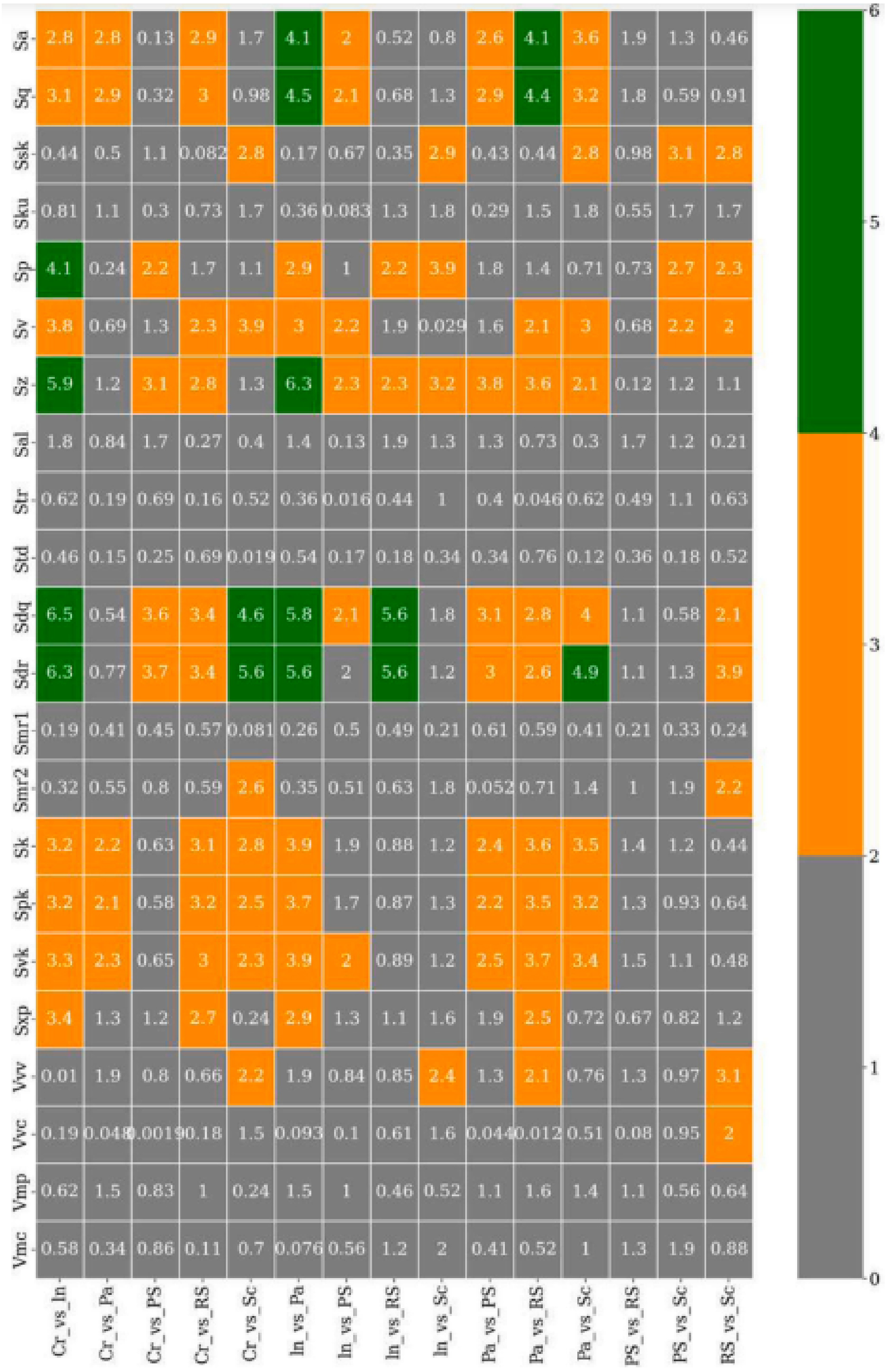
Table 3.2 lists the metrics with the highest d' value for each classification task; for the NEU-CLS case study, they are all either height or hybrid parameters.

Table 3.2: Best metric for each classification task, associate d' values, and the resulting classification accuracy

Task (T_i)	Top metric	d'	Accuracy
Cr vs. In	Sdq	6.5	1.0
Cr vs. Pa	Sq	2.9	.89
Cr vs. PS	Sdr	3.7	.94
Cr vs. RS	Sdq	3.4	.95
Cr vs. Sc	Sdr	5.6	1.0
In vs. Pa	Sz	6.3	.96
In vs. PS	Sz	2.3	.88
In vs. RS	Sdr	5.6	1.0
In vs. Sc	Sp	3.9	.95
Pa vs. PS	Sz	3.8	.94
Pa vs. RS	Sq	4.4	.94
Pa vs. Sc	Sdr	4.9	.94
PS vs. RS	Sa	1.9	.84
PS vs. Sc	Ssk	3.1	.93
RS vs. Sc	Sdr	3.9	.98

3.5.2 Multiclass classification

Figure 3.8 depicts the d' matrix based on the twenty-two ISO metrics in Table 3.1 and the 15 pairwise surface comparisons constructed from the six aforementioned NEU-CLS surface classes. As noted in the previous section, the hybrid metrics Sdq and Sdr, appear to be among the better predictors; in all but four cases (Cr vs. Pa, In vs. Sc, PS vs. RS, and PS vs. Sc) they had d' values greater than 3. PS vs. RS is the only task that does not a have candidate metric with a d' greater than two, i.e. the entire column is gray, indicating no classifier is expected to perform well for this classification challenge.



	Cr_vs_In	Cr_vs_Pa	Cr_vs_PS	Cr_vs_RS	Cr_vs_Sc	In_vs_Pa	In_vs_PS	In_vs_RS	In_vs_Sc	Pa_vs_PS	Pa_vs_RS	Pa_vs_Sc	PS_vs_RS	PS_vs_Sc	RS_vs_Sc
Sa	2.8	2.8	0.13	2.9	1.7	4.1	2	0.52	0.8	2.6	4.1	3.6	1.9	1.3	0.46
Sq	3.1	2.9	0.32	3	0.98	4.5	2.1	0.68	1.3	2.9	4.4	3.2	1.8	0.59	0.91
Ssk	0.44	0.5	1.1	0.082	2.8	0.17	0.67	0.35	2.9	0.43	0.44	2.8	0.98	3.1	2.8
Sku	0.81	1.1	0.3	0.73	1.7	0.36	0.083	1.3	1.8	0.29	1.5	1.8	0.55	1.7	1.7
Sp	4.1	0.24	2.2	1.7	1.1	2.9	1	2.2	3.9	1.8	1.4	0.71	0.73	2.7	2.3
Sv	3.8	0.69	1.3	2.3	3.9	3	2.2	1.9	0.029	1.6	2.1	3	0.68	2.2	2
Sz	5.9	1.2	3.1	2.8	1.3	6.3	2.3	2.3	3.2	3.8	3.6	2.1	0.12	1.2	1.1
Sal	1.8	0.84	1.7	0.27	0.4	1.4	0.13	1.9	1.3	1.3	0.73	0.3	1.7	1.2	0.21
Str	0.62	0.19	0.69	0.16	0.52	0.36	0.016	0.44	1	0.4	0.046	0.62	0.49	1.1	0.63
Std	0.46	0.15	0.25	0.69	0.019	0.54	0.17	0.18	0.34	0.34	0.76	0.12	0.36	0.18	0.52
Sdq	6.5	0.54	3.6	3.4	4.6	5.8	2.1	5.6	1.8	3.1	2.8	4	1.1	0.58	2.1
Sdr	6.3	0.77	3.7	3.4	5.6	5.6	2	5.6	1.2	3	2.6	4.9	1.1	1.3	3.9
Smr1	0.19	0.41	0.45	0.57	0.081	0.26	0.5	0.49	0.21	0.61	0.59	0.41	0.21	0.33	0.24
Smr2	0.32	0.55	0.8	0.59	2.6	0.35	0.51	0.63	1.8	0.052	0.71	1.4	1	1.9	2.2
Sk	3.2	2.2	0.63	3.1	2.8	3.9	1.9	0.88	1.2	2.4	3.6	3.5	1.4	1.2	0.44
Spk	3.2	2.1	0.58	3.2	2.5	3.7	1.7	0.87	1.3	2.2	3.5	3.2	1.3	0.93	0.64
Svk	3.3	2.3	0.65	3	2.3	3.9	2	0.89	1.2	2.5	3.7	3.4	1.5	1.1	0.48
Sxp	3.4	1.3	1.2	2.7	0.24	2.9	1.3	1.1	1.6	1.9	2.5	0.72	0.67	0.82	1.2
Vvv	0.01	1.9	0.8	0.66	2.2	1.9	0.84	0.85	2.4	1.3	2.1	0.76	1.3	0.97	3.1
Vvc	0.19	0.048	0.0019	0.18	1.5	0.093	0.1	0.61	1.6	0.044	0.012	0.51	0.08	0.95	2
Vmp	0.62	1.5	0.83	1	0.24	1.5	1	0.46	0.52	1.1	1.6	1.4	1.1	0.56	0.64
Vmc	0.58	0.34	0.86	0.11	0.7	0.076	0.56	1.2	2	0.41	0.52	1	1.3	1.9	0.88

Figure 3.8: Illustration of d' matrix based on the NEU-CLS dataset and the twenty-two ISO 25178-2 parameters. Columns are indexed by different binary classification tasks with rows indexed by ISO metric. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Figure 3.9 depicts the probability density estimations for selected height (*Sa*, *Sq*, *Sv*, *Sp*, *Sz*, *Ssk*, *Sku*) and hybrid parameters (*Sdr*, *Sdq*), for the six defect classes defined by NEU-CLS dataset, i.e. those that produced among the highest d' values for each classification pair. For viewing clarity, the metric values have been normalized, and the y-axis, representing the density estimation, removed. In nearly all distributions unequal variances exist, with *Sdr* distribution for the crazing defects in the bottom of Fig. 3.9 actually exhibiting a binomiallike distribution. Notice that *sa* and *Sq* are highly correlated metrics, as are *Sdq* and *Sdr*. Since their mathematical definitions are very similar, it is not surprising that their respective distributions are very similar in nature, i.e. compare the top two rows for *Sa* and *Sq*, and the bottom two rows for *Sdq* and *Sdr*. Using a few of the selected metrics to intuitively analyze the different surface types, both scratches (Sc) and inclusion (In) classes exhibit low ââintensity roughnessââ given by the *Sa* parameter, relative to the surfaces with patches (Pa) (where the ratio of light to dark areas is closer to one), spanning a larger intensity range on average. Inclusion (In) surfaces have low *Sdr* and *Sdq* values, indicative of a ââflatterââ and more uniform intensity image compared to crazing (Cr) surfaces which display more complex variations in composition and texture. Scratched (Sc) surfaces exhibit distinct bright streaks creating an asymmetry between the intensity values that represent the surface. Notice skewness *Ssk* (third row from top) can characterize this asymmetrical appearance as a distinct range of quantitative values (i.e. purple distribution) relative to the nominal *Ssk* ranges of the other surface types.

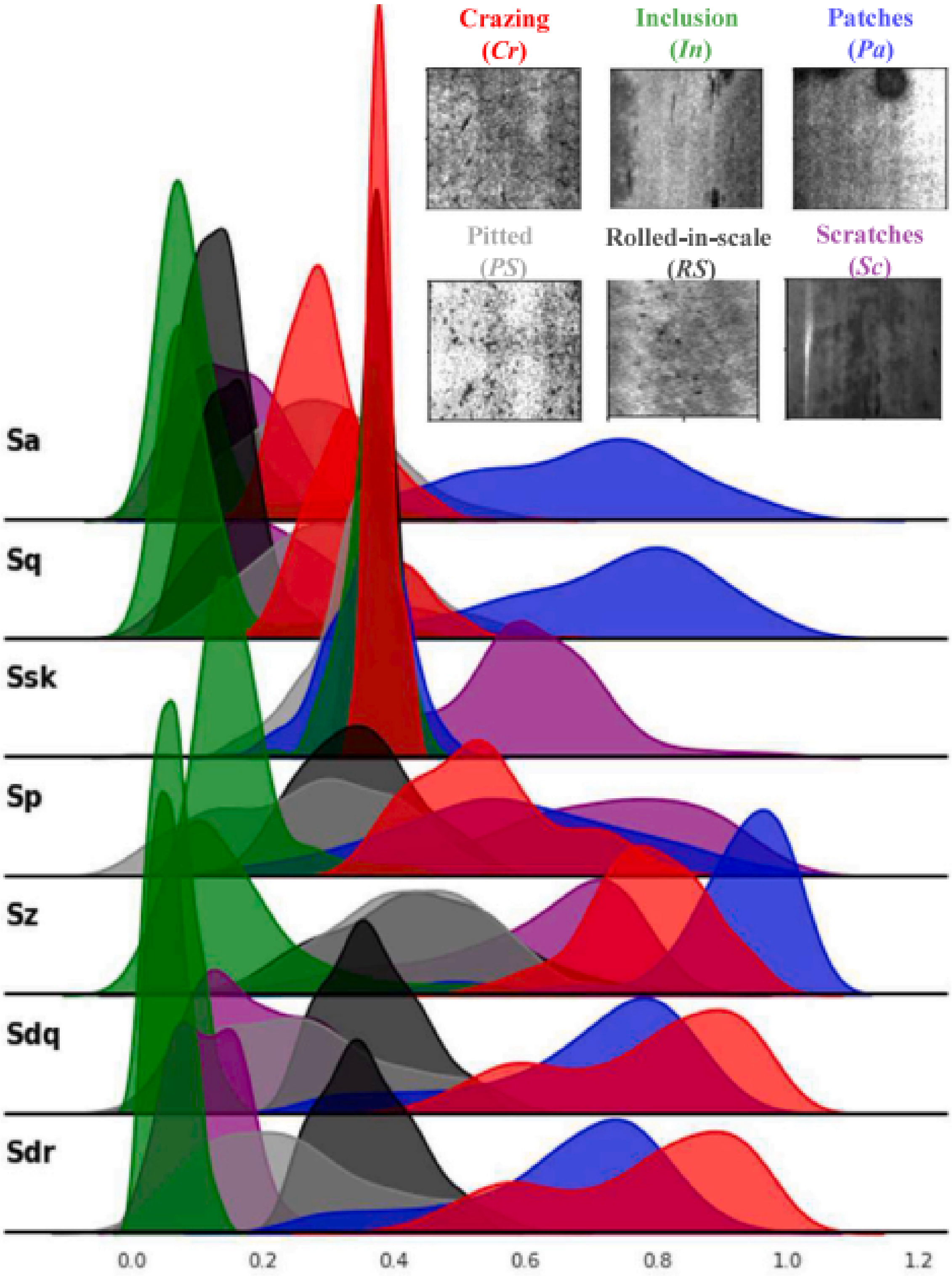


Figure 3.9: Density functions of ISO 25178-2 parameters selected by Algorithm 1 based on d' matrix for NEU-CLS dataset. Parameter values were normalized to a range of 0-1 for visualization, the y-axis corresponds to density estimation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As noted earlier, PS vs. RS did not have any d' values higher than 2 in the d' matrix. The In vs. PS task also has no metrics with a d' value greater than 2.3. Examination of the seven plots in Fig. 9 clearly shows that the distributions for each metric significantly overlap for both PS (light gray) and RS (dark gray) classes. The same is true for In vs. PS. To reliably differentiate between these surface classes, it is necessary to explore additional filtering operations and/or alternative ISO metrics. In comparison d' values of 6.5, 6.3, and 5.9 are recorded for Cr vs. In for metrics *Sdq*, *Sdr*, and *Sz* respectively, hence little to no overlap between these distributions appear in the bottom three rows of Fig. 3.9. It is expected that any reasonably optimized machine-learning model which uses these metrics to make predictions would rarely, if ever, misclassify crazing defects (Cr) for inclusion (In) and vice versa.

3.5.3 Machine learning for multiclass problems

Based on a 0.5/0.5 train/test split of the NEU-CLS dataset, 900 images (150 per class) for training and 900 images for testing, a decision tree model trained on all twenty-two metrics reported in Table 1 yielded a maximum test accuracy of 95.19% and took approximately 3 s to process a 200 × 200-pixel image. Using only metrics with d' less than or equal to 2 to train the model results in a test accuracy of 80.56%. In comparison, a decision tree trained using the seven metrics (*Sa*, *Sq*, *Ssk*, *Sp*, *Sz*, *Sdq*, *Sdr*) isolated by the outlined approach, the lowest d' value being 1.9, achieved a test accuracy of 95.72% and took only 6 ms to process new images. Processing time is further reduced by approximately 1 ms for each additional metric that is withdrawn.

Removing redundant, i.e. highly correlated parameters, Sdq and Sq, and the skewness metric Ssk, the decision tree model trained using the remaining four metrics Sa, Sp, Sz, Sdr achieved 94.07% test accuracy. Using only three metrics Sa, Sp, and Sdr, still provided a respectable test accuracy of 92.78%.

Figure 3.10 depicts how the test accuracy (y-axis) of different machine learning classifiers vary according to different training/testing splits. In all cases the seven metrics (Sa, Sq, Ssk, Sp, Sz, Sdq, Sdr) were used as inputs to the model(s). For each split, the dataset is shuffled and the training and evaluation process is performed a total of five times. The error bars represent the standard deviation of the recorded test accuracy. According to Fig. 3.10, maximizing the proportion of training examples tends to increase classification performance in most cases. However, reducing the number of examples withheld for testing makes it progressively more difficult to assess how a model may perform when classifying unseen examples. Even with the challenging 0.1/0.9 split, where only 30 out of 300 images per class are used for training and the remaining 1620 images are used for testing, the suggested approach achieves accuracies exceeding 89%. This demonstrates noteworthy sample efficiency, which is highly desirable in many manufacturing environments. For the 0.9/0.1 split where 270 out of 300 images per class are used for training and the remaining 180 images for testing, accuracies greater than 96% are reached. The decision tree classifier consistently outperforms other models when averaged across all train/test splits. In comparison, the SVM classifier exhibits significantly lower performance, with an average reduction in accuracy of approximately 20% compared to the decision tree. The Bayes classifier shows slightly better performance for the 0.2/0.8 and 0.1/0.9 train/test splits, but there is no noticeable improvement as the ratio of training samples is increased. On the other hand, the KNN classifier demonstrates considerable performance improvement, with accuracies increasing from 78% to 90% for the 0.1/0.9 and 0.9/0.1 train/test splits, respectively.

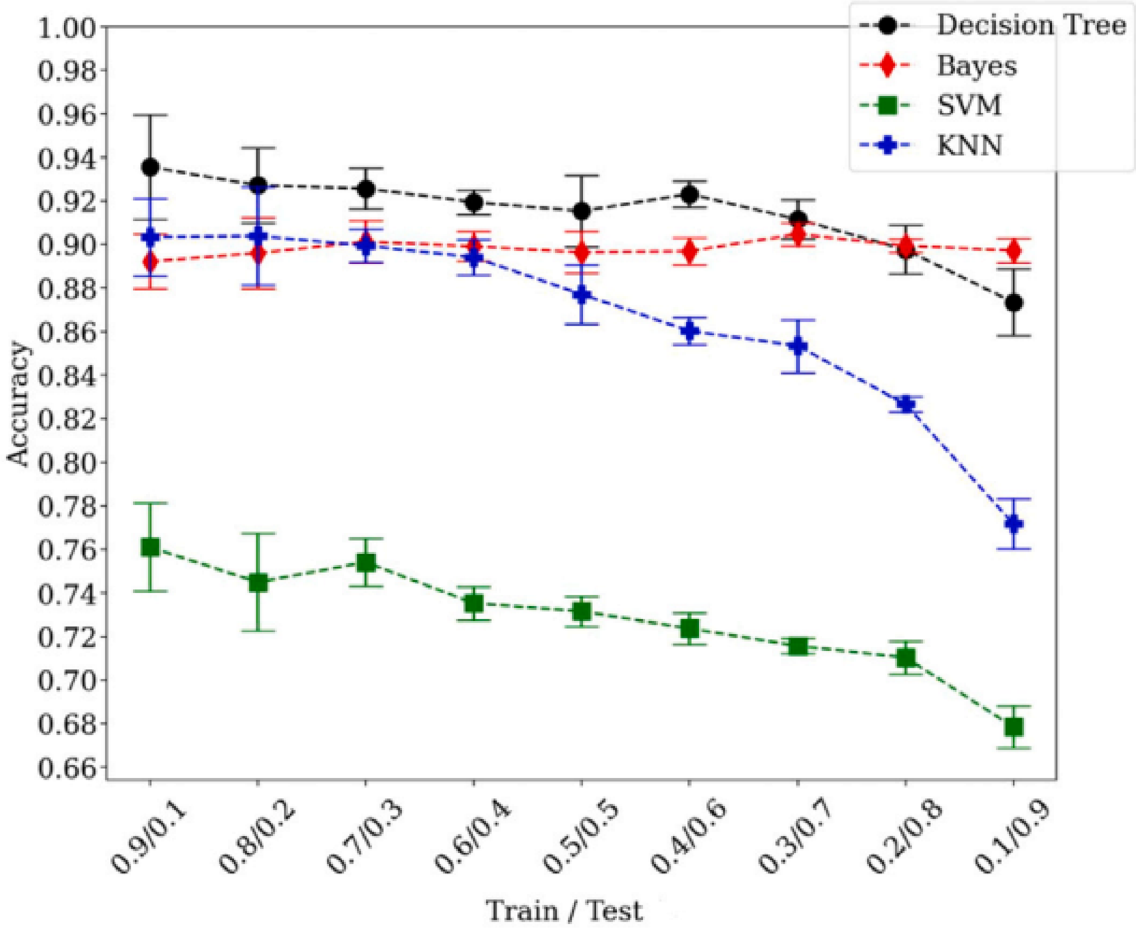


Figure 3.10: Test accuracy of different machine learning classifiers for different train/test splits of the NEU-CLS dataset using seven ISO metrics (Sa, Sq, Ssk, Sp, Sz, Sdq, Sdr) isolated by Algorithm 1.

The average and standard deviation of the number of correct and incorrect predictions on the testing sets made by the decision tree classifier for the 0.5/0.5 split shown in Fig. 3.10 is depicted by the modified confusion matrix in Fig. 3.11. Correct predictions, colored in green, correspond to the average number of instances that the predicted label matched the true label for images in the testing set. The off-diagonal elements colored in red represent instances of incorrect predictions. If all of the test samples were classified correctly each time the training and testing data was shuffled then the confusion matrix in Fig. 3.11 would be a diagonal matrix. The intuition described in Section 3.3, which is empirically quantified in Fig. 3.6, suggests that a diagonal confusion matrix is only expected when each classification task (i.e. Cr vs. In, Cr vs. Pa, RS vs. Sc, etc.) is characterized by at least one ISO metric with a d' greater than seven.

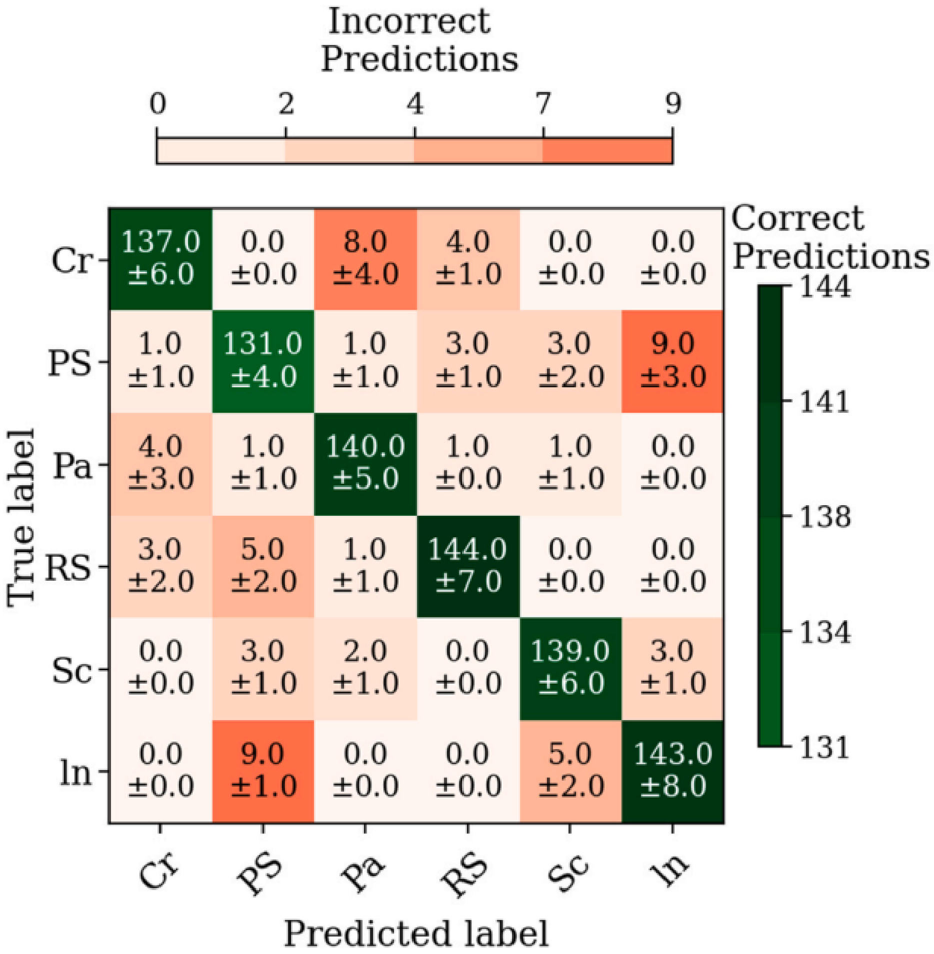


Figure 3.11: Modified confusion matrix representing average and standard deviation of correct and incorrect predictions on the test set after five independent evaluations of the decision tree classifier using seven ISO metrics (Sa, Sq, Ssk, Sp, Sz, Sdq, Sdr) isolated by Algorithm 1. For each evaluation, 150 samples per class are used for training and 150 per class are reserved for testing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Based on Fig. 3.11, there were, on average, a total of 18 instances where surfaces with inclusion (In) were incorrectly classified as pitted surfaces (PS), and vice versa. The d' matrix in Fig. 3.8 supports these findings, indicating that the highest-ranking

ISO metric, Sz, achieved a d' value of 2.3 for the In vs. PS task. Similar trends were observed for other tasks: an average of twelve instances for Cr vs. Pa with Sq achieving a d' value of 2.9, eight instances for In vs. Sc with Sp having a d' value of 3.9, and three instances for Sc vs. Pa with Sdr achieving the highest d' value of 4.9. No incorrect predictions were recorded for Cr vs. In or In vs. Pa. The d' matrix reported Sdq as the highest-ranking ISO metric with a d' value of 6.5 for the former task, and Sz with a value of 6.3 for the latter. Deviations from the trend were observed. For example an average of eight instances of incorrect predictions between PS and RS with Sa achieving a d' of 1.9 and an average of one instance between PS and Cr with Sdr having a value of 3.7. However, the average accuracy of the decision tree for each task was always higher than the conservative estimate obtained from the d' vs. accuracy curve(s) in Fig. 3.6. This is because a reasonably optimized decision tree tends to outperform the curveâs criterion of $c = \frac{\mu_A+\mu_B}{2}$.

3.5.4 Comparison to other approaches

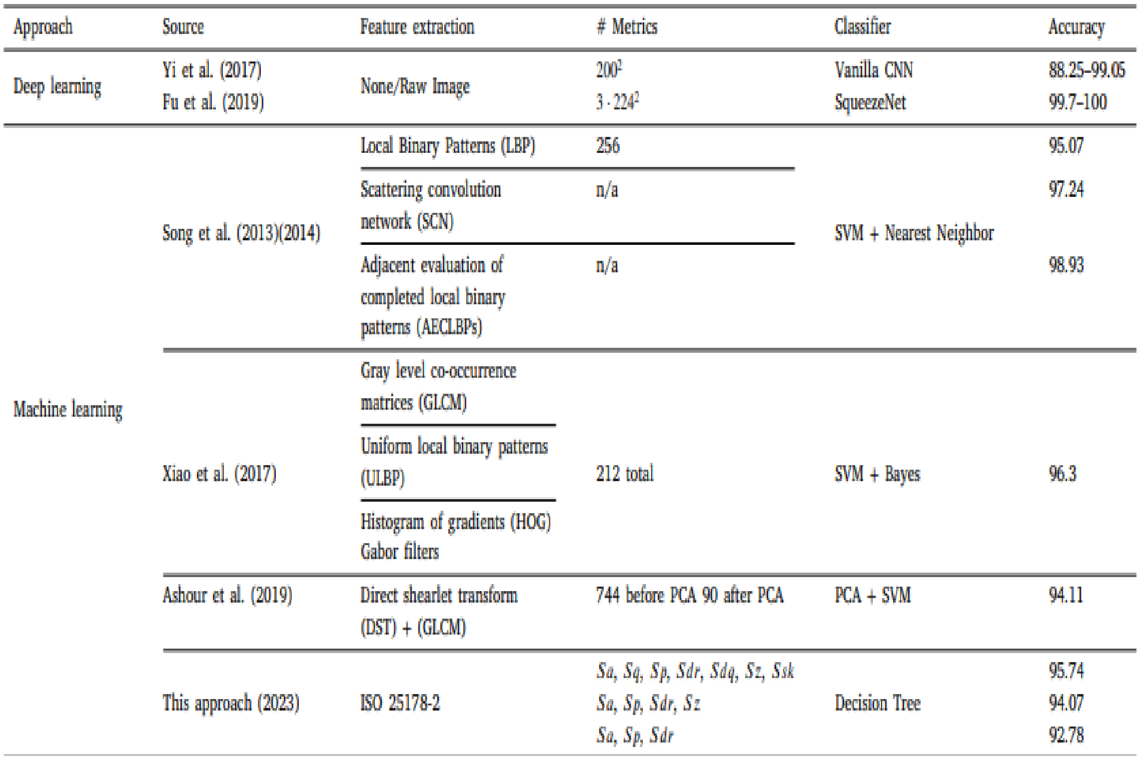
Following the release of the NEU-CLS dataset, several approaches have been explored for automating the classification of the six classes. Table 3.3 summarizes a range of metric extraction techniques and classification models applied to the NEU-CLS found in published literature. For comparison, the result from this work is listed in the last row of the table. Yi et al. [15] demonstrated that a custom CNN deep-learningbased architecture could achieve a test accuracy of 99.05%. However, to achieve this the authors required extensive data augmentation (i.e. artificially increasing the size of a training dataset by applying various transformations to the original data, such as flipping, rotating, cropping, or adding noise) to artificially inflated the training dataset. Without these additional processing steps, accuracy drops to 88.25%. Their reported accuracy is also based on a 0.9/0.1 train-test split, unlike the conventional 0.5/0.5 train-test benchmark introduced by [5]. The perfect 0.8/0.2 train-test result reported by Fu et al. [16] also utilized augmentation and required the input images to be converted into pseudo color images of size 224 × 224 × 3. These types of dataset and image modifications are ubiquitous among reported state-of-the-art deep learning approaches for the NEU-CLS dataset.

Three approaches listed in Table 3.3 utilized machine learning. Song et al. [5,14] combined machine learning with features resulting from local binary patterns (LBP) [48], scattering convolution network (SCN), or adjacent evaluation of completed local binary patterns (AECLBP) image processing techniques to achieve classification accuracies be-tween 95.07% to 98.93%. Xioa et al. [7] combined machine learning with an ensemble of image processing techniques for feature extraction including gray-level co-occurrence matrices (GLCM) [49], uniform local binary patterns (ULBP), histogram of gradients (HOG) [50,51], and Gabor filters resulting in 153 descriptions for each image. The approach achieves high classification performance with an accuracy of 98.24%. Ashour et al. [6] pre-processed the images using direct shearlet trans-forms (DST) [52] and GLCM transforms and then distilled the resulting 744 image features to 90 using principal component analysis (PCA) prior to classification of the surface image to achieve 94.11% accuracy. As with the deep learning approaches, no systematic or generalized framework is outlined to inform the reader why the pre-processing route and chosen metrics work. This limits the readers' ability to assess whether such approaches would translate to their specific application.

3.6 Final summary and conclusions

This paper outlines a systematic, SDT-based framework to assess a metric's ability to differentiate between two different surface classes, or indeed multiple different surfaces. A metric's d' value for two surface types provides *a priori* insights into how well a classification model could perform if a particular surface metric were used. This offers opportunities to significantly reduce the amount of effort required to select appropriate metrics and optimize a classification algorithm. The d' calculation is sufficiently robust to be applied to non-Gaussian datasets. To demonstrate its effec-

Table 3.3: Summary of different approaches for classifying the six types of surfaces included in the NEU-CLS dataset



Approach	Source	Feature extraction	# Metrics	Classifier	Accuracy
Deep learning	Yi et al. (2017)	None/Raw Image	200^2	Vanilla CNN	88.25–99.05
	Fu et al. (2019)		$3 \cdot 224^2$	SqueezeNet	99.7–100
Machine learning	Song et al. (2013)(2014)	Local Binary Patterns (LBP)	256	SVM + Nearest Neighbor	95.07
		Scattering convolution network (SCN)	n/a		97.24
		Adjacent evaluation of completed local binary patterns (AECLBPs)	n/a		98.93
	Xiao et al. (2017)	Gray level co-occurrence matrices (GLCM)	212 total	SVM + Bayes	96.3
		Uniform local binary patterns (ULBP)			
		Histogram of gradients (HOG) Gabor filters			
	Ashour et al. (2019)	Direct shearlet transform (DST) + (GLCM)	744 before PCA 90 after PCA	PCA + SVM	94.11
	This approach (2023)	ISO 25178-2	Sa, Sq, Sp, Sdr, Sdq, Sz, Ssk	Decision Tree	95.74
			Sa, Sp, Sdr, Sz		94.07
			Sa, Sp, Sdr		92.78

tiveness, an empirical relationship between d' values and classification accuracy was constructed using twenty-two ISO metrics and over 15 different binary datasets. Furthermore, a comprehensive evaluation of the performance of the ISO 25178-2 metrics and a decision tree for classifying the six types of surface classes in the NEU-CLS dataset are reported and compared to existing image processing and deep learning approaches. The discrimination matrix and automated metric selection method based on d' was found to be effective in isolating relevant ISO metrics, reducing the number used for classification from twenty-two to seven without decreasing accuracy. A decision tree model that utilized only these seven ISO metrics achieved accuracies exceeding 89% for a 0.1/0.9 train/test split and 96% for a 0.9/0.1 train/test split. Using an even smaller subset of the seven metrics including Sdr, Sa, and Sp achieved 92.78% accuracy on a 0.5/0.5 train/test split and reduced processing time from 166 to 332 frames per second. This speed makes the approach suitable for high-speed in-situ and/or online surface inspection tasks. To the best of our knowledge, this is the first report involving the use of ISO 25178-2 parameters to characterize defects on NEU-CLS dataset or other collections of intensity-based images. The reasoning for their use (as opposed to conventional image processing metrics) is because of their deep linage in manufacturing environments, in addition to their standardized equations and definitions that provide intuitive characterizations of a surface's topography, regardless of the units. Combining the metrics isolated by the d' matrix with a decision tree offers the opportunity for easier interpretation compared to black box outputs of other machine and deep learning approaches. This can be particularly important in a manufacturing setting where it is highly desirable to have traceable decisions and the ability to explain the reasons for (or assign numerical definitions to) defects or conformance specifications. Knowing the d' values of the metrics a classifier uses provides an indication of the level of confidence or skepticism one should have in relation to its predictions. A machine learning model that utilizes multiple metrics with low d' values is not expected to make accurate or reliable classifications. Another advantage of this approach is the ability to adapt and preserve the original input image resolution. ISO metrics can be computed on images of any size and aside from standard form removal and optional digital filtering methodologies, no numerical or dimensional resizing of the image is required prior to classification. Moreover, the suggested d' matrix could be extended to any other univariate metrics, such as profile metrics defined by ISO 4287 [53] or other image processing metrics. The proposed methodology is widely applicable to other classification tasks, such as surfaces grouped by functionalities such as wettability, aesthetic appearance, or haptic sensation. Investigations like these will be carried out and reported in future work to better understand the relationships between surface texture and function.

3.7 Acknowledgments

This work was funded by the Center for Precision Metrology (CPM) at the University of North Carolina at Charlotte.

3.8 References

[1] Luo Q, Fang X, Su J, Zhou J, Zhou B, Yang C, Liu L, Gui W, Tian L, "Automated visual defect classification for flat steel surface: a survey," *IEEE Trans Instrum Meas*, vol. 69, no. 12, pp. 9329–49, 2020.

[2] Mullany B, Savio E, Haitjema H, Leach R, "The implication and evaluation of geometrical imperfections on manufactured surfaces," *CIRP Ann*, vol. 71, no. 2, pp. 717–39, 2022.

[3] Rosen B-G, Eriksson L, Bergman M, "Kansei, surfaces and perception engineering," *Surf Topogr: Metrol Prop*, vol. 4, no. 3, pp. 033001, 2016.

[4] Brown CA, Hansen HN, Jiang XJ, Blateyron F, Berglund J, Senin N, Bartkowiak T, Dixon B, Le Goic G, Quinsat Y, et al., "Multiscale analyses and characterizations of surface topographies," *CIRP Ann*, vol. 67, no. 2, pp. 839–62, 2018.

[5] Song K, Yan Y, "A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects," *Appl Surf Sci*, vol. 285, pp. 858–64, 2013.

[6] Ashour MW, Khalid F, Abdul Halin A, Abdullah LN, Darwish SH, "Surface defects classification of hot-rolled steel strips using multi-directional shearlet features," *Arab J Sci Eng*, vol. 44, no. 4, pp. 2925–32, 2019.

[7] Xiao M, Jiang M, Li G, Xie L, Yi L, "An evolutionary classifier for steel surface defects with small sample set," *EURASIP J Image Video Process*, vol. 2017, no. 1, pp. 1–13, 2017.

[8] Chondronasios A, Popov I, Jordanov I, "Feature selection for surface defect classification of extruded aluminum profiles," *Int J Adv Manuf Technol*, vol. 83, pp. 33–41, 2016.

[9] Yang S-W, Lin C-S, Lin S-K, Chiang H-T, "Automatic defect recognition of TFT array process using gray level co-occurrence matrix," *Optik*, vol. 125, no. 11, pp. 2671–6, 2014.

[10] Shanmugamani R, Sadique M, Ramamoorthy B, "Detection and classification of surface defects of gun barrels using computer vision and machine learning," *Measurement*, vol. 60, pp. 222–30, 2015.

[11] Tao X, Zhang Z, Zhang F, Xu D, "A novel and effective surface flaw inspection instrument for large-aperture optical elements," *IEEE Trans Instrum Meas*, vol. 64, no. 9, pp. 2530–40, 2015.

[12] Smith ML, Smith G, Hill T, "Gradient space analysis of surface defects using a photometric stereo derived bump map," *Image Vis Comput*, vol. 17, no. 3–4, pp. 321–32, 1999.

[13] Luo Q, Fang X, Liu L, Yang C, Sun Y, "Automated visual defect detection for flat steel surface: A survey," *IEEE Trans Instrum Meas*, vol. 69, no. 3, pp. 626–44, 2020.

[14] Song K, Hu S, Yan Y, "Automatic recognition of surface defects on hot-rolled steel strip using scattering convolution network," *J Comput Inf Syst*, vol. 10, no. 7, pp. 3049–55.

[15] Yi L, Li G, Jiang M, "An end-to-end steel strip surface defects recognition system based on convolutional neural networks," *Steel Res Int*, vol. 88, no. 2, pp. 1600068, 2017.

[16] Fu G, Sun P, Zhu W, Yang J, Cao Y, Yang MY, Cao Y, "A deep-learning-based approach for fast and robust steel surface defects classification," *Opt Lasers Eng*, vol. 121, pp. 397–405, 2019.

[17] Jansons E, Puganovs G, Panina LV, Kovalenko I, "Tribological and mechanical properties of the nanostructured superlattice coatings with respect to surface texture," *Lubricants*, vol. 10, no. 11, p. 285, 2022.

[18] Das J, Linke B, "Evaluation and systematic selection of significant multi-scale surface roughness parameters (SRPs) as process monitoring index," *J Mater Process Technol*, vol. 244, pp. 157–65, 2017.

[19] Å»ak K, Grzesik W, "Metrological aspects of surface topographies produced by different machining operations regarding their potential functionality," *Metrol Meas Syst*, vol. 24, no. 2, pp. 325–35, 2017.

[20] Grzesik W, "Prediction of the functional performance of machined components based on surface topography: state of the art," *J Mater Eng Perform*, vol. 25, no. 10, pp. 4460–8, 2016.

[21] Song K, Yan Y, "Northeastern university surface defect database," 2012.

[22] Puntous T, Pavan S, Delafosse D, Jourlin M, Rech J, "Ability of quality controllers to detect standard scratches on polished surfaces," *Precis Eng*, vol. 37, no. 4, pp. 924–8, 2013.

[23] Gruber DP, Berger G, Pacher G, Friesenbichler W, "Novel approach to the measurement of the visual perceptibility of sink marks on injection molding parts," *Polym Test*, vol. 30, no. 6, pp. 651–6, 2011.

[24] Blateyron F, "The areal field parameters," in: *Characterisation of areal surface texture*, Springer, 2013, pp. 15–43.

[25] Pawlus P, Reizer R, Wieczorowski M, "Functional importance of surface texture parameters," *Materials*, vol. 14, no. 18, p. 5326, 2021.

[26] ISO 25178-2:2012: Geometric Product Specification (GPS). Surface texture: Areal. Part 2: Terms, definitions and surface texture parameters, 2012, [Online; accessed February 2023]. URL `https://www.iso.org/standard/52019.html`.

[27] ISO. Geometrical product specifications (GPS)âSurface texture: ArealâPart 2: Terms, definitions and surface texture parameters, 2021, URL `https://www.iso.org/standard/74591.html`.

[28] Bui S. NIST - Internet based Surface Metrology Algorithm Testing System, URL `https://physics.nist.gov/3DVSC/jsp/FileOpen3D.jsp`.

[29] Czifra Ã, BarÃ¡nyi I. SDQ-SDR topological map of surface topographies. *Front Mech Eng* 2020;6:50, URL `https://www.frontiersin.org/articles/10.3389/fmech.2020.00050/full`.

[30] Das A, Geisler WS. A method to integrate and classify normal distributions. *J Vis* 2021;21(10):1, URL `https://arxiv.org/pdf/2012.14331.pdf`.

[31] See JE. Visual inspection: a review of the literature, 2012, URL `https://www.osti.gov/servlets/purl/1055636`.

[32] Craig A. Human engineering: The control of vigilance. In: Sustained attention in human performance. Chichester, UK: Wiley; 1984, p. 247-91.

[33] Green DM, Swets JA, et al. Signal detection theory and psychophysics, Vol. 1. New York: Wiley; 1966.

[34] Whitehouse D. The parameter rashâis there a cure? *Wear* 1982;83(1):75-8.

[35] learn contributors S. Decision trees, 2021, [Online]. Available: `https://scikitlearn.org/stable/modules/tree.html#tree-mathematical-formulation`.

[36] Bramer M. Avoiding overfitting of decision trees. In: Principles of data mining. Springer; 2007, p. 119-34.

[37] Mingers J. An empirical comparison of pruning methods for decision tree induction. *Mach Learn* 1989;4:227-43.

[38] Redford J. Surface quality and inspection descriptions, 2022, `https://jesseredford-squid-main-otrfh7.streamlit.app/`, [Accessed: Feb. 23, 2023].

[39] Redford J. Surface quality and inspection descriptions, 2022, `https://github.com/Jesse-Redford/SQuID.git`, [Accessed: Feb. 23, 2023].

[40] learn contributors S. `sklearn.tree.DecisionTreeClassifier`. scikit-learn; 2021, [Online]. Available: `https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html`.

[41] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. 2011-, URL `https://scikit-learn.org/stable/index.html`.

[42] Cortes C, Vapnik V. A training algorithm for optimal margin classifiers. *Mach Learn* 1995;20(3):273-97.

[43] Cover T, Hart P. Pattern classification and scene analysis. *IEEE Trans Inform Theory* 1967;13(1):21-7.

[44] Rish I. The optimality of naive Bayes. In: Proceedings of the 17th international florida artificial intelligence research society conference (FLAIRS 2004). 2004, p. 196-200.

[45] scikit-learn contributors. `sklearn.svm.SVC`. 2023, `https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html` (Accessed: July 7, 2023).

[46] scikit-learn contributors. `sklearn.neighbors.KNeighborsClassifier`. 2023, `https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html` (Accessed: July 7, 2023).

[47] scikit-learn contributors. `sklearn.naive_bayes.GaussianNB`. 2023, `https://scikitlearn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html` (Accessed: July 7, 2023).

[48] Ojala T, PietikÃ€inen M, MÃ€enpÃ€Ã€ T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 2002;24(7):971-87.

[49] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3(6):610-21.

[50] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPRâ05), Vol. 1. IEEE; 2005, p. 886-93.

[51] McConnell RK. Method of and apparatus for pattern recognition. 1986.

[52] Guo K, Labate D, Lim W-Q, Weiss G. A direct algorithm for the construction of compactly supported shearlet frames. *IEEE Trans Image Process* 2007;16(6): 1720-36.

[53] ISO. Geometrical product specifications (GPS)-Surface texture: Profile method-Terms, definitions and surface texture parameters. 1997.

CHAPTER 4: (Paper 2) CLASSIFICATION OF VISUAL SMOOTHNESS STANDARDS USING MULTI-SCALE AREAL TEXTURE PARAMETERS AND LOW-MAGNIFICATION COHERENCE SCANNING INTERFEROMETRY

Overview of paper 2

The work in this chapter has been published in the Journal of Materials [7]. To further assess the framework introduced in Chapter 3 while also addressing multi-scale aspects of surface characterization, in this chapter SQuID^{TM} is leveraged to classify different grades of surface finish appearances. This chapter also provides an end-to-end example of the SQuID^{TM} workflow incorporating data collection, dataset preparation, processing, and assessment of the results. In this work, ISO 25178-2 areal surface metrics extracted from bandpass filtered measurements of a set of ten visual smoothness standards obtained from low magnification coherent scanning interferometry are used to quantify different grades of powder-coated surface finish. The ability to automatically classify the standard tiles using multi-scale areal texture parameters is compared to parameters obtained from a hand-held gloss meter. The results indicate that the ten different surface finishes can be automatically classified with accuracies as low as 65% and as high as 99%, depending on the filtering and parameters used to quantify the surfaces. The highest classification accuracy is achieved using only five multi-scale topography descriptions of the surface. In this case, spatial and hybrid parameters were selected over commonly prescribed height parameters such as Sa, which proved ineffective in characterizing differences between the surface grades. It was also found that standard appearance parameters provided by the RhoPoint-IQ gloss meter successfully differentiated only four of the ten tiles, underscoring the superior accuracy of areal surface texture measurements in linking topography to subjective visual grading. Furthermore, the results demonstrate how the developed framework can be leveraged to automate visual inspection tasks, proving the means to not only select the required parameters that enable successful classification but also provide the explicit decision logic used to determine the grade(s) of the surfaces.

The contributions of this chapter include: 1) a benchmark assessment of the SQuID^{TM} framework on challenging multi-classification tasks; 2) a comprehensive evaluation and comparison of texture parameters and gloss parameters for classifying visual smoothness standards; 3) the derivation of a unique multiscale description and decision logic for classifying a set of ten visual smoothness standards.

Abstract

The ability to objectively specify surface finish to ensure consistent visual appearance addresses a vital need in surface coating engineering. This work demonstrates how a computational framework, called surface quality and inspection descriptors (SQuID^{TM}), can be leveraged to effectively rank different grades of surface finish appearances. ISO 25178-2 areal surface metrics extracted from bandpass-filtered measurements of a set of ten visual smoothness standards taken on a coherent scanning interferometer are used to quantify different grades of powder-coated surface finish. The ability to automatically classify the standard tiles using multi-scale areal texture parameters is compared to parameters obtained from a hand-held gloss meter. The results indicate that the ten different surface finishes can be automatically classified with accuracies as low as 65% and as high as 99%, depending on the filtering and parameters used to quantify the surfaces. The highest classification accuracy is achieved using only five multi-scale topography descriptions of the surface.

4.1 Introduction

The surface texture can impact a manufactured surfaceâs functionality and aesthetic quality. Some examples of where surface texture affects functional performance include wettability [1-3], reflectance [4,5], corrosion resistance [6,7], fatigue [7,8], and heat exchange [9,10]. While the surface aesthetic (visual and tactile) includes human perception and thus encroaches on the psychophysics domain [11-14], the topographical amplitudes and spatial relationships between different features on the surface, along with other traits such as color, thermal conductivity, and modulus, affect how the surface is perceived and valued [12,13,15-18]. Central to understanding how surface texture affects functionality and aesthetics is to isolate metrics capable of quantifying relevant surface topographies at the appropriate length scales. Similarly, quantitative metrics are needed to assist in fine-tuning processes used to generate the surfaces and for subsequent process control and part acceptance. That said, the identification of appropriate qualitative metrics is far from trivial [19,20]. While surfaces can be easily measured in digital environments, merely assigning a numerical assessment to a surface that does not correlate with its intended function provides incomplete and potentially misleading information. Additionally, while visual comparisons to defined standards are utilized in many instances, especially regarding aesthetics, human visual inspection (HVI) processes are inherently subjective, as discussed in [11]. There is a strong desire to replace HVI with automated systems and a longstanding goal of definitively linking functionality to sur-face topographies. However, there remains a need to make it easier to find suitable metrics that can reliably discriminate between surfaces grouped or ranked by functionality, subjective appearance, or, indeed, processing-induced differences [21-23].

4.1.1 Challenges in quantifying surface quality

Information on surface quality can be obtained by multiple modes of acquisition and processing, giving different outputs. For example, functional gloss meter readings quantify a surfaceâs specular reflectance and/or the directional distribution of light reflected by surfaces illuminated under specified conditions [24]. The gloss unit (GU) reports the amount of luminous flux reflected from a specimen to that of a reference glass tile with a known refractive index (1.567) having a specular reflectance of 100 GU at a specified angle and wavelength (λ = 587.6 nm or 546.1 nm) [25]. Typical gloss meter systems are designed around standard measurement angles of 20° for high-gloss surfaces, 60° for semi-gloss surfaces, and 85° for matte surfaces [25,26]. Although multiple surface characteristics influence gloss values derived from these measurement angles, they may fail to characterize certain aspects that reduce appearance quality (haze, orange peel, etc.). As such, additional appearance terminology and parameters characterizing surface quality, such as the distinctness of image DOI, the "aspect of gloss charac-terized by the sharpness of images of objects produced by reflection at a surface", have been developed [27] and used to quantify the severity of orange peel. DOI values close to 100 indicate a very high sharpness, and values close to 0 indicate a very high induced waviness in the reflected image [28]. Other industry standards have been developed by Rho-Point, two of which are Rspec for smoothness quantification and the reflected image quality RIQ parameter [29]. On the other hand, conventional imaging techniques provide intensity images of how light interacts with the different spatial regions of the surface. While fast and effective in providing insights into the nature of the surface texture, they are sensitive to lighting conditions and lack topographical height data to provide full three-dimensional representations. Conventional image-processing metrics used to quantify surface textures, such as grayscale histograms, local binary patterns (LBP), and gray-level co-occurrence matrices (GLCMs), are often difficult to interpret with respect to the physical nature of the surface and thus have limited value in understanding what drives a particular functionality. As such, they will not be the focus of this paper. A surfaceâs composition and geometric topography ultimately dictate its functional performance. Tactile or optical-based instruments can provide either line or areal height information pertaining to the geometric nature of the surface topography [30]. The speed at which a measurement is taken will vary depending on the system, with coherence scanning interferometry (CSI) being among the fastest but still slower than high-speed intensity imaging systems. Datasets from these systems can provide a wealth of geometrical topographical data with the scale of resolvable features dependent on system-limiting factors, tip radii for tactile measurements, and numerical aperture considerations for optical systems [31]. While the wealth of the available data is advantageous, effectively processing the data to give useful metrics still poses challenges [11,21]. Statistical parameters used to describe surfaces within the manufacturing community include height, spatial, hybrid, functional, and other parameters detailed in the ISO 25178-2 standard [32]. Even though there is a shift away from relying on two of the most commonly utilized areal metrics [33], the arithmetic average roughness, Sa, and the root mean square roughness, Sq, they remain prevalent despite having potentially similar values for surfaces that drastically differ in their topographical arrangements. Using other standardized parameters is promising, but questions about how alternative parameters should be selected for a specific application still arise. Surface geometries can be complicated; using only one or a small finite number of statistical parameters may fail to provide a full description. Increasing the number of parameters used can result in a more accurate description [34], but excessive reporting of redundant and/or irrelevant parameters leads to what Whitehouse coined in 1982 as "parameter rash" [35]. Further complicating the selection process, surface texture and its relationship to different functions is inherently multi-scale [22]. Single or multiple parameters calculated only at one specific spatial bandwidth may not always provide an adequate description. Even for the same surface, texture parameter values can change by orders of magnitude depending on the definition area of the measurement, sampling interval, form-removal operations, and cutoff frequencies used for high-, low-, or band-pass filtering of the data. If the search space is left unbounded, the number of parameter and length scale combinations are seemingly infinite. While particular combinations of areal parameters may provide a strong foundation for correlating process variations, function, or subjective visual appearance with statistical representations of the surfaceâs topography, it is crucial to understand that a heuristic approach to parameter selection might not always yield the most useful surface descriptions for a particular application. Objective and data-driven methodologies should complement heuristic approaches to ensure more accurate and effective surface characterizations, especially when aiming to capture subjective aspects of surface quality. However, using black-box machine learning (ML) models to carry out data-driven methodologies for classifying surface quality can limit the ability of designers, manufacturers, and researchers alike to use texture parameters to unambiguously communicate surface quality expectations and specifications for in-spection. Therefore, achieving a balance between traditional methods and AI-driven approaches is essential. Integrating the strengths of both can lead to more robust and effective surface quality assessments while addressing the limitations of each approach.

4.1.2 Paper description and organization

To address the challenges of selecting the relevant length scale and parameters to reliably differentiate between different surface texture classes, this paper outlines an approach demonstrated on a set of ten ranked visual smoothness panels available from the Powder Coating Institute (PCI). These panels are commonly used in industry for HVI of powder coatings. Currently, there is no clear pathway for automatically classifying a workpiece finish as one of the ten PCI textured samples. The ASTM D3451-06 standard [36] for testing powder coatings notes that subjective HVI comparison is the most common method to quantify differences in the surface profile, primarily the levels of reflectance and orange peel of cured powder coatings. A second method refers to a portable instrument that, when scanned across the surface, acts like the human eye and detects differences in reflectance (light → dark areas), then transforms them into a numerical number relating to orange peel. A third method, alluded to within the standard, mentions a combination of surface amplitude and spatial content but clearly defines neither within. To circumvent the need for subjective visual analysis, a previously published systematic framework, SQuID^{TM} [21], will be built upon to derive quantitative descriptions using areal maps subdivided into multiple length scale measurements and a set of logical rules to objectively classify the aforementioned set of subjectively graded visual smoothness panels. Section 4.2.1 includes an overview of the standardized visual smoothness tiles investigated in this study. Section 4.2.2 includes acquisition and data collection details of the tiles using a CSI and a gloss meter. Gloss meter readings are included in the analysis due to their simplicity. Areal parameters considered and spatial bandwidths used in generating the datasets for subsequent analysis are covered in Section 4.2.3. The parameter selection and classification methodology employed for down-selecting metrics and classifying the standard appearance tiles is covered in Section 4.3. The performance of a decision tree classifier and a summary of the down-selected parameters using this algorithmic methodology is covered in Section 4.4. A summary of the results and areas for future work is outlined in Section 4.5.

4.2 Samples and measurements

4.2.1 Visual smoothness standards

The PCI visual smoothness standards shown in Figure 1 cost approximately 650 USD and consist of ten 100 mm × 150 mm black powder-coated tiles with varying degrees of smoothness. The samples are numbered such that tile 01 appears to be the "roughest" sample, containing the most orange peel, and tile 10 is the "smoothest", visually exhibiting the least orange peel and the highest reflectance levels. The orange peel texture appears as its name suggests: visible local surface undulations or waviness. It is most visible in tiles 01 and 02 and within the region of the reflected overhead light in tiles 04 and 05 (see the indicated regions in Figure 4.1). Industrial equipment manufacturers pre-scribe different quality designations to surfaces with an orange peel rating of 05-09 versus those with a rating of 02-09 [37]. Typically, classification is carried out via HVI in illuminated cells by comparing the standard tile grades to the part surfaces. Under non-ideal lighting conditions, only PCI tiles 01, 02, and potentially tile 03 can be reliably differentiated. Resolving the differences between tiles 04 to 10 is more challenging, requiring specific lighting conditions, viewing angle, and visual acuity to resolve minor variations in spatial periods and specular reflection. To correctly classify all ten tiles over long periods of time would be considered a difficult inspection task and would undoubtedly lead to inconsistency between inspectors.

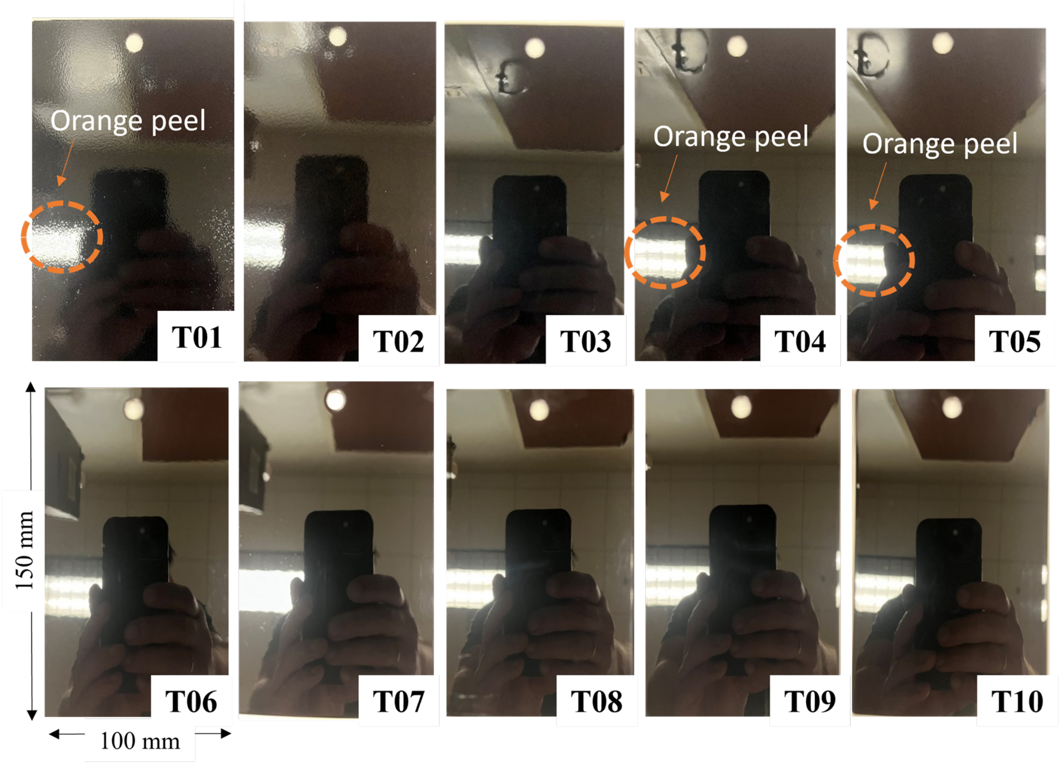


Figure 4.1: Images of PCI visual smoothness standards captured by an iPhone under non-ideal lighting conditions. Circular regions shown in T01, T04, and T05 illustrate the reflected image distortion caused by the severity of the orange peel.

The PCI tiles in Figure 4.1 were kept in a protective case during shipping and storage. Direct handling was limited to the edges of each tile to avoid damage or contamination of the measurement area. Before each measurement described in the next section, the tiles were gently wiped with a clean microfiber cloth to remove any dust or contaminates on the surface.

4.2.2 Gloss meter and CSI measurements

The different measurement strategies for the two different measurement approaches are depicted in Figure 4.2. Thirty measurements of each PCI tile were collected using a RhoPont-IQ meter [29]. This was achieved by rotating about the z-axis normal to the surface and translating the device across the central region of the 100 mm × 150 mm tiles. The assessment areas were uniformly distributed over approximately 50 mm × 50 mm center region of the tile, maintaining a buffer around the edges to prevent inaccurate readings. The PCI tiles would generally be considered "high-gloss" surfaces; as such, 20° angle gloss measurements are most appropriate for quantifying the tiles compared to 85° and 60° angles. By design, DOI, Rspec, and RIQ assessments are expected to correlate with the severity of orange peel, which is assumed to determine the ascending ordering of the tiles, i.e., from tile 01 to 10. Surface measurements of the PCI visual smoothness standards were obtained using a Zygo NexView CSI (Middlefield, CT, USA) configured with a Michelson 2.75× objective and 0.5× tube lens, providing a numerical aperture (NA) of 0.08 and sampling interval of 5.91 μm. The objective chart for the instrument uses the term spatial sampling in units of μm per pixel to describe the latter, noting it is the pixel size on the sample and is derived from the camera pixel size divided by the system magnification. The instrument is housed in a laboratory environment temperature controlled to 20 ± 0.1 degrees C. Plastic toe clamps were used to secure and flatten each tile to the stage of the CSI. The CSI measurements were taken from the center region of the tile. The measurements comprise partially overlapping 6.05 mm × 6.05 mm scans stitched together to increase the nominal field of view to 30.25 mm × 30.25 mm. Stitching enables a larger field of view without comprising the measurementâs spatial resolution. The stitching process was carried out using Zygo MX^{T}M (v8.0.0) software using 20% overlap between sequential scans in the adaptive adjust mode.

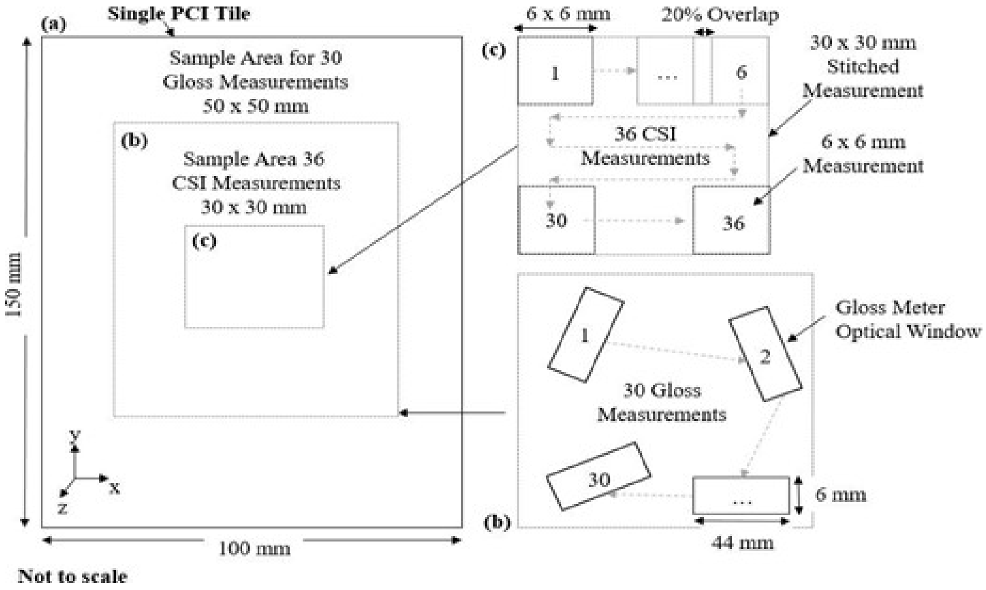


Figure 4.2: (a) Illustration of sampling locations on the PCI tiles for gloss meter (b) and CSI measurements (c).

4.2.2.1 Measured gloss metric values

Figure 4.3 depicts box and whisker plots highlighting the relationship between the values of selected gloss metrics (y-axis) and PCI tile number (x-axis). Gray boxes represent the interquartile (IQR) range, with whiskers denoting the minimum and maximum quartiles. Each box displays the median and mean of the data. The median is indicated by a horizontal line, while the mean is represented by "×" symbol(s). Outliers are shown as circles.

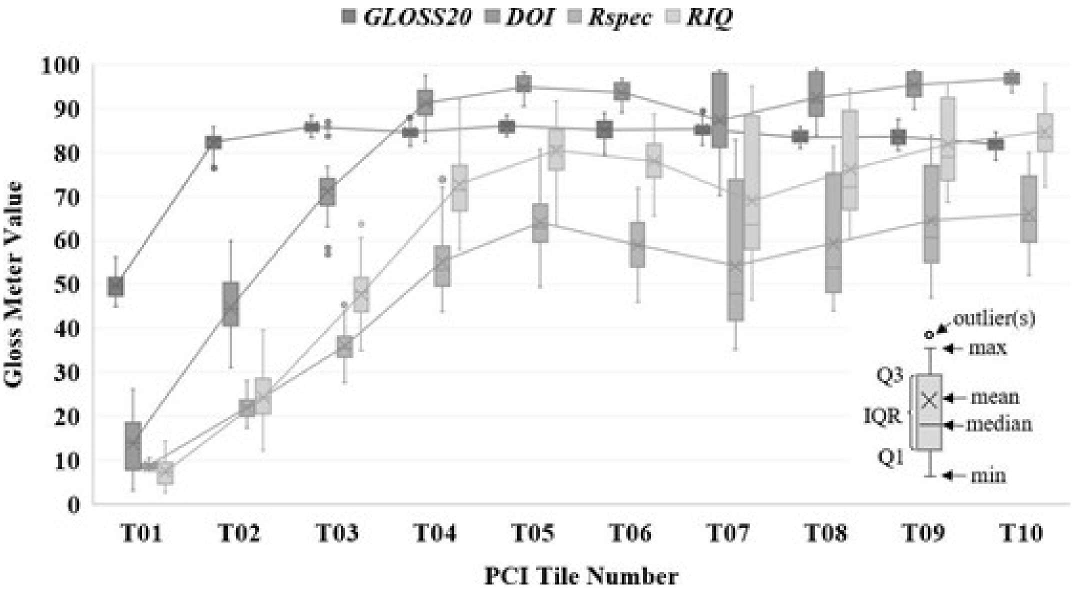


Figure 4.3: Box and whisker plots of select Rho-Point IQ gloss meter readings vs. PCI tile number.

For Gloss20, there is a clear distinction between tile 01 and tile 02, with average values of approximately 50 and 83 gloss units, respectively. Metrics DOI, Rspec, and RIQ demonstrate a positive correlation for tile 01 through tile 04 (the mean value of the measurements increases as a function of the tile number.). However, even using the data illustrated in Figure 4.3, it is clear that no one of the functional parameters (Gloss20, DOI, Rspec, RIQ, etc.) would enable the ability to distinguish between all ten tiles, nor facilitate the ability of a machine learning model to automate the classification of new tiles. This is because there is too much overlap between metric values computed on tile 04 through tile 10. For example, the range of Gloss20 values for T04, T05, T06, and T07 are almost identical.

4.2.2.2 Measured CSI data and the power spectral density curves

The images at the top of Figure 4a show stitched 30.25 mm × 30.25 mm CSI measurements, with typical 6 mm × 6 mm single CSI measurements taken from each tile shown below (Figure 4b). The lower section, Figure 4.4c, depicts the Power Spectral Density (PSD) for each of the ten stitched measurements (30.25 mm × 30.25 mm).

The PSD of a surface is the average of the square of the Fourier transform for each row (or column) of an areal measurement [38]. The graph provides insights into the spatial content of a surface, whereby the x-axis covers the range of spatial wavelengths possible to capture within a measurement. In Figure 4.4, a non-directional PSD was calculated using Zygo MX^{TM} (v8.0.0) for each 30.25 mm × 30.25 mm stitched CSI measurement (see Figure 2c) using default settings. The non-directional PSD is calculated by summing the 2D PSD data for a single frequency in all directions for each frequency for which there is data. Readers are referred to the MX^{TM} (v8.0.0) software manual for complete details. In this case, the highest frequency coincides with the sampling interval of 5.9 μm, limited by the pixel size of the CSI detector and system magnification. The lowest frequency is bounded by the 30.25 mm dimension(s) of the stitched measurement. However, the longest feasible wavelength discernible with any degree of confidence is approximately 7 mm, i.e., one-quarter of the overall length of the stitched measurement. This is slightly larger than the nominal field-of-view of the single 6 × 6 mm CSI measurements (see Figure 4.2c). Different surface features (micro-scale surface roughness affecting reflectance or longer-scale surface undulations referred to as orange peel) occupy different length scales on the surface; both influence the classification of the ten different tiles. However, exactly which surface features and length scales correlate with the visual ranking of the tiles is unknown. Therefore, the surface is split into five different spatial bands for further analysis. The bands (cut-off-frequencies) chosen are those commonly used to access paint finish in the automotive industry [39-41]: W_A (0.1 mm $\rightarrow$ 0.3 mm), W_B (0.3 mm $\rightarrow$ 1 mm), W_C (1 mm $\rightarrow$ 3 mm), W_D (3 mm $\rightarrow$ 10 mm), and W_E (10 mm $\rightarrow$ 30 mm). The additional band W_R (0.018 mm $\rightarrow$ 0.1 mm) is included to account for the shortest spatial content on the surface, i.e., between three times the sampling interval and the lowest 0.1 mm cutoff of the W_A band. The additional P bands indicated the unprocessed surface data of either the larger stitched file $P_{Stitched}$ or a single measurement from the stitch, P_{single}. Integrating the PSD curve(s) shown in Figure 4.4 across the entire range of spatial frequencies would result in the total power of the primary surface(s) comparable to the squared root-mean-square roughness (Sq) of the stitched measurement(s). Band-limited calculations of Sq can be obtained using the stated cut-offs of a given band as the bounds for integration. The W_D band shows a more "ideal" (strictly negative) correlation between amplitude parameters and tile number (relative to other bands); T01 has the highest band-limited Sq value (0.37 μm), and Sq generally decreases with every other subsequent tile number. However, the Sq value for T10 (0.124 μm) is approximately equal to that for T09 (0.123 μm), the Sq value of T04 (0.276 μm) is greater than T03 (0.261 μm), similar for T06 (0.243 μm) and T05 (0.230 μm). W_D does not provide the required amount of separation that is needed for reliable classification. Using cut-offs for the W_A or W_R band would result in the Sq of T03 greater than T02. Bands W_B and W_C would rank T09 and T10 above T04, T05, and T06. Band W_E does not rank the sample correctly, either. For subsequent analysis, only W_A, W_B, and W_C bands are used on the 360 individual smaller measurements (6.05 mm × 6.05 mm). The additional band W_R that captures shorter wavelengths (0.1 mm → 0.018 mm) and a P band (the primary surface) will also be used. Limiting the evaluation to these bandwidths eliminates the need to obtain stitched measurements to accommodate the field of view (FOV) required for bands W_E, W_D, which in turn reduces overhead in data storage (24 MB vs. 262 MB), measurement time (4 s vs. 10 min), and subsequent processing time (1 s vs. 10 s).

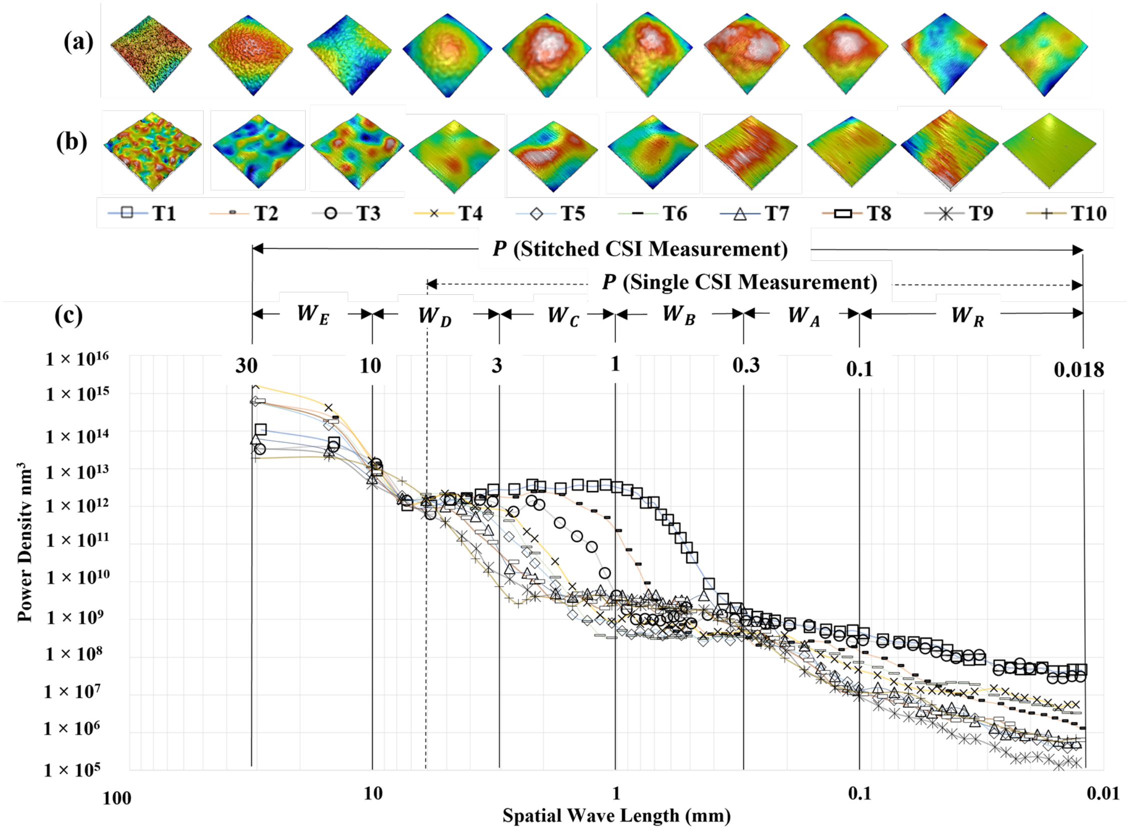


Figure 4.4: (a) Examples of stitched 30.25 mm × 30.25 mm CSI measurements. (b) Examples of single 6.05 mm × 6.05 mm CSI measurements. (c) Power spectrum density curves of stitched 30.25 × 30.25 mm CSI measurements of PCI tiles 01-10 with spatial bands W_E, W_D, W_C, W_B, W_A, and W_R overlaid.

4.2.3 Dataset generation

Table 4.1 summarizes the acquisition details, measurements, and parameters used in the subsequent analysis. The field of view (FOV) for Rhopoint-IQ measurements varies depending on the calculated metric. For the 85° gloss metric, it corresponds to 4.4 mm × 44 mm; for the 60° gloss, a 6 mm × 12 mm area is required; and for the 20° gloss, DOI, RIQ, and Rspec 6 mm × 6.4 mm definition areas are needed. The gloss dataset consists of only these eight parameters provided by the Rhopoint-IQ. Twenty-one ISO 25178-2 parameters were computed for each CSI measurement after the pre-processing operations listed below in Table 4.2 were applied. The specified filtering operations and subsequent parameter calculations were performed using Zygo

Table 4.1: Instrument specifications and datasets collected for analysis.

Instrument	**Zygo NexView**	**Rho-Point IQ**
Objective	2.75×	n/a
Tube Lens	0.5×	n/a
Field of View (FOV)	6.05 mm × 6.05 mm	6.04 mm × 6 mm 6 mm × 12 mm 4 mm × 44 mm
Sampling Interval	5.9 µm	n/a
Sampling Area	30.25 mm × 30.25 mm	50.0 mm × 50.0 mm
Number of Measurements	36/Sample	30/Sample
Metrics	*Sa*, *Sq*, *Ssk*, *Sku*, *Sp*, *Sv*, *Sz*, *Sdq*, *Sdr*, *Sal*, *Str*, *Std*, *Sk*, *Spk*, *Smr*1, *Smr*2, *Svk*, *Vmc*, *Vmp*, *Vvc*, *Vvv*, *Sxp*	*Gloss*20, *Gloss*60, *Gloss*85, *Haze*, *LogHaze*, *DOI*, *RIQ*, R_{spec}
References	[32,42]	[27,29]

Note: n/a refers to either not available or not applicable.

MX^{TM} (v8.0.0) software [42]. Each unique dataset, P, W_C, W_B, W_A, and W_R, consists of 360 individual measurement sites (6.05 mm × 6.05 mm), 36 per sample, and 22 ISO parameters calculated for each of the 360 areal maps. The tabulated results were exported in .csv format for subsequent analysis. For brevity, only the shorthand abbreviations of the ISO 25178-2 areal parameters are included in Table 4.1; readers are referred to the standard itself [32] or to references [19,43] for detailed verbal descriptions and numerical definitions of each parameter.

Table 4.2 includes the processing information for each dataset, specifying the lowpass filter, bandpass filter, form removal, spike removal, and edge clipping applied to data prior to computing ISO 25178-2 parameters. The cut-off wavelength of the initial lowpass Gaussian spline filter used in each case is set to three times the sampling interval to remove high-frequency noise from the data. Note that the sampling interval remains constant, irrespective of the filtering applied to the data. A few false height readings were observed in the measurements, likely caused by dust particles or contamination on the surface. Although these will have little effect on the values

Table 4.2: Detailed information on datasets and filtering operations: low-pass, bandpass, form re-moval, spike removal, and edge clipping.

Dataset	Low-Pass	Bandpass	Form Removal	Spike Removal	Edge Clipping
P	18.0 μm	-	Best fit plane	6σ from mean plane	No
W_C	18.0 μm	1.0 mm–3.0 mm	Best fit plane	6σ from mean plane	24 pixels
W_B	18.0 μm	0.3 mm–1.0 mm	Best fit plane	6σ from mean plane	24 pixels
W_A	18.0 μm	0.1 mm–0.3 mm	Best fit plane	6σ from mean plane	24 pixels
W_R	18.0 μm	0.018 mm–0.1 mm	Best fit plane	6σ from mean plane	24 pixels

of texture parameters such as the average roughness (Sa) or root mean square gradient (Sdq), as the million other pixel values will average them out, a single outlier in the measurement can significantly affect the extreme value parameters such as the maximum peak height (Sp), maximum pit height (Sv), and maximum height (Sz); therefore, spike clipping is used to ensure these parameters characterize the actual topography and not single-pixel measurement artifacts. The spike filter is set at 6Ï; the maximum number of data points removed from any measurement was less than 0.3% and visually has no impact on the topography image or values of other ISO parameters. Edge cropping was employed to remove filtering artifacts. Examples of processed height maps after bandpass filtering are shown in Figure 4.5.

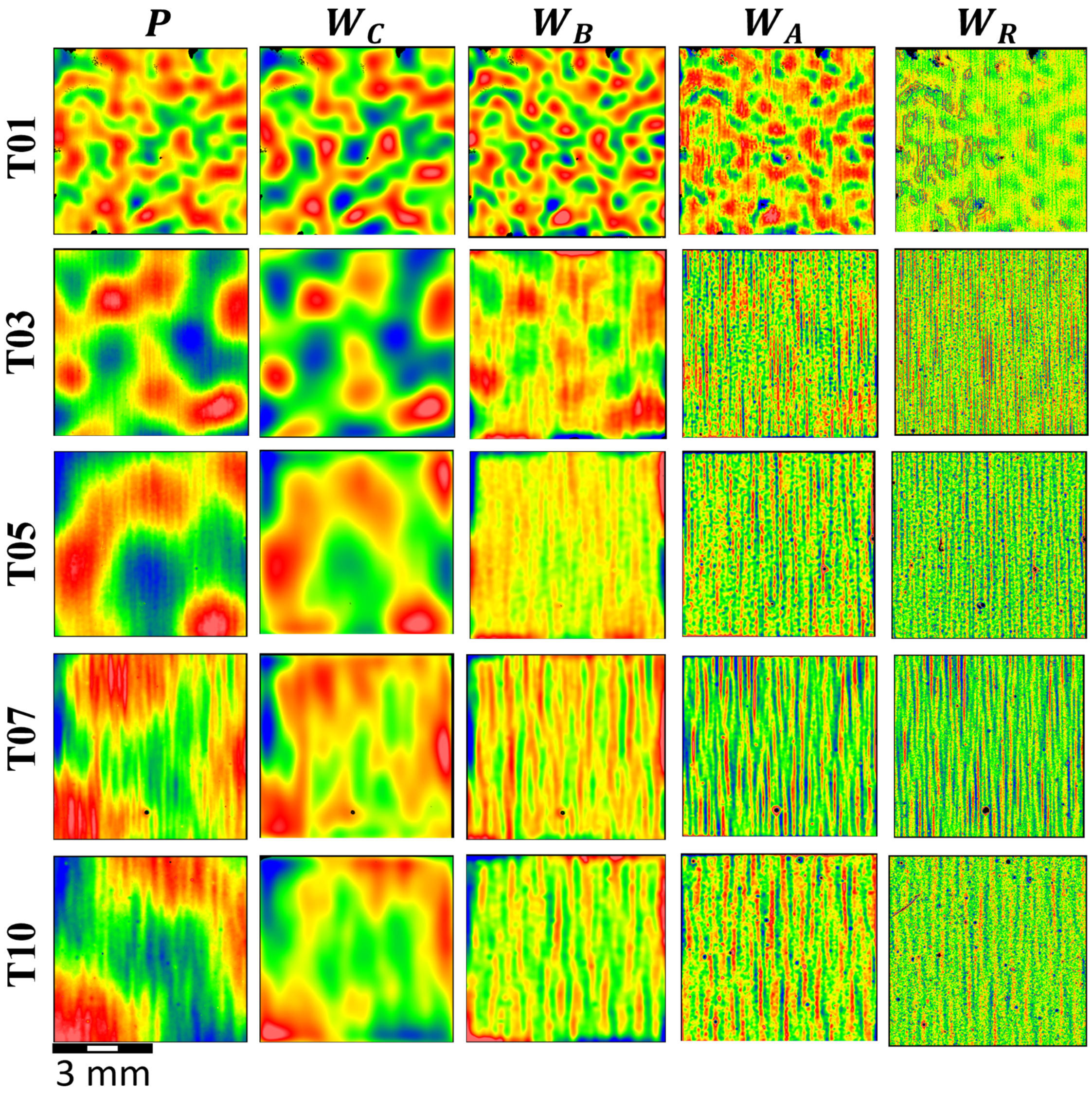


Figure 4.5: Examples of CSI measurements after pre-processing operations used to create datasets P, W_C, W_B, W_A, and W_R. Color ranges have been optimized for each image to improve viewing quality. Red and blue colors correspond to areas of higher or lower elevation, respectfully.

Optimized color representation for each height map shown in Figure 4.5 makes it much easier to see differences in the spatial aspects of the topography than if a fixed color bar (height scale) was used for the image or any row/column; as such, the colors depict different height ranges for each image. The primary surface (P) of tile 01 exhibits an isotropic topography (no apparent surface structure or directionality). However, traveling down the same column from T01 to T10, notice that as the degree of isotropic orange peel (longer undulations) decreases, there is an increase in anisotropic vertical striations. Still, identifying unique surface characteristics that distinguish between tiles 05, 07, and 10 is extremely difficult, regardless of the applied filtering. Figure 4.6 depicts the areal roughness (Sa) and the root-mean-square-gradient (Sdq) of the tile surfaces at the different spatial bands. Notice the lack of a strictly negative correlation between parameter values and the tile number. For the primary surface (the P dataset shown in Figure 4.5), the areal surface roughness Sa of the 10 tiles ranged from 0.25 μm to 2 μm, Sq from 0.15 μm to 2.02 μm, and Sdq ranged from 0.07° to 0.77°. A fourth-order fit between Sdq and tile number yields correlation coefficient R2 values of 0.98, 0.99, 0.96, and 0.96 for P, W_C, W_B, and W_A datasets, respectively. Similar R^2 values were observed for Sa. Although there is a high correlation, the fitted function will not enable explicit classification criteria; there is insufficient separation in the metricsâ values for the higher-numbered tiles. While not shown here, none of the other 20 ISO 25178-2 parameters considered delivered clear distinctions and/or strictly linear trends between all ten tiles.

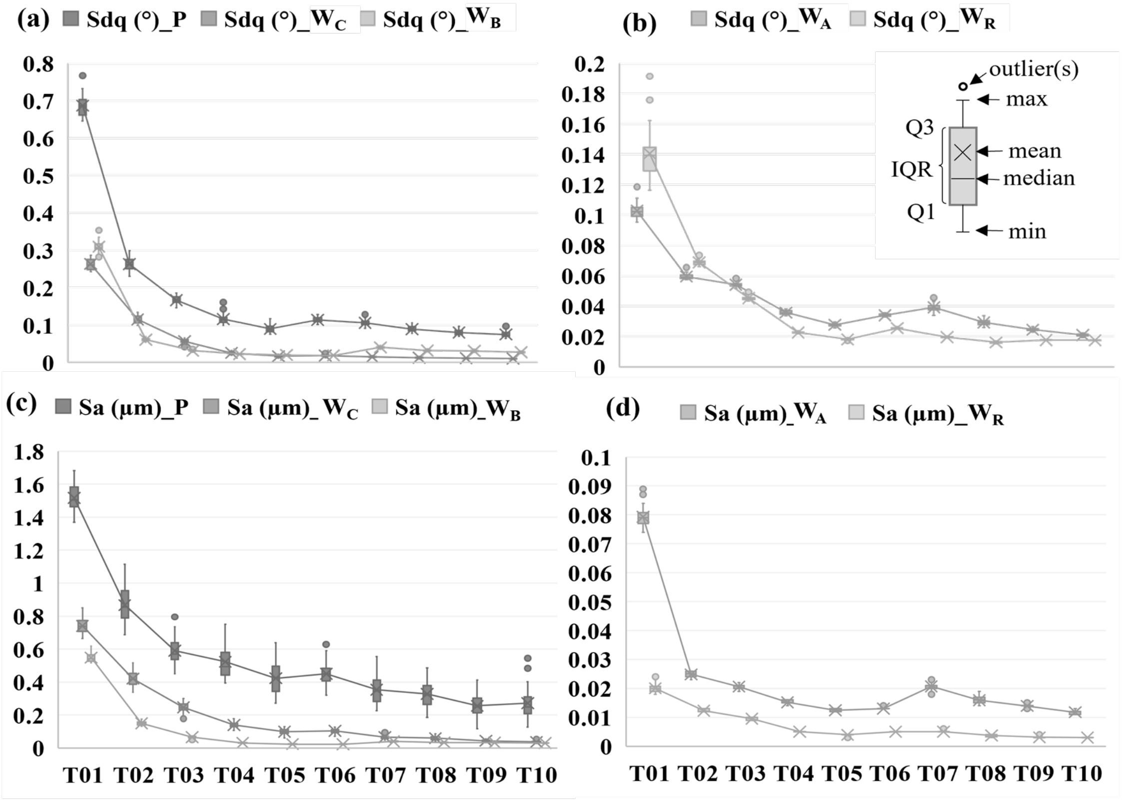


Figure 4.6: (a) Sdq (P, W_C, W_B) vs. tile number. (b) Sdq (W_A, W_R) vs. tile number. (c) Sa (P, W_C, W_B) vs. tile number. (d) Sa (W_A, W_R) vs. tile number.

4.3 Parameter selection and classification methodology

Surface Quality and Inspection Descriptors (SQuID^{TM}) Instead of relying on preconceived notions of parameter selection for accurately classifying the PCI tiles, an algorithmic method outlined in [21] called SQuID^{TM} (Surface Quality and Inspection Descriptors) is used to down-select optimal surface parameters for classification. SQuID^{TM} uses a selection method that breaks down a user-defined multi-class classification task into a series of binary classification tasks. For instance, in this case, the multi-class task involves using areal texture or gloss descriptions from Nexview CSI or Rhopoint-IQ to determine a measured tile designation (01, 02, 03 ... 10). The procedure begins by decomposing the multiclass task into a finite series of binary classification tasks (T01 vs. T02, T01 vs. T03, and so on). For each binary task, a reference set of measurements quantified by a single areal or gloss description is sampled from the dataset. The difference in means between the two groups is divided by the root-mean-square of the standard deviations. Taking the absolute value of this resulting quantity yields a discriminability index referred to as d', describing the separation between the two groups in terms of normalized standard deviations. The process is repeated for each parameter on each binary classification task derived from the multi-class dataset, resulting in a d' matrix. In this matrix, parameters are indexed as rows, and binary classification tasks are represented as columns. The optimal set of parameters for the multi-class task is derived by selecting the parameter (row index) with the maximum d' value for each task (column index). If, for every binary task, the corresponding selected parameter exhibits a d' value of approximately seven or higher, a reasonably optimized decision tree classifier trained using these selected parameters is expected to achieve near-perfect classification accuracy on the multi-class dataset. Following the down-selection process, SQuID^{TM} employs a decision tree as the default machine-learning model to determine a logical mapping between the selected parameters and the target surface classes. The fitted decision tree can intuitively communicate how the selected surface texture parameters and associated thresholds can be employed as rules for classifying a newly measured surface, providing an objective basis for assessing surface quality in accordance with user-defined grades of surface quality. Readers are referred to the original paper [21] for complete details, including examples of the selection procedure, construction of the d' matrix, and graphs related to the expected classification outcomes versus d' values of the selected parameters.

4.4 Results

4.4.1 Summary of d' matrix

Figure 4.7 represents the combined d' matrix generated for the P, W_C, W_B, W_A, and W_R dataset(s). Columns are indexed as tasks following the sequence T01 vs. T02, T01 vs. T03, etc., and rows are indexed as ISO parameters for each band (see Table

4.1 for ordering). For this discussion, reading the fine print associated with the rows and columns on the matrix is not important; what is important is the distribution of the gray values on the d' matrix. The shading represents the magnitude of the d' values computed for each row (parameter)-column (task) combination. Ideally, one row (metric) would have a d' value greater than seven (black shading) across all columns (tasks), though this is not the case for any datasets considered. Notice that the d' values for most ISO parameters along the first column (task T01 vs. T02) tend to be greater than those of the last column (task T09 vs. T10). This implies that differentiating between the topography of tiles 01 and 02 using ISO parameter values is relatively trivial compared to 09 and 10, regardless of the pre-processing routine applied to the data. The d' matrix also suggests that a classification model that uses metrics computed for the P and W_C bands will perform worse than a model that uses ISO metrics computed for the W_B and W_A bands. This is because there is an increased coverage of higher d' values across the entire task space in the W_A and W_B bands. A combination of metrics with dark values ($d' \geq 7$) across all tasks implies the data are likely separable, and near-perfect classification performance is expected. The expanse of white cells ($d' < 2$) in the P and W_C bands implies less-than-ideal performance ($\approx 80\%$ classification accuracy or less); see the experimental curve in [21] for the empirical relationship between d' and classification accuracy. The relevant regions are highlighted by the dashed red and green boxes in Figure 4.7.

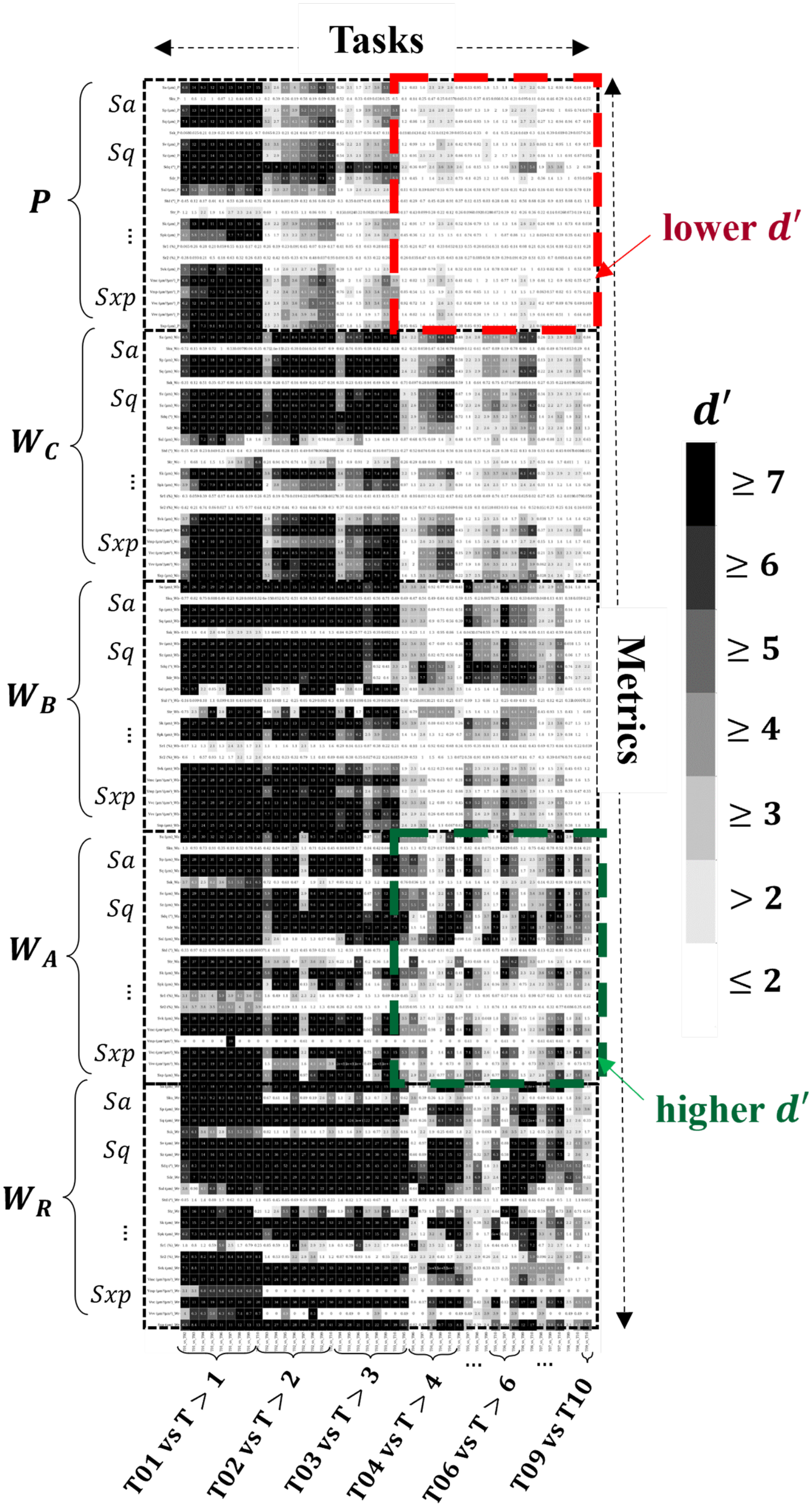
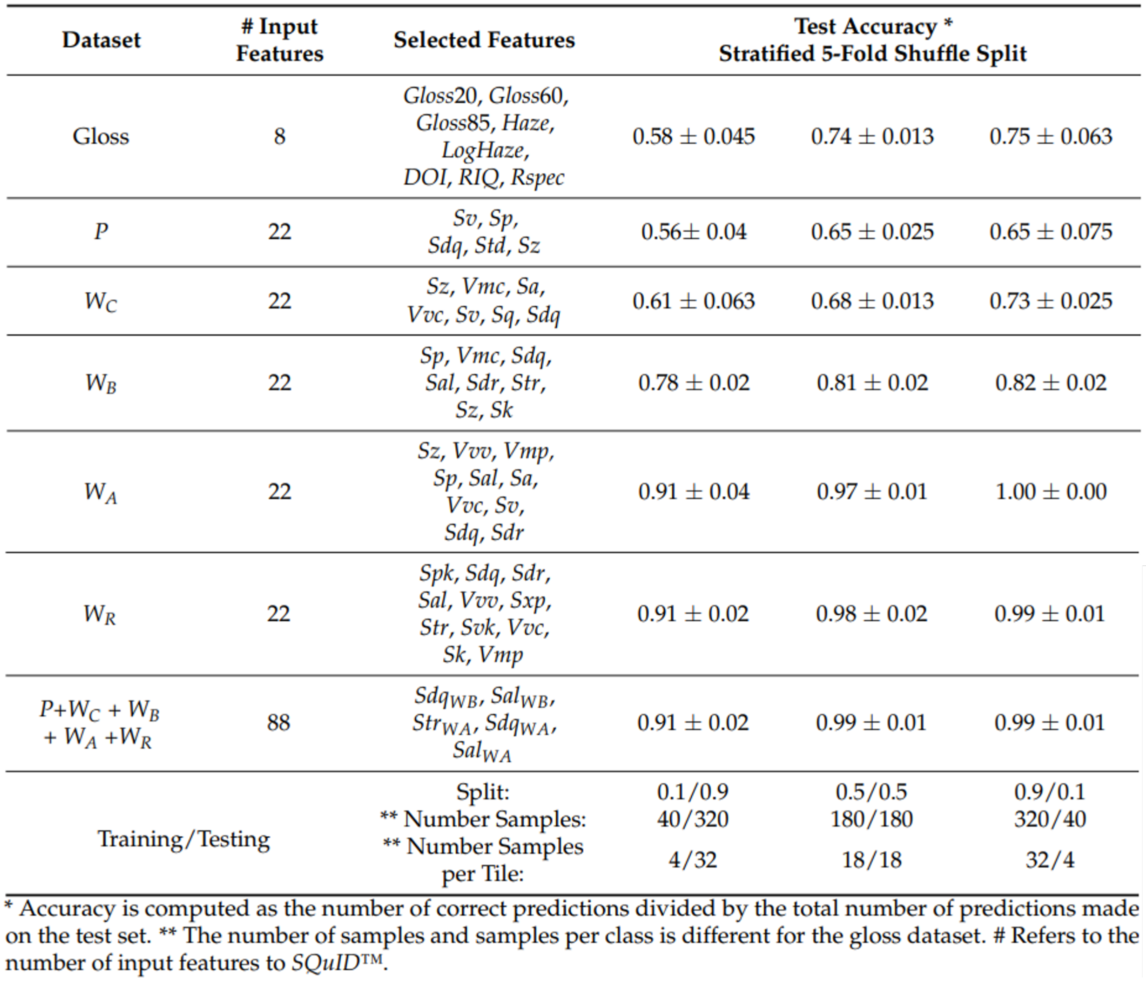


Figure 4.7: Example of the d' matrix generated for combined datasets. P, W_C, W_B, W_A, and W_R comprise 110 metrics (rows) and 45 tasks (columns).

4.4.2 Identified parameters and classification performance

Table 4.3 indicates the top metrics identified by SQuID^{TM} for each analysis set and the classification performance of a fitted decision tree that used the selected metrics to predict the tile number (i.e., surface grade). The average and standard deviation of the test accuracy based on stratified five-fold shuffle split cross-validation [44] using train/test splits of 0.1/0.9, 0.5/0.5, and 0.9/0.1 for each dataset is shown in the rightmost columns. The evaluation provides a more comprehensive estimate of the expected classification performance compared to using a single dataset for training and testing. The first row of Table 4.3 indicates that gloss metrics fail to provide enough information to discern between the ten standard surface tiles. This comes as no surprise given the overlap of the gloss data illustrated in Figure 4.3; no reduction in the number of selected features is observed, and average accuracies as low as 58% are recorded for the 0.1/0.9 split, 74% for the 0.5/0.5 split, and 75% for the 0.9/0.1 split. For the primary dataset P, a total of five features are selected by SQuID^{TM}. Still, they fail to perform any better than the gloss metrics, averaging around 65% classification accuracy or less across all train/test splits. This is unsurprising given that only 2% of metrics had values of $d' > 3$ for tasks T04 vs. T > 04 (see the dashed red box in the top right corner of Figure 4.7). Coincidentally, this is where most incorrect classifications occurred. Classification accuracy marginally improves between 5% and 8% depending on the train/test split for the W_C dataset, where seven features are selected. Again, better performance is expected; the d' values in Figure 4.7 for this dataset tend to be greater than those of the ISO metrics of the P dataset. For the latter, only 31% of all entries had d' values greater than three, with the former having 50%. For tasks T04 vs. T05 to T10, 29% of entries were greater than three. Most incorrect classifications occur between T08, T09, and T10. Seven metrics (with higher d' values) were also selected for the W_B dataset; the classification performance ranges between 78% to 82%, approximately a 10% increase over the W_C dataset. However, again, most incorrect classifications occur for tasks T08 vs. T09 and T10, where none of the d' values of the selected metrics were greater than four. A drastic increase in performance is observed for the W_A and W_R datasets. The former achieved an average accuracy of 97% using ten selected features, and the latter reached 98% based on eleven selected parameters for the 0.5/0.5 split. Both consist of selected metrics with $d' > 4$ for tasks T08 vs. T09 and T10. Less than five incorrect classifications in total were recorded in both cases. For the combined dataset (i.e., P, W_C, W_B, W_A), which includes 88 ISO features, near-perfect classification is achieved for the 0.5/0.5 and 0.9/0.1 splits using only five multi-scale metrics selected from the W_B and W_A datasets. Less than five incorrect predictions were made, each being off by a single tile designation and occurring only on tasks T07 vs. T08 to T10. The fact no ISO metrics were selected from the P and W_C datasets indicates there was always an ISO metric computed on either the W_B or W_A dataset that possessed a greater d' value for a given classification task (column(s) on the d' matrix).

Table 4.3: Classification accuracy of decision tree based on selected features for seven different datasets.



Dataset	**# Input Features**	**Selected Features**	**Test Accuracy * Stratified 5-Fold Shuffle Split**		
Gloss	8	*Gloss20, Gloss60, Gloss85, Haze, LogHaze, DOI, RIQ, Rspec*	0.58 ± 0.045	0.74 ± 0.013	0.75 ± 0.063
P	22	*Sv, Sp, Sdq, Std, Sz*	0.56± 0.04	0.65 ± 0.025	0.65 ± 0.075
W_C	22	*Sz, Vmc, Sa, Vvc, Sv, Sq, Sdq*	0.61 ± 0.063	0.68 ± 0.013	0.73 ± 0.025
W_B	22	*Sp, Vmc, Sdq, Sal, Sdr, Str, Sz, Sk*	0.78 ± 0.02	0.81 ± 0.02	0.82 ± 0.02
W_A	22	*Sz, Vvv, Vmp, Sp, Sal, Sa, Vvc, Sv, Sdq, Sdr*	0.91 ± 0.04	0.97 ± 0.01	1.00 ± 0.00
W_R	22	*Spk, Sdq, Sdr, Sal, Vvv, Sxp, Str, Svk, Vvc, Sk, Vmp*	0.91 ± 0.02	0.98 ± 0.02	0.99 ± 0.01
P+W_C + W_B + W_A +W_R	88	Sdq_{WB}, Sal_{WB}, Str_{WA}, Sdq_{WA}, Sal_{WA}	0.91 ± 0.02	0.99 ± 0.01	0.99 ± 0.01
Training/Testing		Split: ** Number Samples: ** Number Samples per Tile:	0.1/0.9 40/320 4/32	0.5/0.5 180/180 18/18	0.9/0.1 320/40 32/4

* Accuracy is computed as the number of correct predictions divided by the total number of predictions made on the test set. ** The number of samples and samples per class is different for the gloss dataset. # Refers to the number of input features to *SQuID*™.

4.4.3 Interpretation of classification criteria

Figure 4.8 depicts the decision tree generated using down-selected parameters, including the auto-correlation length Sal and root-mean-square gradient Sdq computed for the W_B and W_A bands, in addition to the texture aspect ratio Str computed on the W_A band. Figure 9 plots the values of the ISO metrics selected for use in the decision tree for all ten tiles. Examination of the three graphs illustrates that a combination of metrics can exist to differentiate the tiles from each other but that no one metric could differentiate between all ten; the decision tree is required to achieve the latter. Verification of the decision tree logic used for classifying the data is depicted in

Figure 4.9. Following the logic of the tree (Figure 8), the first split for differentiating tiles 01-06 from 07-10 is by comparing autocorrelation length Sal on the W_B surface to a threshold of 133 μm (the root node top of the tree). Based on inspection of the data in Figure 9a, it is obvious that if the SalW_B value of the surface in question is less than or equal to this value, a classification of grades 07-10 is most appropriate. The next split to the left of the root node (decision node a) compares the Sal of the surface to a value of 64.4 μm but, in this case, on the W_A surface, differentiating a tile 07 or 08 from a 09 or 10 designation. Finally, a minor difference in the Sdq of the surface for spatial band W_A is used to assign a final designation of 07, 08, 09, or 10 (see decision nodes b and c). Following the logic to the right of the root node (decision node d), the Sdq of the W_A band is used to split the group of designations (04, 05, 06) from (01, 02, 03). In the case of the latter group, the presence of what appears to be rolling marks on the surface, exhibiting a dominant lay pattern, becomes more noticeable as the degree of orange peel is reduced, resulting in a departure from an isotropic texture (see Figure 4.5). This trend in increasing anisotropy is captured by the texture aspect ratio (Str) at the W_A band shown at the bottom of Figure 4.9c, which can be used to resolve the individual differences. For the former group (04, 05, 06), it is important to note that the Sdq for the W_B band is less for tile 05 than for 04 and 06. This initial split (decision node e), followed by comparing the Sdq value for the W_B band, enables individual designations (see Figure 4.9b and decision node g in Figure 4.8).

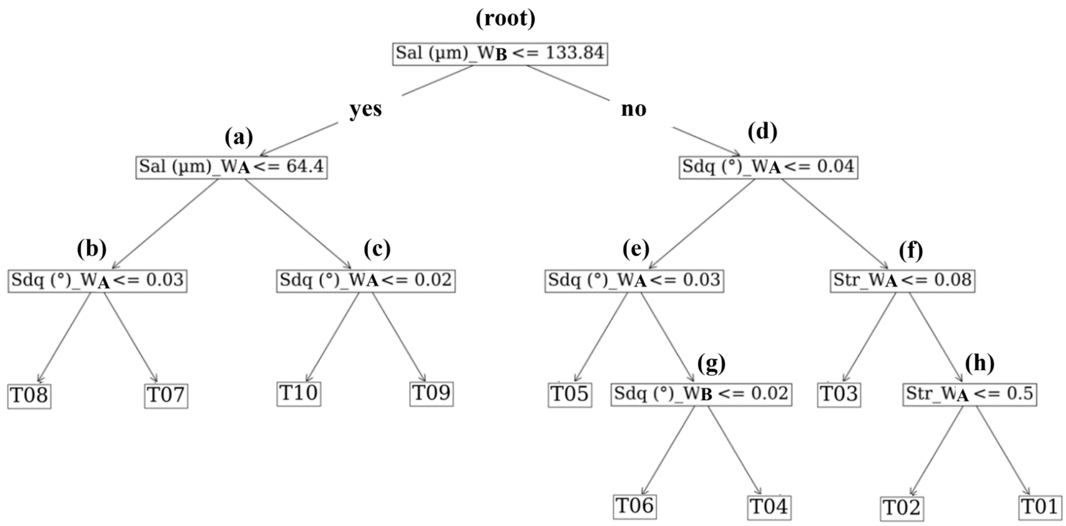


Figure 4.8: Decision tree generated from SQuID^{TM} based on five parameters selected from combined P + W_C + W_B + W_A datasets. For visualization, parameter values are truncated to two decimal places. Callouts (a-h) represent the logic used to classify a surface as one of the ten tile grades.

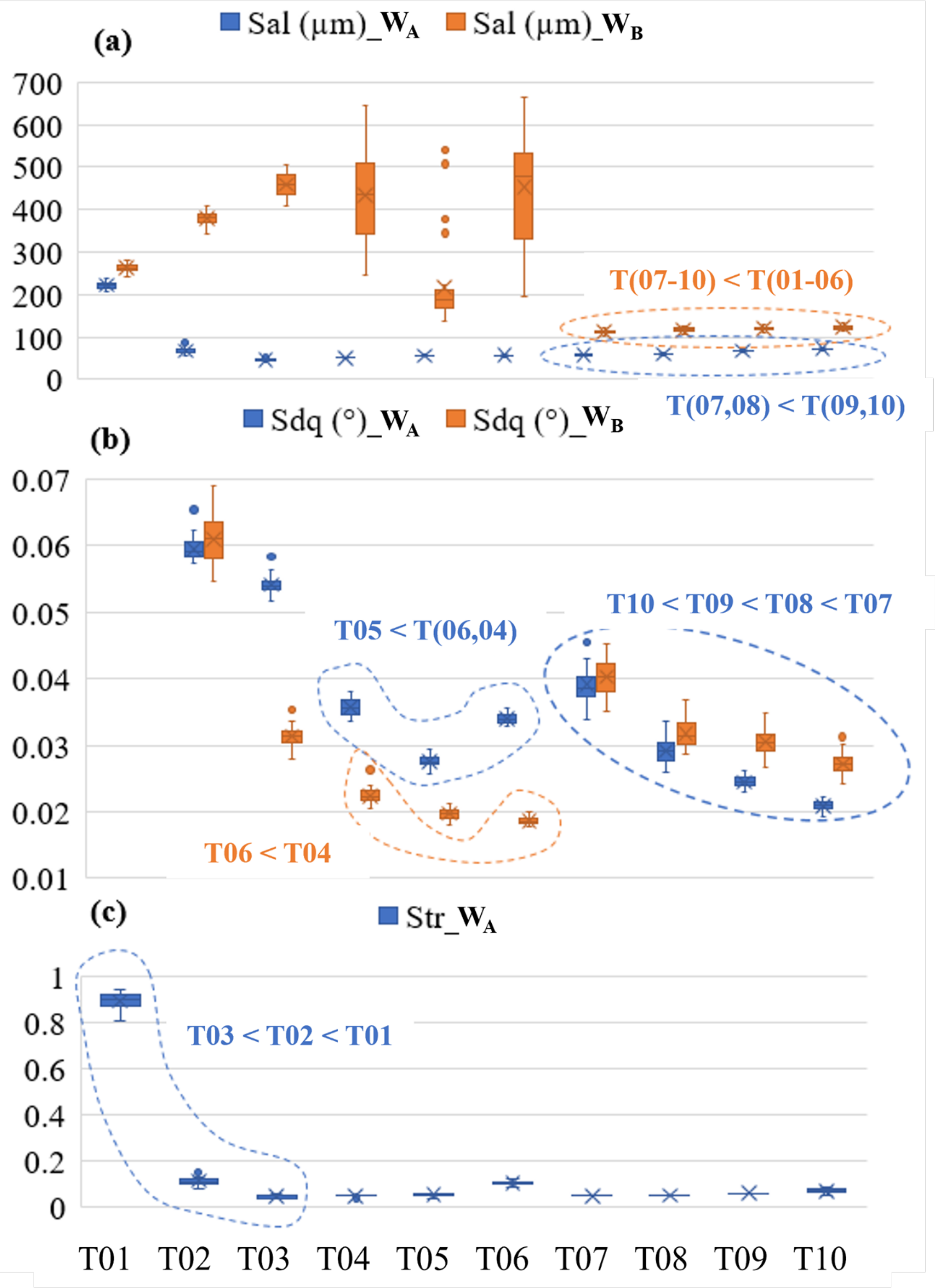


Figure 4.9: Decision tree verification by graphical assessment of the parameters selected by SQuID^{TM} for classification. (a) Measured auto-correlation length Sal and (b) rms-slope Sdq values for W_A and W_B bands plotted against tile designation T01-T10. (c) Texture aspect ratio Str values computed for the W_A band of the ten different tiles.

This result highlights a unique approach for characterizing topography with ISO 25178-2 parameters, which does not limit the evaluation to a single pre-processing routine. Using a script configured to run the CSI and automate the data processing/evaluation steps would enable the classification of new surfaces in less than 10 s, bypassing visual inspection. The parameter values in Figure 9 and decision tree logic in Figure 4.8 are subject to change for different measurement conditions and data processing. For example, the calculated values of hybrid parameters Sdq and the interfacial area ratio Sdr depend on the sampling interval. This can be modified by adjusting the system magnification of the CSI or increased digitally after the measurement is acquired. The use of different filter types, instrument settings, and environmental disturbances such as temperature and vibrations can also affect parameter values. Further study is required to evaluate the impact of these conditions on classification performance in addition to different modes of acquisition, such as confocal and profilometry.

4.5 Summary and conclusions

This work demonstrates the ability of a multi-scale methodology called SQuID^{TM} to down-select both ISO 25178-2 and gloss metrics to quantify ten different grades of powder-coated surfaces objectively. A set of only five multi-scale quantitative descriptions of the surface coupled with interpretable decision logic was provided to classify ten PCI visual smoothness standard tiles with near-perfect classification accuracy. This performance was achieved despite a limited sample size and nonlinear correlation between the qualitative ranking of the tiles (grades 01-10) and the values of parameters obtained from CSI measurements of each sample. Moreover, in cases where incorrect classification occurred, the predictions were only off by a single sample grade. Central to achieving this is leveraging the fact that a tileâs texture comprises features of varying spatial wavelength and amplitude; this is apparent from the varied PSD trends for each tile illustrated in Figure 4c and the bandwidth-filtered measurements depicted in Figures 4.5 and 4.6. While features of different wavelength variations may not always be re-solvable by a humanâs eye (if the amplitude is too low or the wavelength is too short), they all affect a lightâs interaction with a surface and, hence, how the same is perceived. Filtering at selected bandwidths enables their isolation, and subsequent quantification provides the additional metrics necessary for successful correct classification. For this work, filtering bandwidths used in the automotive industry for powder-coating and painting processes were utilized. No effort was expended on isolating optimal band-widths; however, pursuing this avenue could yield even better outcomes. Incorporating additional segmentation methods and feature parameters defined within the ISO 25178-2 standard may also yield even more robust and accurate classification. Of equal importance is to note that the presented classification criteria(s) provide the explicit logic to make predictions; it is not a black box function. It is possible to use this decision rule diagram in an automated way with a machine vision or measurement system and/or as a reference to support visual inspection effects when subjectivity arises. Overall, this approach demonstrates great potential in removing the need for visual inspection. In conclusion, the SQuID^{TM} framework is shown to be a practical, systematic, interpretable, multi-scale approach for selecting standardized statistical parameters to transform subjective grades of surface quality into objective descriptions of surface topography.

4.6 References

[1] Shi, Y.; Jiang, Z.; Cao, J.; Ehmann, K.F. Texturing of metallic surfaces for superhydrophobicity by water jet guided laser micromachining. *Appl. Surf. Sci.* 2020, 500, 144286.

[2] Nakae, H.; Inui, R.; Hirata, Y.; Saito, H. Effects of surface roughness on wettability. *Acta Mater.* 1998, 46, 2313-2318.

[3] Kubiak, K.J.; Wilson, M.C.; Mathia, T.G.; Carval, P. Wettability versus roughness of engineering surfaces. *Wear* 2011, 271, 523-528.

[4] Arumugam, K.; Smith, S.T.; Her, T.-H. Limitations caused by rough surfaces when used as the mirror in displacement measurement interferometry using a microchip laser source. In Proceedings of the American Society of Precision Engineering- 33rd Annual Meeting, Las Vegas, NV, USA, 4-9 November 2018. *Materials* 2024, 17, 1653.

[5] Vessot, K.; Messier, P.; Hyde, J.M.; Brown, C.A. Correlation between gloss reflectance and surface texture in photographic paper. *Scanning* 2015, 37, 204-217. [PubMed]

[6] Hilbert, L.R.; Bagge-Ravn, D.; Kold, J.; Gram, L. Influence of surface roughness of stainless steel on microbial adhesion and corrosion resistance. *Int. Biodeterior. Biodegrad.* 2003, 52, 175-185.

[7] Grzesik, W. Prediction of the functional performance of machined components based on surface topography: State of the art. *J. Mater. Eng. Perform.* 2016, 25, 4460-4468.

[8] Gockel, J.; Sheridan, L.; Koerper, B.; Whip, B. The influence of additive manufacturing processing parameters on surface roughness and fatigue life. *Int. J. Fatigue* 2019, 124, 380-388.

[9] Mandloi, K.; Allen, A.; Cherukuri, H.; Miller, J.; Duttrer, B.; Raquet, J. CFD and experimental investigation of AM surfaces with different build orientations. *Surf. Topogr. Metrol. Prop.* 2023, 11, 034001.

[10] Fox, J.C.; Evans, C.; Mandloi, K. Characterization of laser powder bed fusion surfaces for heat transfer applications. *CIRP Ann.* 2021, 70, 467-470.

[11] Mullany, B.; Savio, E.; Haitjema, H.; Leach, R. The implication and evaluation of geometrical imperfections on manufactured surfaces. *CIRP Ann.* 2022, 71, 717-739.

[12] Hekkert, P. Design aesthetics: Principles of pleasure in design. *Psychol. Sci.* 2006, 48, 157.

[13] Yanagisawa, H.; Takatsuji, K. Effects of visual expectation on perceived tactile perception: An evaluation method of surface texture with expectation effect. *Int. J. Des.* 2015, 9, 39-51.

[14] Tymms, C.; Zorin, D.; Gardner, E.P. Tactile perception of the roughness of 3D-printed textures. *J. Neurophysiol.* 2018, 119, 862-876. [PubMed]

[15] Stylidis, K.; Wickman, C.; Söderberg, R. Defining perceived quality in the automotive industry: An engineering approach. *Procedia CIRP* 2015, 36, 165-170.

[16] Simonot, L.; Elias, M. Color change due to surface state modification. *Color Res. Appl.* 2003, 28, 45-49.

[17] Dalal, E.N.; Natale-Hoffman, K.M. The effect of gloss on color. *Color Res. Appl.* 1999, 24, 369-376.

[18] Briones, V.; Aguilera, J.M.; Brown, C. Effect of surface topography on color and gloss of chocolate samples. *J. Food Eng.* 2006, 77, 776-783.

[19] Jiang, X.J.; Whitehouse, D.J. Technological shifts in surface metrology. *CIRP Ann.* 2012, 61, 815-836.

[20] Whitehouse, D.J. *Handbook of Surface and Nanometrology*; Taylor & Francis: New York, NY, USA, 2002.

[21] Redford, J.; Mullany, B. Construction of a multi-class discrimination matrix and systematic selection of areal texture parameters for quantitative surface and defect classification. *J. Manuf. Syst.* 2023, 71, 131-143.

[22] Brown, C.A.; Hansen, H.N.; Jiang, X.J.; Blateyron, F.; Berglund, J.; Senin, N.; Bartkowiak, T.; Dixon, B.; Le Goic, G.; Quinsat, Y.; et al. Multiscale analyses and characterizations of surface topographies. *CIRP Ann.* 2018, 67, 839-862.

[23] Das, J.; Linke, B. Evaluation and systematic selection of significant multi-scale surface roughness parameters (SRPs) as process monitoring index. *J. Mater. Process. Technol.* 2017, 244, 157-165.

[24] Hunter, R.S. Methods of determining gloss. *NBS Res. Pap. RP* 1937, 18, 958.

[25] ISO 2813:2014; Paints and Varnishes-Determination of Gloss Value at 20°, 60°, and 85°. International Organization for Standardization: Geneva, Switzerland, 2014. Available online: `https://www.iso.org/standard/56807.html` (accessed on 1 January 2024).

[26] ASTM D523; Standard Test Method for Specular Gloss. ASTM International: West Conshohocken, PA, USA, 2018. Available online: `https://www.astm.org/d0523-14r18.html` (accessed on 1 January 2024).

[27] ASTM E284-17; Standard Terminology of Appearance. ASTM International: West Conshohocken, PA, USA, 2017. Available online: `https://cdn.standards.iteh.ai/samples/97708/7e9a7cafb4bb432f9458a1eeff471dcd/ASTM-E284-17.pdf` (accessed on 1 January 2024).

[28] Weber, C.F.; Spiehl, D.; Dörsam, E. Comparing measurement principles of three gloss meters and using them for measuring gloss on metallic embellishments produced by the printing industry. *Lux Jr.* 2021, 15, 327-341.

[29] R. Instruments. Rhopoint iq Goniophotometer Datasheet. 2023. Available online: `https://www.rhopointamericas.com/download/rhopoint-iq-goniophotometer-datasheet-english/` (accessed on 1 January 2024).

[30] Leach, R. (Ed.) *Optical Measurement of Surface Topography*; Springer: Berlin/ Heidelberg, Germany, 2011.

[31] Conroy, M.; Armstrong, J. A comparison of surface metrology techniques. *J. Phys. Conf. Ser.* 2005, 13, 458.

[32] ISO 25178-2; Geometrical Product Specifications (GPS)-Surface Texture: Areal-Part 2: Terms, Definitions and Surface Texture Parameters. ISO: Geneva, Switzerland, 2021. Available online: `https://www.iso.org/obp/ui/en/#!iso:std:74591:en` (accessed on 1 January 2024).

[33] Todhunter, L.D.; Leach, R.K.; Lawes, S.D.; Blateyron, F. Industrial survey of ISO surface texture parameters. *CIRP J. Manuf. Sci. Technol.* 2017, 19, 84-92.

[34] Gadelmawla, E.S.; Koura, M.M.; Maksoud, T.M.; Elewa, I.M.; Soliman, H.H. Roughness parameters. *J. Mater. Process. Technol.* 2002, 123, 133-145.

[35] Whitehouse, D.J. The parameter rash-Is there a cure? *Wear* 1982, 83, 75-78.

[36] ASTM E284-17; Standard Terminology of Appearance. ASTM International: West Conshohocken, PA, USA, 2017. Available online: `https://cdn.standards.iteh.ai/samples/97708/7e9a7cafb4bb432f9458a1eeff471dcd/ASTM-E284-17.pdf` (accessed on 1 January 2024).

[37] Weber, C.F.; Spiehl, D.; Dörsam, E. Comparing measurement principles of three gloss meters and using them for measuring gloss on metallic embellishments produced by the printing industry. *Lux Jr.* 2021, 15, 327-341.

[38] R. Instruments. Rhopoint iq Goniophotometer Datasheet. 2023. Available online: `https://www.rhopointamericas.com/download/rhopoint-iq-goniophotometer-datasheet-english/` (accessed on 1 January 2024).

[39] Leach, R. (Ed.) *Optical Measurement of Surface Topography*; Springer: Berlin / Heidelberg, Germany, 2011.

[40] Conroy, M.; Armstrong, J. A comparison of surface metrology techniques. *J. Phys. Conf. Ser.* 2005, 13, 458.

[41] ISO 25178-2; Geometrical Product Specifications (GPS)-Surface Texture: Areal-Part 2: Terms, Definitions and Surface Texture Parameters. ISO: Geneva, Switzerland, 2021. Available online: `https://www.iso.org/obp/ui/en/#!iso:std:74591:en` (accessed on 1 January 2024).

[42] Todhunter, L.D.; Leach, R.K.; Lawes, S.D.; Blateyron, F. Industrial survey of ISO surface texture parameters. *CIRP J. Manuf. Sci. Technol.* 2017, 19, 84-92.

[43] Gadelmawla, E.S.; Koura, M.M.; Maksoud, T.M.; Elewa, I.M.; Soliman, H.H. Roughness parameters. *J. Mater. Process. Technol.* 2002, 123, 133-145.

[44] Whitehouse, D.J. The parameter rash-Is there a cure? *Wear* 1982, 83, 75-78.

[45] ASTM D3451-06; Standard Guide for Testing Coating Powders and Powder Coatings. ASTM International: West Conshohocken, PA, USA, 2006. Available online: `https://compass.astm.org/document/?contentCode=ASTM%7CD3451-06%7Cen-US&proxycl=https://secure.astm.org&fromLogin=true` (accessed on 1 January 2024).

[46] Deere, J. Appearance Standard JDHZ610. Available online: `https://studylib.net/doc/25779977/610` (accessed on 9 November 2023).

[47] Muralikrishnan, B.; Raja, J. Computational Surface and Roundness Metrology; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008.

[48] Michigan Metrology, L. Multiscale Surface Texture Analysis in Action. Available online: `https://michmet.com/multiscaleanalysis-in-action/` (accessed on 9 November 2023).

[49] BYK Instruments. Appearance of a Class A Surface: Orange Peel and Brilliance. Available online: `https://www.byk-instruments.com/en/t/knowledge/Orange-Peel-Paint` (accessed on 9 November 2023).

[50] DigitalMetrology. Bandify Multi-Band Surface Texture Analysis. Available online: `https://digitalmetrology.com/solution/bandify/` (accessed on 9 November 2023).

[51] Zygo. Mxâ¢ Surface Texture Parameters. Available online: `https://www.zygo.com/insights/blog-posts/-/media/ccf08d090bbe4c8eb98c1292e17d1d65.ashx?la=en&revision=662471a5-53c8-4839-beaa-8c00f88e6238` (accessed on 9 November 2023).

[52] FranÃ§ois, B. The areal field parameters. In *Characterisation of Areal Surface Texture*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 15-43.

[53] FPedregosa, G.; Varoquaux, A.; Gramfort, V.; Michel, B.; Thirion, O.; Grisel, M.; Blondel, P.; Prettenhofer, R.; Weiss, V.; Dubourg, J.; et al. Sklearn.Modelselection.Stratifiedshufflesplit. Available online: `http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedShuffleSplit.html` (accessed on 9 November 2023).

CHAPTER 5: (Paper 3) CENTER-LINE-TIME FUNCTIONS AND CRITICAL CONSTANTS FOR PREDICTING LASER POWDER BED FUSION MELT POOL DISTORTION USING ONE SURFACE TOPOGRAPHY MEASUREMENT

Overview of paper 3

The work in this chapter was presented at the 2023 joint special interest group meeting between EUSPEN and ASPE on advancing precision in additive manufacturing and is published in the proceedings [8].

This chapter includes the paper titled "Center-line-time functions and critical constants for predicting laser powder bed fusion melt pool distortion using one surface topography measurement," which introduces a novel approach to both measure and predict overmelting during laser powder bed fusion (LPBF). Using a center-line-time function (CLTF) to describe the time-dependent position(s) of the laser during part fabrication, the study evaluates the length of distorted solidified melt pools in eight rapid turnarounds (RTR) samples manufactured from nickel super alloy 625 using a commercial LPBF machine with vendor-recommended build parameters. To isolate the topography of interest, the programmed layer height for the build is used to inform the placement of a reference plane that enables the surface to be classified into distorted and non-distorted regions. Next, the size of the distorted region is related to the scanning strategy to determine a critical time/cooling constant. The constant describes the track-to-track processing conditions that result in the observed atypical surface formation during part fabrication. Using the analytical equations derived to describe the time-dependent locations of the laser relative to the parts center axis, the critical time constant (CTC) is used to predict the initiation, continuation, and termination locations of melt pool distortion in the different sample geometries. This research represents a significant step towards a quantitative procedure for characterizing LPBF surfaces by linking process conditions to topography measurements. The results offer valuable insights into understanding and addressing challenges in laser powder bed fusion processes.

Abstract

This work introduces a center-line-time function (CLTF) to characterize common scan strategies used in laser powder bed fusion (LPBF) that result in melt pool distortion. Eight rapid turnaround (RTR) samples, manufactured from nickel super alloy 625 using a commercial laser powder bed fusion machine with vendor-recommended build parameters, were utilized. The CLTF, in conjunction with a definition of melt pool distortion and a corresponding measurement procedure, was employed to evaluate the length of distorted melt pool regions in these samples. A critical time constant (CTC) was derived from the CLTF and measurement procedure, enabling the prediction of initiation, continuation, and termination locations of melt pool distortion for each sample geometry. Comparing the predicted and measured distorted melt pool lengths, an average error of 0.19 mm ± 0.77 mm was observed, with measured lengths ranging from 1.72 mm to 14.24 mm. The calculated CTC and CLTF values may vary depending on the material and machine parameters used for manufacturing the RTR samples. However, the methodology for determining the CTC and CLTF remains consistent, irrespective of material and machine parameters. These results demonstrate a step towards a quantitative procedure capable of characterizing the occurrence and location of melt pool distortion in both past and future builds. This approach provides practical insights that can aid in understanding and addressing melt pool distortion in laser powder bed fusion processes.

5.1 Introduction

Commercial laser powder bed fusion (LPBF) machines often employ fixed process parameters for specific materials, including laser power, scan velocity, and hatch spacing. [1] The vendor-recommended settings offer a general process window for the bulk regions of a build. However, previous studies by [2] have demonstrated that these fixed parameters lead to melt pool distortion in rapid turnaround regions (RTRs), typically found near stripe boundaries or narrow geometric features of the part. This melt pool distortion, resembling swelling [2-5], has been further investigated in recent work [6], which focused on novel part geometries to study the scan strategies and build conditions causing distorted melt pools. The resulting superelevation, where solidified regions exceed the powder layer thickness, poses a risk of impact with the recoater blade during subsequent layer spreading. [7,8] Moreover, variations in the depth of the melt pool, as observed in [6], can introduce irregularities in the subsurface microstructure, impacting both machine health and part quality. Despite the prevalence of distorted melt pool regions in LPBF manufacturing, there remains a need for a quantitative ex-situ procedure to consistently identify and measure these features. By combining prior knowledge of scan strategies and build parameters with reproducible measurements of distorted melt pool regions, it becomes feasible to enhance and evaluate the effectiveness of optimization efforts targeted at eliminating these regions

5.2 Methology

This paper builds upon existing conceptual models [2,6] of distorted melt pool formation in rapid turnaround (RTR) samples. Using coherent scanning interferometry (CSI), we manufactured and measured eight samples with different geometries. To characterize the distorted melt pool regions and laser trajectories used in sample production, we developed a center-line-time function (CLTF) and a corresponding measurement procedure. From these, we derived a critical time constant (CTC) to predict the length of distorted melt pool regions for each sample geometry. By comparing the predicted quantities with manual measurements, we assessed the accuracy of the predictions when utilizing a CTC value based on a distorted melt pool length measurement from a single or multiple RTR samples.

5.2.1 Rapid turn-around artifacts vs samples

RTR artifacts, such as the one shown in figure 5.1, consist of two single layer RTR samples connected by a rectangular waist built on top of a rectangular pedestal. The surface topography of an RTR artifact and a profile section along the center x-axis of the part are shown in figure 5.1(a) and figure 5.1(b), respectively. Elevated regions in the build direction caused by distorted melt pools on either end of the sample are shaded in red.

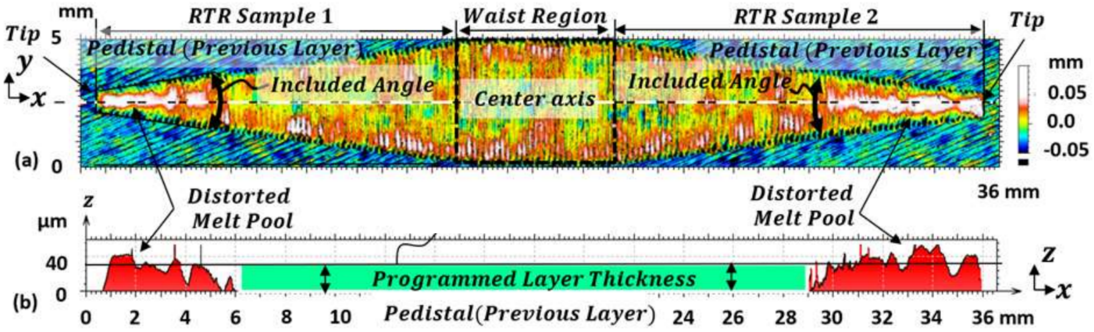


Figure 5.1: (a) Stitched CSI measurement of manufactured RTR artifact using Zygo Zegage Pro HR with a 5.5× objective showing elevated topography and melt pool distortion at the narrow ends of RTR samples. (b) Profile along the center axis of the artifact showing only isolated regions of elevated topography due to severe melt pool distortion

The length of a single RTR sample, denoted L_{RTR}, is controlled by a prescribed included angle, expressed as θ_{RTR}. Equation 5.1 provides a formal expression for the RTR sample length L_{RTR} as a function of the prescribed included angle θ_{RTR}, w, the width of the narrow region, and W, the width of the waist region.

$$L_{RTR}(\theta_{RTR}) = \frac{.5\sin\left(\frac{90^\circ-\theta_{RTR}}{2}\right)W}{\cos\left(\frac{90^\circ-\theta_{RTR}}{2}\right)} - \underbrace{\left[\frac{.5w}{\tan\left(\frac{\theta_{RTR}}{2}\right)}\right]}_{\Delta x}$$

(5.1)

The Δx term in equation 5.1 accounts for the RTR sample trapezoidal geometry. Dimensions of the pedestal upon which the RTR samples and waist region are built are presented in [5].

5.2.2 Rapid turn-around artifacts and acquisition details

A total of eight RTR samples were considered within included angles of 5°, 10°, 15°, 20°, 25°, 30°, and 35°. W was held constant at 5 mm and w held at 1 mm. Each sample is manufactured from nickel super alloy 625 (IN625) using an EOS M290 LPBF machine with vendor-recommended build parameters: laser power of 285 W, scan velocity of 960 mm/s, hatching spacing of 110 μm, and programmed layer thickness of 40 μm. Contouring and stripe boundaries were turned on. The scan strategy is designed such that during the manufacturing of the RTR sample the laser step-over direction is parallel to the x-axis of the RTR sample. The surface topography measurements of the entire sample area were obtained by stitching together multiple individual measurements using Zygo Nexview CSI with a 10× objective and 0.5× tube lens. Data processing was performed in the instrument's native software, using Mx ver. 8.0.0.26, with a 20% stitching overlap. Distorted melt pool regions were isolated, and their lengths were measured following the procedure described in the next section, utilizing MountainsMap^{TM} Version 10.0 software.

5.2.3 Measurement procedure for distorted melt-pool lengths

The following outlines a definition of a distorted melt pool and a procedure for measuring the length of a distorted melt pool in RTR samples, as depicted in figure 5.2. The purpose of this definition and procedure is to minimize variability in the measurement of the distorted melt pool length by eliminating operator subjectivity in determining the start and end points of the distorted melt pool with respect to the narrow end of an RTR sample. A distorted melt pool is defined as a continuous region of solidified material that exceeds a reference surface. This reference surface is parallel to a best-fit plane through the pedestal region and is offset by the programmed layer thickness ($H = 40\ \mu$m) along the z-axis. The distorted melt pool length is determined as the maximum perpendicular distance between the narrow end (i.e., tip) of the RTR sample and the end of the distorted melt pool region. By implementing this definition and measurement procedure, the assessment of the distorted melt pool length in RTR samples can be consistently and objectively performed, reducing variability in the measurement process.

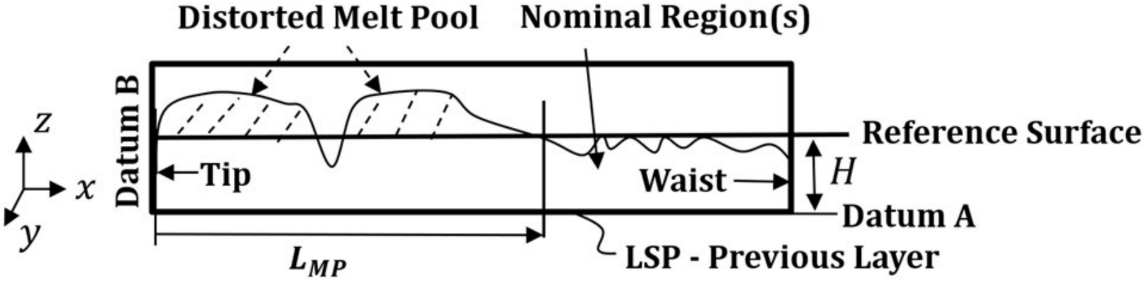


Figure 5.2: Schematic defining the distorted melt-pool length ($\hat{L_{RTR}}$). The tip of the RTR geometry is datum B. The end of the distorted melt pool region is the intersection between the surface topography (dashed region) and reference surface offset from a least-squares plane (LSP) through pedestal region (datum A) by the programmed layer thickness H.

5.2.4 Derivation of a center-line-time function and critical constants for predicting distorted melt pool regions

By leveraging the mirror symmetry about the center x-axis of the RTR artifacts, exemplified in figure 5.1, a CLTF that characterizes the time needed for the laser to cross the user- defined center-line-axis twice, at any sequential scan increment n, is derived using prior knowledge of the designed laser trajectories, scan velocity (v_s), hatch spacing (Δh), and nominal variables controlling the geometry of the RTR artifact. As shown in figure 5.3, this is achieved by summing the durations of times t_a, t_b, t_c that describe the three sequential paths the laser must traverse before arriving back at the same nominal position along the center y-axis of the artifact but displaced along the x-axis by a fixed distance, Δh. If the laser begins manufacturing the RTR artifact starting at the left end (i.e., diverging case) it follows that $t_a < t_c$ for each track $n \in [0, \frac{L_{RTR1}}{\Delta h}]$, and for the waist region it follows $t_a = t_c$ on the interval $n \in [\frac{L_{RTR1}}{\Delta h}, \frac{L_{RTR1}}{\Delta h} + \frac{W}{\Delta h}]$ until reaching the wide end of the RTR sample on the right (converging case) where $t_a > t_c$ for each sequential track, $n \in [\frac{L_{RTR1}}{\Delta h} + \frac{W}{\Delta h}, \frac{L_{RTR1}}{\Delta h} + \frac{W}{\Delta h} + \frac{L_{RTR2}}{\Delta h}]$. The skywriting time ($t_b$) where the laser is turned off may vary according to the part geometry and build parameters. Relative to t_a and t_c, small variation in t_b is considered negligible and the average time is assumed constant (i.e., $\Delta h \cdot v_s^{-1} \in n$). However, further study will be required to measure and quantify the full range and impact of skywriting time variations.

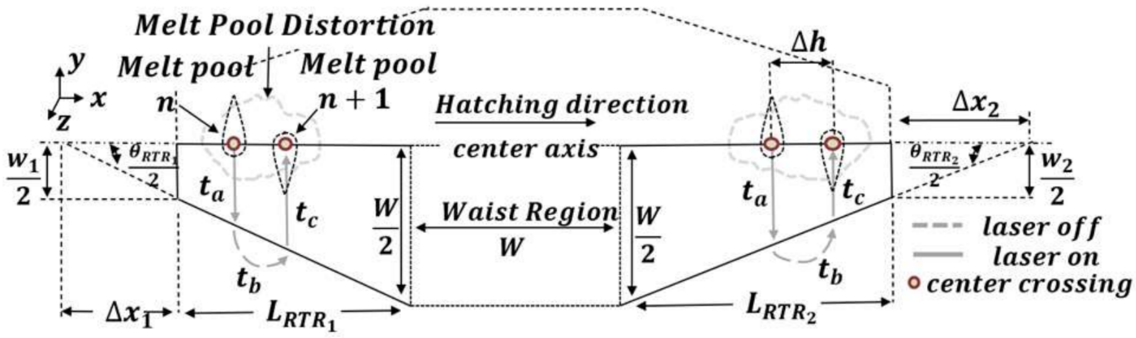


Figure 5.3: Conceptual illustration of scan strategy used to manufacture RTR samples and examples of diverging (left) and converging (right) cases.

The CLTF for any RTR sample geometry in the case of a diverging scan strategy, depicted in figure 5.3, is described by equation 5.2.

$$t(n) = \underbrace{\frac{\tan(\theta_{RTR})\left(\frac{w}{2\tan(\theta_{RTR}/2)} + \Delta h(n)\right)}{2v_s}}_{t_a} + \underbrace{\frac{\Delta h}{v_s}}_{t_b} + \underbrace{\frac{\tan(\theta_{RTR})\left(\frac{w}{2\tan(\theta_{RTR}/2)} + \Delta h(n+1)\right)}{2v_s}}_{t_c}$$

(5.2)

In this equation, $t(n)$ represents the CLTF for the sample and n is a user-defined scan line number that ranges from 0 to $\frac{L_{RTR}}{\Delta h}$. The scan line number acts as an index to track the evolving laser path trajectories along the center axis of the sample. All the variables for calculating the CLTF (e.g., included angle θ_{RTR}, and width W, defining the RTR sample geometry and build parameters including the scan velocity v, and hatch spacing Δh are known or assigned prior to the build. Intuitively, if the duration of time for the laser to cross the center axis of the part twice is less than the time for the melt pool to completely solidify from the previous scan line (i.e., melt pool n shown in figure 5.3) some amount of melt pool distortion is expected. However, this may be an underestimate of the minimum duration of time to completely avoid melt pool distortion, as the previous statement neglects effects from residual heat in fully solidified material. Still, a conservative assumption regarding the LPBF build process is if any incremental scan strategy trajectory evaluated at an increment n results in a CLTF solution that is less than the time for a nominal melt pool (i.e., quasi-steady state melt pool size in regions free from rapid turnarounds and residual heating effects) to solidify, then some melt pool distortion between adjacent tracks is expected. This is also consistent with the underling mechanisms hypothesized to create the "double wide" weld tracks along stripe boundaries and RTR regions of the build shown in [2]. Calculating a geometry dependent critical constant, n_{crit}, (i.e., the number of scan lines that can be divided into measured length of the RTR sample's distorted melt pool length, L_{MP}) and substituting the value into equation 5.2, a critical time constant (CTC) value, $t_{crit} = t(n)$, where $n = \frac{L_{MP}}{\Delta h} = n_{crit}$, is acquired and assumed to be independent of the RTR sample geometry. Instead, the calculated value of t_{crit} is a result that characterizes the thermal conditions created by the alloy, machine, and build parameters described in section 5.2.2. This is because t_{crit} serves as a measurand that encompasses all these factors. By equating the right-hand side of equation 5.2 to the value of t_{crit} and rearranging the equation to solve for n, we obtain an expression for the track number at which melt pool distortion terminates. Multiplying this non-dimensional expression by the hatch spacing (Δh) yields a closed-form solution to calculate the length of the distorted melt pool in millimeters. This calculation is applicable to samples with various geometries but manufactured with the same build parameters as the RTR sample used to determine t_{crit}. The predicted length of the distorted melt pool (L_{MP}) for any RTR sample geometry, denoted as ($\widehat{L_{MP}}$), is determined using equation 5.3, which leverages the critical time constant (CTC) value. $\widehat{L_{MP}}$ represents the distance between the narrow end of an RTR sample (datum B) and the point where the distorted melt pool originates or ends, depending on whether it is converging or diverging. Equation 5.3 is as follows:

$$\widehat{L}_{MP} = (\frac{-0.5\Delta h - \frac{\Delta h}{tan(\theta_{RTR})} - \Delta x + \frac{t_{crit} v_s}{tan(\theta_{RTR})}}{\Delta h})\Delta h \tag{5.3}$$

In this equation, Δh, v_s, and t_{crit} are constants, while θ_{RTR} and Δx from equation 5.1 can be adjusted to predict the length of the distorted melt pool for different RTR sample geometries.

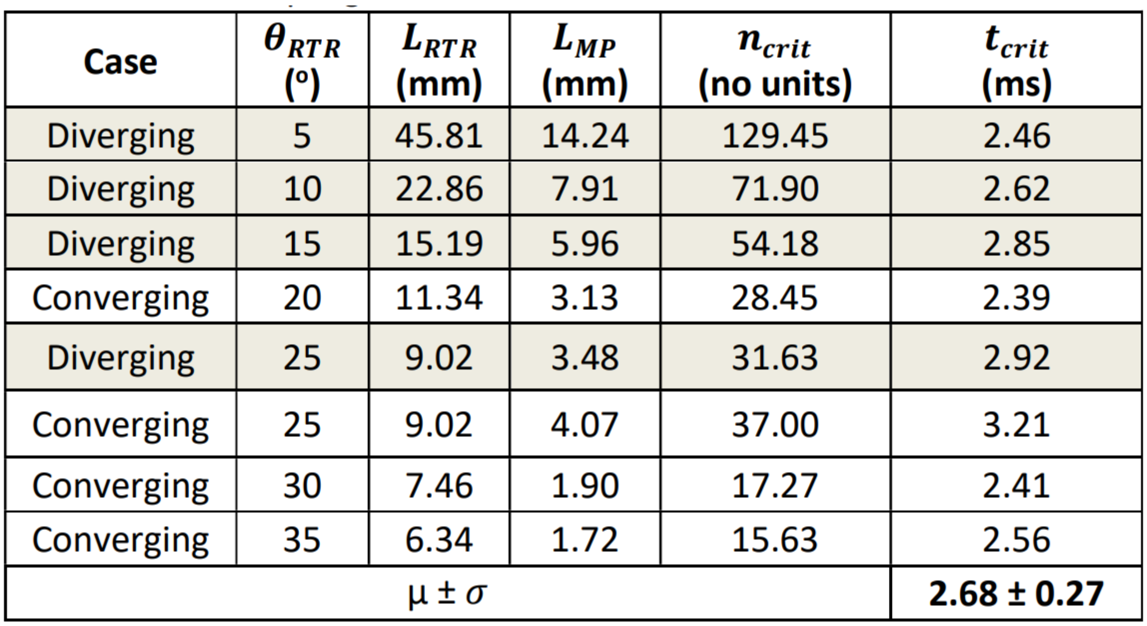
5.3 Results

5.3.1 Tabulated measurements & critical constants

Table 5.1 lists the distorted melt pool length measurements (L_{MP}) for the eight RTR samples using the definition and procedure described in section 2.3. Measurements related to converging cases (i.e., 20°, 25°, 30°, 35°) are shown in white and diverging cases are highlighted as grey rows (i.e., 5°, 10°, 15°, 25°). The critical constants (n_{crit}) were calculated by dividing the measured distorted melt pool lengths (L_{MP}) by the programmed hatch spacing (Δh) of 110 μm. The individual CTC, t_{crit}, values were obtained by evaluating equation 5.2 according to the calculated n_{crit} and prescribed θ_{RTR} for each sample.

From table 5.1, the longest reported distorted melt pool length of 14.24 mm occurs for the 5° included angle and the shortest recorded length of 1.72 mm occurs for the 35° sample (i.e., the widest angle). The distorted melt-pool measurements exhibit an inverse relationship to the included angle of the RTR sample. Comparing the relative ranges of t_{crit} and n_{crit} the variation of the latter is large and dependent on the sample geometry and build parameters, whereas the standard deviation of the former is small and independent of the sample geometry, supporting the assumption that the CTC value, t_{crit}, describes a geometry- agnostic estimate of the minimum duration of time, 2.68 ms ± 0.27 ms in this case, that needs to elapse between sequentially formed melt pools to avoid melt pool distortion when manufacturing IN625 using the machine and build parameters described in section 5.2.2. A disparity of 0.59 mm

Table 5.1: Tabulate measurements of distorted melt pool lengths for different RTR sample geometries



Case	θ_{RTR} (°)	L_{RTR} (mm)	L_{MP} (mm)	n_{crit} (no units)	t_{crit} (ms)
Diverging	5	45.81	14.24	129.45	2.46
Diverging	10	22.86	7.91	71.90	2.62
Diverging	15	15.19	5.96	54.18	2.85
Converging	20	11.34	3.13	28.45	2.39
Diverging	25	9.02	3.48	31.63	2.92
Converging	25	9.02	4.07	37.00	3.21
Converging	30	7.46	1.90	17.27	2.41
Converging	35	6.34	1.72	15.63	2.56
$\mu \pm \sigma$					**2.68 ± 0.27**

is observed between measurements of the diverging and converging cases for the 25° sample geometry that could be related to changes in residual heat effects caused by the laser step-over direction. However, whether this difference exceeds the dispersion of values that can reasonably be attributed to the realization of the measurand (i.e., distorted melt pool length) is yet to be determined. Ongoing work is being conducted to provide a provisional assessment of the Type A and Type B uncertainties associated with the measurement results.

5.3.2 Predictions vs measurements for other RTR geometries

The dashed blue line in figure 5.4 provides a plot of RTR sample length according to a prescribed included angle given by equation 5.1. Blue text/diamonds correspond to the eight manufactured samples. For example, the manufactured sample with an included angle, θ_{RTR}, of 5° has a length, $_{RTR}$, of 45.81 mm. The initiation/end of a distorted melt pool region relative to the narrow end of the sample is indicated with white triangles, △ and , representing converging and diverging cases, respectively. The vertical distance between a triangle and blue diamond corresponds to the measured distorted melt pool length (L_{MP}) tabulated in table 5.1. Values are given as black text. The predicted initiation/end regions, according to equation 5.3, are shown as red crosses ($\times$) with predicted melt pool lengths, $\widehat{L_{MP}}$, given as red text. For example, the measured distorted melt pool length (L_{MP}) for the 5° sample is 14.24 mm and the predicted length ($\widehat{L_{MP}}$) is 15.39 mm. The predicted distorted melt pool lengths given by the dashed red line are based on the CTC value, $t_{crit} = 2.57$ ms, obtained from a single distorted melt pool length (L_{MP}) measurement of 1.72 mm corresponding to the 35° RTR sample. The shaded red regions describe the solution space of predictions if any one of the t_{crit} values shown in table 5.1 were used as the basis for predicting the distorted melt pool length for all other RTR geometries.

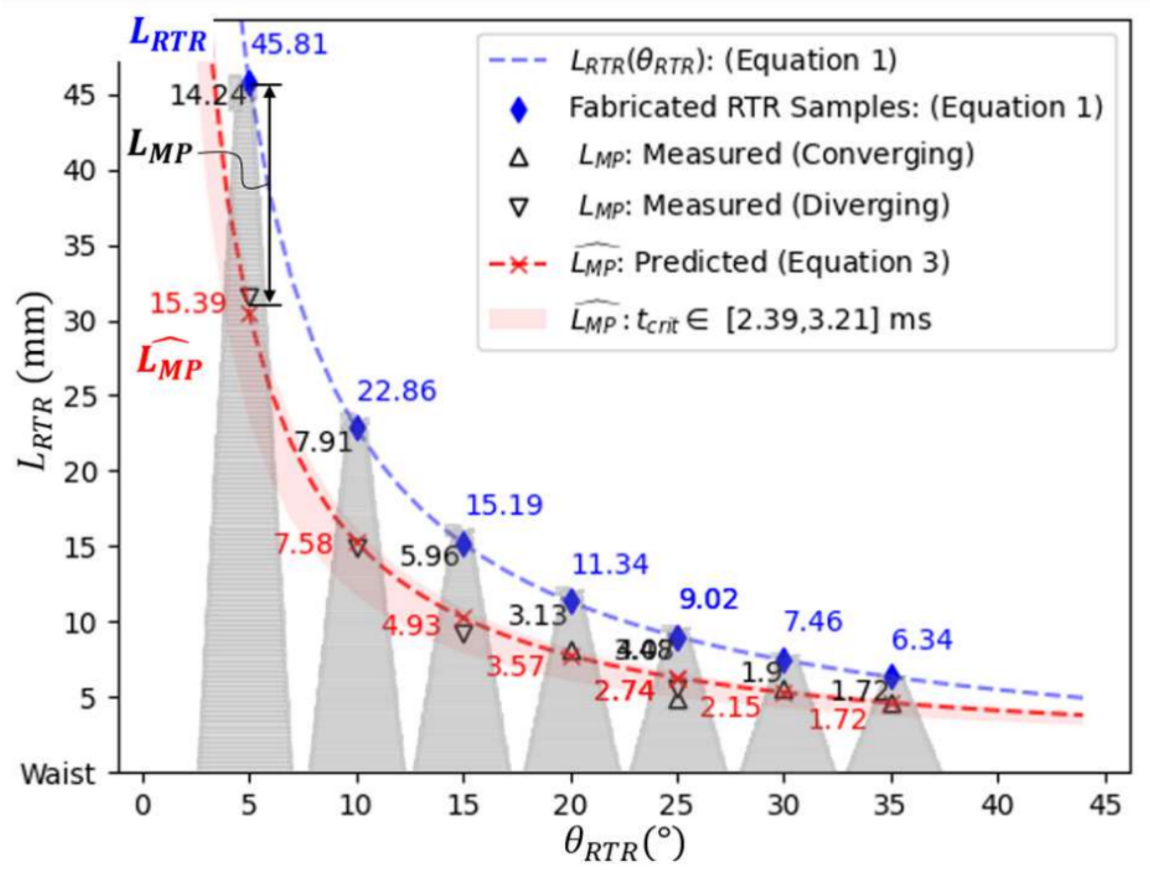


Figure 5.4: RTR sample length as a function of included angle, 1 mm tip and 5 mm waist, with a comparison between the measured versus predicted initiation/end regions and distorted melt pool lengths relative to the tip of the fabricated RTR samples

Comparing the measurements of distorted melt pool length with the predictions (i.e., L_{MP} - $\widehat{L_{MP}}$) for the eight RTR samples reveals an average error of 0.19 mm with a standard deviation of 0.77 mm. This indicates symmetric errors centered around the mean, with a standard deviation of approximately 7× the hatch spacing. The largest discrepancy of 1.15 mm is observed for the RTR sample with a 5° included angle. Utilizing the average t_{crit} value of 2.68 ms from table 5.1 results in an average error of -0.22 mm with a standard deviation of 1 mm. Employing the maximum calculated value of 3.21 ms for t_{crit} shifts the predicted trend line to the lower end of the shaded region, while using the minimum value of 2.39 ms shifts the predictions to the upper end. The former CTC value tends to overestimate the distorted melt pool lengths compared to the latter

5.4 Summary and conclusions

A distorted melt pool length measurement of a single RTR sample was used to derive a CTC value that is directly linked to the thermal conditions created by the alloy, machine, and build parameters described in section 5.2.2. The CTC value is used in conjunction with a CLTF to predict the locations melt pool distortion is expected to initiate, continue, or terminate over a large range of RTR sample geometries. An average error of 0.19 mm ± 0.77 mm was recorded between predicted and measured melt pool lengths, the latter ranging from 1.72 mm to 14.24 mm. The outlined methodology for deriving CLTFs, measuring distorted melt pool lengths, and obtaining a CTC value provides an intuitive approach to characterize why and where melt pool distortion may occur in future or past builds. For example, if manufacturing IN625 using an EOS M290 with vendor recommended settings, the duration of time for the laser to turn around (i.e., skywriting time) at each sequential scan track is consistently less than the range of CTC values obtained using the outlined approach. This implies that the distorted melt pools observed in [2] at stripe boundaries and RTR regions of the build are/can be expected. Increasing the skywriting time to be greater than the calculated CTC value is hypothesized to eliminate melt pool distortion in these areas. The calculated value for this CTC and CLTF may vary depending on the material and machine parameters used to manufacture the RTR samples, but the method of determining the CTC and CLTF is the same regardless of material and machine parameters. The results of this work demonstrates a practical step toward a quantitative procedure that is independent of the specific build, alloy, and machine used. Researchers and manufacturers can utilize this procedure to derive a critical time constant (CTC) value for the characterization, evaluation, and prediction of build parameters and scan strategies that either produce or avoid distorted melt pool regions for different materials. Continuing efforts involve further development, evaluation, and utilization of this CLTF and CTC. Future investigations will explore the links between the calculated CTC and CLTF values with in-situ thermal conditions, as well as melt pool morphology. Additionally, the research will encompass ex-situ measurements of surface topography and subsurface microstructures in both distorted melt pool and nominal regions of RTR samples.

5.5 References

[1] Yeung, Ho, and Brandon Lane. "A residual heat compensation based scan strategy for powder bed fusion additive manufacturing." *Manufacturing Letters* 25 (2020): 56-59.

[2] Fox, Jason, Christopher Evans, Aarush Sood, Romaine Isaacs, Brigid Mullany, Angela Allen, and Ed Morse. "Weld Track Distortion in Laser Powder Bed Fusion of Nickel Superalloy 625." *Proceedings of the ASPE-euspen Special Interest Group Meeting: Advancing Precision in Additive Manufacturing*, Knoxville, TN, US, (2022).

[3] Vasileska, Ema, et al. "Layer-wise control of selective laser melting by means of inline melt pool area measurements." *Journal of Laser Applications* 32.2 (2020).

[4] Sames, William. "Additive manufacturing of Inconel 718 using electron beam melting: processing, post-processing, & mechanical properties." Diss. (2015).

[5] Scime, Luke, et al. "Layer-wise anomaly detection and classification for powder bed additive manufacturing processes: A machine-agnostic algorithm for real-time pixel-wise semantic segmentation." *Additive Manufacturing* 36 (2020): 101453.

[6] Fox, Jason C., Chris J. Evans, Jordan S. Weaver, and Jesse K. Redford. "Surface Topography and Melt Pool Behavior in Rapid Turnaround Regions of Laser Powder Bed Fusion Additive Manufacturing of Nickel Superalloy 625." *CIRP Annals* (2023).

[7] Grasso, Marco, et al. "In-process monitoring of selective laser melting: spatial detection of defects via image data analysis." *Journal of Manufacturing Science and Engineering* 139.5 (2017): 051001.

[8] zur JacobsmÃŒhlen, Joschka, Stefan Kleszczynski, Gerd Witt, and Dorit Merhof. "Elevated region area measurement for quantitative analysis of laser beam melting process stability." In *2015 International Solid Freeform Fabrication Symposium.* University of Texas at Austin, (2015).

CHAPTER 6: (Paper 4) A CRITICAL COOLING CONSTANT TO CLARIFY TRACK-TO-TRACK OVERHEATING IN LPBF AM OF IN625

Overview of paper 4

The work in this chapter will be submitted to the 2024 joint special interest group meeting between EUSPEN and ASPE on advancing precision in additive manufacturing in Golden, USA. This chapter generalizes the analytical centerline time functions described in Chapter 5 to the entire scan strategy, introducing a new point-wise characterization of LPBF scan strategies referred to as "cross-over time." In this case, cross-over time maps of the scanning strategies used to fabricate a new set of parts are used in combination with the derived critical cooling constant(s) determined in Chapter 5 to isolation all the areas within a layer that are expected to exhibit atypical melt-pool-surface formation, as opposed to only one region of the part. This chapter also clearly demonstrates how the methodology can be leveraged for feed-forward control and the determination of tuning parameters employed by commercial LPBF software to mitigate thermally induced flaw formation. The contributions of this work include the linking of topography observations directly to process outcomes that can be calculated directly from scan strategy data. Furthermore, the developed methodology does not depend on arbitrary constants or coefficients making it interpertable and easy to understand. Overall, a single surface measurement enabled useful predictions of thermal flaws expected to occur in new geometries.

Abstract

This study introduces a new point-wise parameterization of laser powder bed fusion (LPBF) scan strategies. Preliminary results suggest the approach could be a scalable

approach to predict areas within layers of a part known to elicit heterogeneity in micro-structure and surface topography due to overheating.

6.1 Introduction

In laser powder bed fusion (LPBF) processes, the back-and-forth raster scanning method produces varying melt pool sizes along the laser's travel path. Specifically, turn-around areas near the ends of tracks encounter localized increases in residual heat, causing the merging of melt pools from adjacent tracks [1, 2]. To mitigate these well-known overheating effects and avoid entering a keyhole mode at the ends of tracks due to slow down of the laser, the practice of "skywriting" has emerged, where the laser is intentionally turned off upon reaching the edges of parts to maintain a consistent laser power input and scanning speed throughout the trajectory. This simple alteration to LPBF scan strategies has proven effective in reducing near-edge porosity and has become widely adopted in industry [3,4,5]. However, even with the incorporation of skywriting, systemic residual heat effects, and melt pool inconsistencies within a layer or at track ends persist. For example, scanning progressions within narrow regions of a layer can lead to continuous merging of melt pools over 50+ tracks and increase the solidified surface height by more than 125% [2]. Newer software features developed by EOS aim to further mitigate this type of interlayer overheating created by short-hatching vectors. This includes the power reduction factor (PRF), which is a percentage set by the user that linearly reduces power depending on the length of a scan vector. The other, known as time homogenization, deals with residual heating effects by introducing waiting times between vectors (i.e., additional skywriting time). Works by [6] provides an investigation of these parameters and their effect on part/build quality relative to default scanning parameters for maraging steel. Aside from different implementations of the heat equation and machine learning techniques notable nonproprietary methods developed to combat residual heating effects include the geometric conductance factor (GCF), which approximates the effect of geometry-based heat accumulation due to small, solidified volumes enclosed by powder [7]. The second is the residual heat factor (RHF) introduced by [8] that characterizes dynamic reheating effects by parameterizing the relationship between each programmed point and its neighbors within the scan strategy using a time and distance threshold. Both methods normalize the GCF or RHF values to modulate laser power and reduce melt pool variability within a layer. However, there's still no clear quantitative measure for determining optimal settings used by all these feed forward control methods. This poses a challenge for users seeking to quickly optimize their processes. Prior work in [9] demonstrated a methodology to quantitively measure and analytically predict the formation of enlarged melt pools in a series of single layer trapezoidal geometries built on top of a 10.76 mm tall pedestal (see Figure 1 in [2]). Localization of the overheated regions in these samples was achieved by determining a critical time constant (t_{crit}) from a single process informed surface topography measurement. Building on this work, the next sections describe how the methodology and derived t_{crit} value translate to completely different sample geometries produced with the same machine, alloy, and process parameters.

6.2 Methodology

6.2.1 Cross over time mapping

Cross over time (COT) is a pointwise parameterization of the scan strategy that refers to the amount of time that elapses between the formation of adjacent (track to track) melt pools. For example, in Figure 6.1, as a typical "serpentine" scan strategy is executed, consider the duration of time Δt_i that elapses before the laser arrives back in the same position in x but offset by the programmed hatch spacing in the y-direction. This is referred to as the "cross over time" or COT for short.

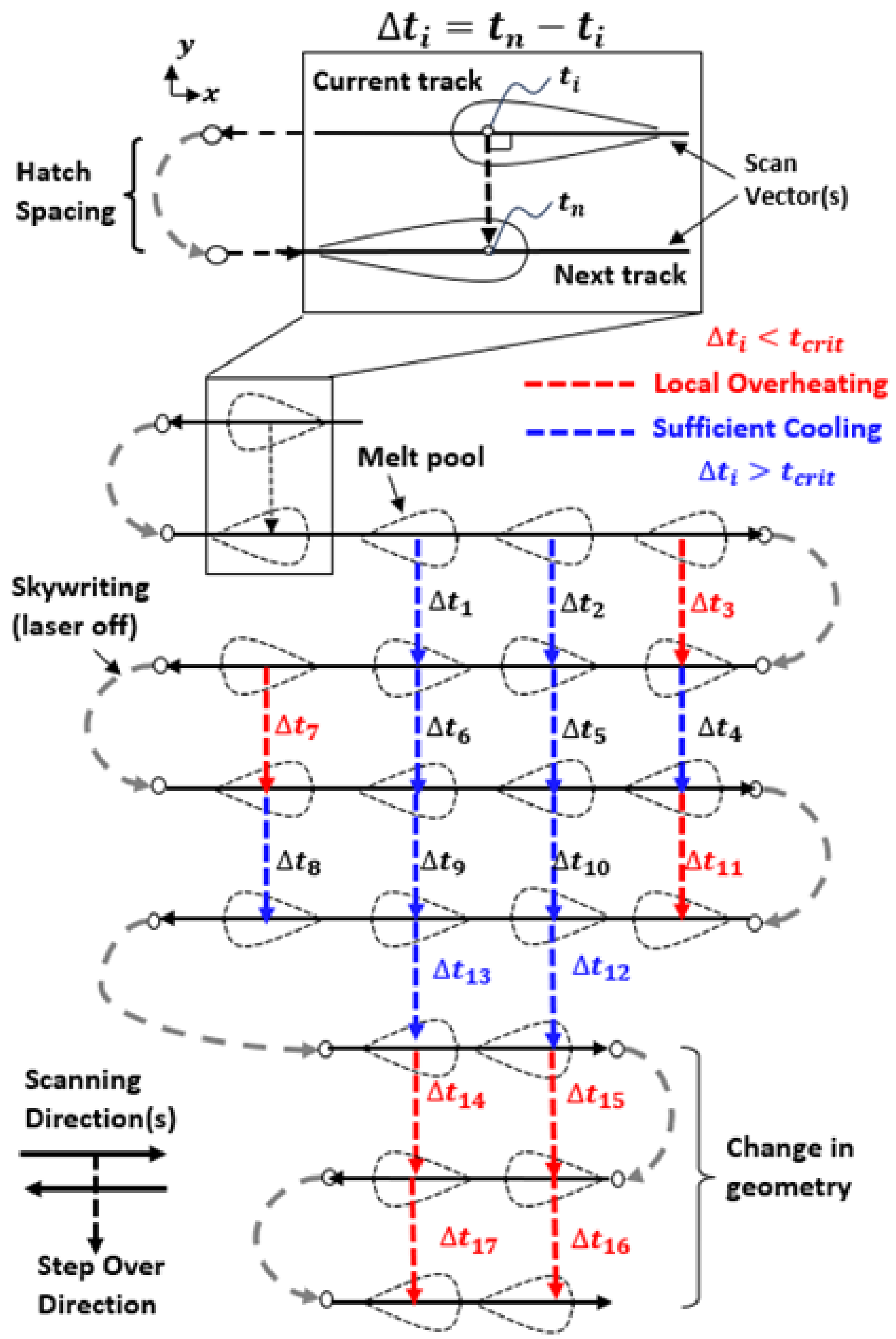


Figure 6.1: (top) Example of cross over time calculation for a programmed position in the scan strategy. (bottom) Illustration of cross over time map representing local cooling times at each programmed position for an arbitrary scan strategy.

A COT map (see bottom of Figure 6.1) for a scan strategy is generated by calculating a COT value at each programmed position where the laser is on (e.g., Δt_1,Δt_2 ... etc.). Notice that local COT values near turn arounds (Δt_3,Δt_7,Δt_{11}) or areas with shortening scanning vectors (Δt_{14},Δt_{15},Δt_{16},Δt_{17}) shown as dashed red arrows will be less than areas far from turn around regions (blue dashed arrows). If local COT values are necessarily small, there will be insufficient cooling time to enable the deposited heat to properly evacuate. Results presented in [9] suggest that at any given scan position where the COT is less than a critical value, local overheating and melt pool development along the step over direction will occur. The critical value (t_{crit}) represents the minimum cooling time required to prevent the latter.

6.2.2 Process parameters & data collection

The IN625 parts designed for this study, detailed in Figure 6.2, were manufactured using the same EOS M290 machine and vendor-recommended process parameters used to fabricate samples detailed in prior work [2,9]. This includes skywriting, a laser power 195 W, scanning speed of 965 mm/s, hatch spacing of 0.11 mm, programmed layer thickness of 0.04 mm, and 67 degree layer wise rotations. Each part (P1, P2, P3, P4) consists of one or more unique single-layer geometries built on top of a 10.76 mm pedestal.

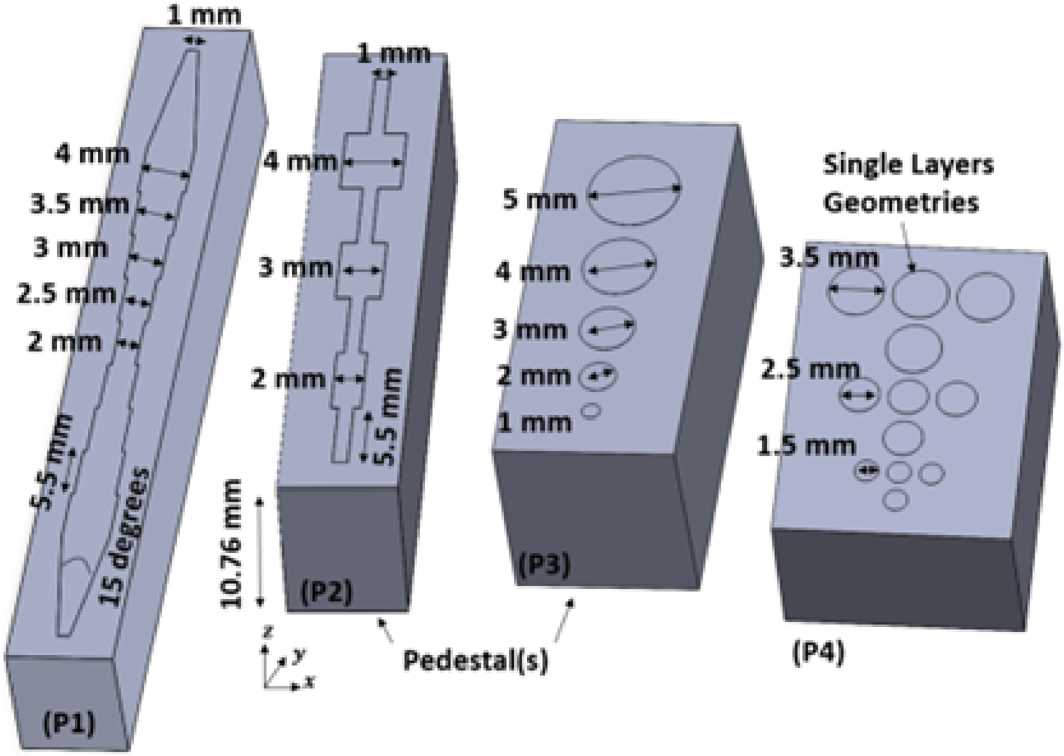


Figure 6.2: Examples of the single layer parts built on top of a 10.76 mm pedestal manufactured for the study.

Scan strategy data was acquired from a Sigma Labs system installed on the EOS M290 machine. Commanded signals sent to the laser and galvos are sampled at 100 khz (i.e., every 10 μs) during part fabrication. For convenience, the data was post-processed in Python to calibrate the voltage signals into commanded x-y positions. However, COT values can be determined directly from the voltage signals. Next, COT maps were calculated for each scan strategy used to produce the parts, as detailed in Figure 6.2. After thresholding the COT maps using t_{crit} values derived in [9] the result is compared to stitched top surface bright field images of the samples collected with a Zeiss AxioImager.Z2 microscope equipped with a 5× objective (sampling interval of 3.45 μm) and processed using native ZenCore 3.2 software.

6.3 Results

Figure 6.3 shows a modified version of the single layer part geometry introduced in [2] that was used to derive t_{crit} in [9]. The two ends have 15 degrees included angles and converge to a 1 mm tip. The top image represents the COT map computed for the scan strategy. The center image is the COT map after removing points where the COT value is less than the largest t_{crit} value of 3.5 ms derived in [9]. The remaining dark red regions indicate the expected areas to form large melt pools that persist across multiple tracks along the step over direction.

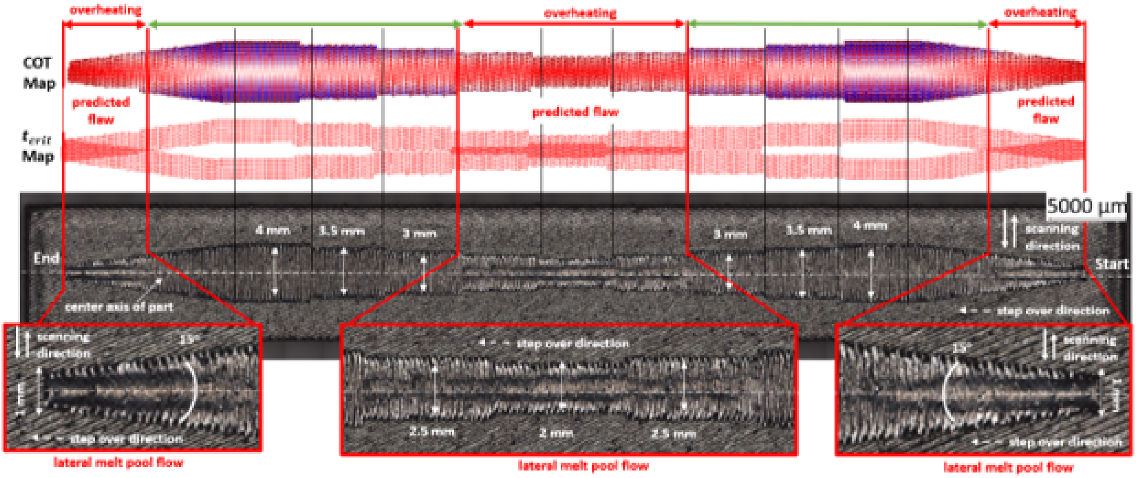


Figure 6.3: Examples of predicted and observed overheated in fabricated samples. (top) Cross over time map. (center) Cross over time map after thresholding values using t_{crit}= 3.5 ms. (bottom) Bright-field image of the fabricated part.

The predicted length of overheated regions in Figure 6.3 given by the COT map after thresholding show good agreement with the fabricated sample. Notice the initiation and termination sites of the overheated regions that develop in the far left and right side of the sample coincide with the same locations predicted by the t_{crit} map. This is also the case for the center most region, where the stripe width is less than 3 mm.

The sample in Figure 6.4 consists of a series of 2-, 3- and 4-mm stripes connected by a smaller 1 mm stripe. As expected, the lateral flow of material along the stepover direction persists in both the 1 mm and 2 mm stripe sections but immediately dissipates at the transition point from the 1 to 3 mm stripe. This is to be expected, as the cross-over times (taken along the center axis) of the 3 mm and 4 mm stripe is greater than t_{crit}.

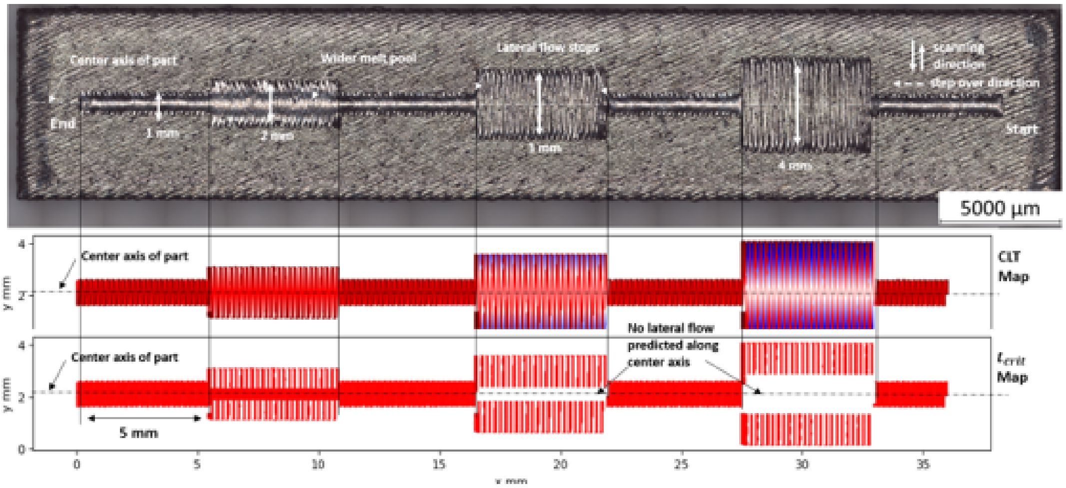


Figure 6.4: (top) Series of 2-,3-,4 mm stripes connected by a 1 mm stripe. (center) Cross over time map. (bottom) COT map after thresholding using t_{crit}= 3.5 ms.

The abrupt transition in lateral flow of the melt pool at the intersection of the 1 mm and 3- or 4-mm stripe depicted in Figure 6.5 suggests this melt pool behavior can be controlled on a track-to-track basis. Although the lateral flow of material along the center axis of the 3- and 4-mm stripes is eliminated; notice the t_{crit} map (bottom image Figure 6.4) suggests track to track melting is still expected to take place near the ends of tracks. This agrees with the systemic a-periodic "doublewides" show in Figure 6.5 that also appear in all previously manufactured samples [2,9]. Notice that the lateral melt pool flow at the edges of the parts does not persist across more than 1 track because every other track has time to sufficiently cool according to t_{crit}. Previous works by [10] referred to these cases as type 1 and type 2 melt pools.

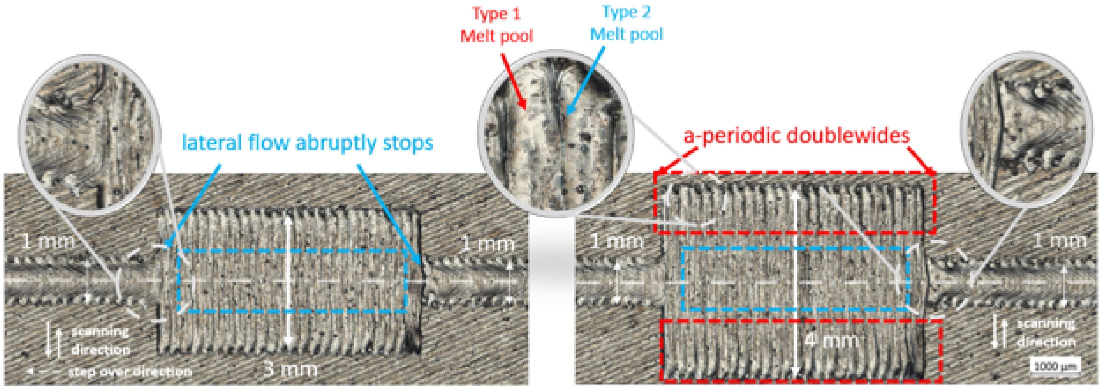


Figure 6.5: (left) Example of transitions from 1 to 3 and 3 to 1 mm. (right) Example of 1 to 4 mm and 4 to 1 mm transition.

Asymmetries in the scanning condition, stepping out of or into a narrow region (diverging vs converging) create different melt pools along the stepover direction. The converging cases tend to show less lateral flow than diverging cases. This can be observed in Figure 6.6, which includes a series of disks with diameters of 1-,2-,3,4-, and 5 mm.

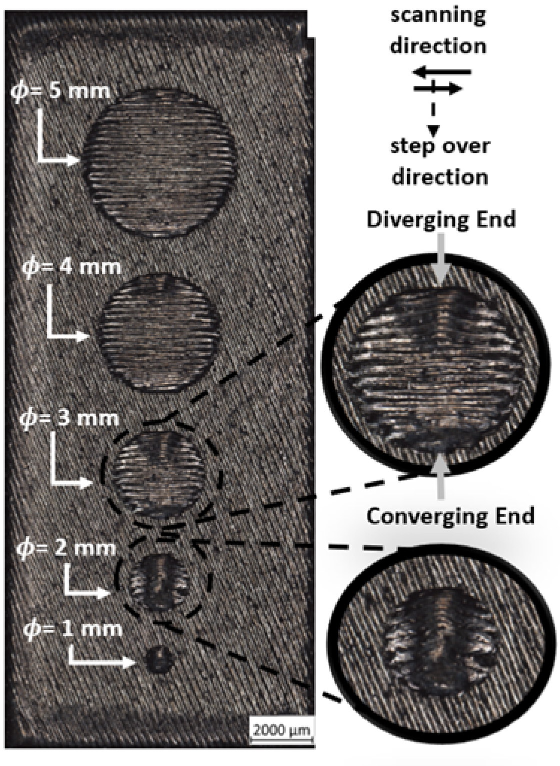


Figure 6.6: Brightfield image of the fabricated 1-,2-,3-,4-, and 5 mm disks.

In the case of the converging scanning condition (stepping into the narrow region), a larger amount of solidified material trails behind the stepover compared to the diverging case. This is because for the diverging case, the initial track begins in powder, and all newly solidified material that trails behind the step over direction was just recently formed and is presumably still at an elevated temperature. The local diffusivity is decreased in this case. This could explain the difference between the asymmetric overheating which occurs at the top and bottom of the disks in Figure 6.6.

Lateral melt pool development in the top center 3.5 mm disk show in Figure 6.7 is reduced compared to the single 3.5 mm disk below. Adding the additional disks to the left and right artificially increases the COT along the center axis of the middle disk, reducing the lateral flow.

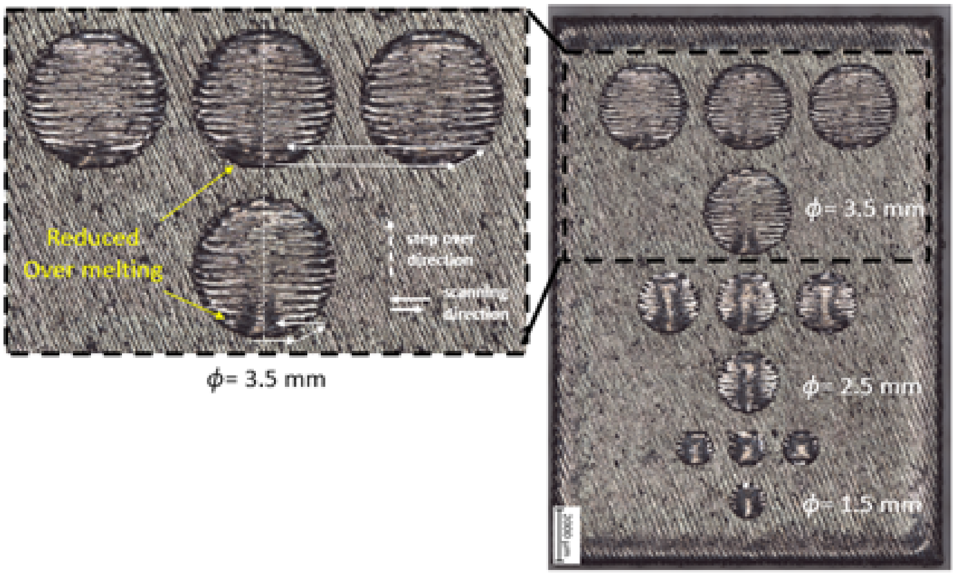


Figure 6.7: (right) Examples of fabricated 3.5-,2.5,1.5 mm disks. (Left) Example of reduced melt pool flow along the stepover direction in 3.5 mm diameter disks due to artificially increasing skywriting time.

6.4 Discussion

The initial derivation of t_{crit} in [9] assumed that skywriting time at each turn around could be approximated as the hatch spacing divide by the programmed scan velocity, giving a constant value of 0.12 ms. After collecting the scan strategy data from the machine, it was found that the average turnaround time is closer to 0.3 ms. Using this information to recalculate the t_{crit} values described in [9] increased the largest calculated t_{crit} value from 3.1 ms to 3.5 ms. Other factors such as build height, exposure time, and overhang conditions are expected to change the effective value of t_{crit}. Further study is required to understand how much of an effect these conditions have on the predictions and whether a straightforward analytical relationship exists. However, the current constant value approximation of t_{crit} still demonstrates the ability to characterize the machine/alloy/parameter combination used to fabricate the analyzed parts. This is because t_{crit} is a measurand that depends on all these factors. While a COT map provides insight into the scanning conditions, t_{crit} provides the means to directly link COT values to process outcomes (i.e. merging of melt pools between adjacent tracks). This is an important aspect to consider in the context of interoperability as LPBF machines with different galvo systems will have different response characteristics. For example, the large, overheated regions that span multiple tracks observed in the samples shown here have not been observed in similar geometries produced with same alloy, power, and scan speed on the NIST additive manufacturing and metrology test bed (AMMT). This is to be expected since the nominal skywriting time on the AMMT is 4.24 ms, which is not only greater than t_{crit} but about 14 times slower than the EOS M290 machine. However, when the AMMT is upgraded with galvos that are 100 times faster than the previous system itâs expected that the skywriting time will need to be artificially increased to avoid systemic overheating without changing power and scan speed. The t_{crit} parameter does not depend on or use any non-physical coefficients to make predictions, instead surface topography signatures are directly linked to a process condition which can be calculated and measured.

6.5 Summary and conclusions

Deriving an estimate of t_{crit} for the given process (e.g., EOSM290 machine, IN625 alloy, vendor recommended parameters) required a single build of one or more hexagonal samples detailed in [2]. Next, the surface topography measurement and calculation described in [9] is carried out to determine t_{crit}. If t_{crit} and a scan strategy is known, COT maps can be calculated and subsequently thresholded to isolate overheated regions within a layer. Pointwise COT information can also be used to mitigate local overheating by increasing the skywriting time in the predicted areas to be greater than t_{crit}. Homogenizing a typical LPBF scan strategy once t_{crit} is known is straightforward. Determine the minimum COT of a programmed position along a given scan vector and then subtract this value from t_{crit}. The result is the amount of additional skywriting time that needs to be added at the end of the track to avoid overheating, or the skywriting time can simply be set to 3.5 ms. The EOS time homogenization software feature already performs this type of scan strategy correction. Works by [6] found that the use of time homogenization to manufacture EOS Maraging Steel MS1 decreases the maximum residual stress by 70% and increases surface hardness compared to default parameters. Type 1 melt pools and merging of adjacent track near turnarounds were also eliminated. All of this was achieved by simply adding approximately 3.91 ms to the skywriting time at each turnaround. However, this value was determined arbitrarily by subtracting the minimum cross-section dimension of the sample geometry from the programmed stripe width and dividing the results by the scan velocity. The derivation of the critical cooling constant t_{crit} proposed here could provide a quantitative procedure for determining this value, with the goal being to determine the minimum skywriting time that enhances part quality but minimizes the increase in build time due to the added delays at the end of each track.

6.6 References

[1] Shrestha, Subin, and Kevin Chou. "Residual heat effect on the melt pool geometry during the laser powder bed fusion process." *Journal of Manufacturing and Materials Processing* 6 (2022): 153.

[2] Fox J, Evans C., Weaver J, & Redford J. "Surface topography and melt pool behavior in rapid turnaround regions of laser powder bed fusion additive manufacturing of nickel superalloy 625." *CIRP Annals* (2023).

[3] Taylor, Hunter C., and Ryan B. Wicker. "Impacts of microsecond control in laser powder bed fusion processing." *Additive Manufacturing* (2022): 60: 103-239.

[4] Diehl, Brett, and Abdalla Nassar. "Reducing near-surface voids in metal (Ti-6Al-4V) powder bed fusion additive manufacturing: the effect of inter-hatch travel time." *Additive Manufacturing* (2020): 236: 101-592.

[5] Mancisidor, A. M., F. Garciandia, M. San Sebastian, P. Ãlvarez, J. DÃaz, and I. Unanue. "Reduction of the residual porosity in parts manufactured by selective laser melting using skywriting and high focus offset strategies." *Physics Procedia* 83 (2016): 864-873.

[6] Bakir, Yusuf. "Investigation of SLM process Phenomena by means of In Situ Monitoring Systems." (2022).

[7] Yeung, Ho, Brandon Lane, and Jason Fox. "Part geometry and conduction-based laser power control for powder bed fusion additive manufacturing." *Additive manufacturing* (2019).

[8] Yeung, Ho, and Brandon Lane. "A residual heat compensation-based scan strategy for powder bed fusion additive manufacturing." *Manufacturing letters* (2020): 56-59.

[9] Redford J, Fox J, Evans C, Mullany B, Allen A, Morse E. "Center line-time functions and critical constants for predicting laser powder bed fusion melt pool distortion using one surface topography measurement." Proceedings of joint special interest group meeting between euspen and ASPE Advancing Precision in Additive Manufacturing, KU Leuven, Belgium. (2023)

[10] Criales, Luis E., Yiğit M. Arısoy, Brandon Lane, Shawn Moylan, Alkan Donmez, and Tuğrul Özel. "Laser powder bed fusion of nickel alloy 625: Experimental investigations of effects of process parameters on melt pool size and shape with spatter analysis." *International Journal of Machine Tools and Manufacture* (2017): 121: 22-36.

CHAPTER 7: CONCLUSIONS

7.1 Summary and major contributions

The first objective of this dissertation was to develop, implement, and validate a systematic framework for identifying surface descriptions capable of classifying textures based on process or user-defined differences. The second was to quantify the ability of signal detection theory's measure d' to assess texture parameter discrimination abilities across various surfaces. Both of these objectives were addressed in Chapter 3 and Chapter 4. SQuID^{TM} provides a new interpretable systematic framework that automates and can replace the Edisionaion nature of identifying discriminatory features for surface texture classification. The framework provides the first reported use of a d' metric, originating from the theory of signal detection, specifically for binary-multi-class surface texture classification and parameter selection. The capabilities and robustness of the SQuID^{TM} framework for surface classification were demonstrated through two challenging case studies. 1) ISO 25178-2 parameters to classify intensity images of rolled steel defects, and 2) low magnification coherence scanning interferometry measurements to classify ten powder-coated visual smoothness standards. The proposed discrimination matrix, presented in a heatmap format, and automatic parameter selection algorithm based on this matrix, introduces a new down-selection approach. In addition, examining the magnitudes of the d' values that make up the matrix entries enables direct insights into the separability of surface data in a structured and interpretable way.

The final objective of this dissertation was to develop, implement, and assess a surface metrology-based framework to predict anomalous melt pool-surface topography formation in the laser powder bed fusion process. Chapter 5 successfully demonstrates an application of a single process-informed surface measurement to derive a temporal process constant to predict track-to-track overheating during LPBF AM of IN625. Chapter 6 generalized the methodology by introducing a pointwise characterization of the scan strategy informed by the critical cooling constant derived in Chapter 5 to predict overheated regions in different sample geometries. The methodology and results provide a new, powerful application of surface metrology in LPBF manufacturing. Furthermore, in line with this dissertation's overall scope and aims, each paper introduced a systematic methodology for linking quantitative surface topography measurements to user-defined categories or process outcomes.

The main contributions of each article are summarised concisely by chapter; they consist of the key findings and critical observations of the work within the dissertation. The major contributions of this dissertation are outlined below.

Paper 1:

- The use of signal detection theory measures d' effectively identified ISO 25178-2 parameters that enable correct classification of different surface types despite violations of normality and equal variance regarding the data.
- The d' matrix, overlayed heatmap, and metric selection algorithm provided the means to systematically down-select and assess the expected classification outcomes of machine learning models based on the features used to describe the surfaces.
- The ISO 25178-2 parameters, devised for surface height map analysis, effectively characterized intensity-based surface images using significantly fewer metrics than deep learning and image processing approaches with comparable classification accuracy.

Paper 2:

- Bandpass filtering is an effective way to augment a dataset; multiscale ISO 25187-2 texture parameters provided a more accurate classification of visual smoothness standards than metrics obtained on primary unfiltered dataset.
- The visual intuition gained from the d' matrix successfully informed a user of the expected classifier performance before evaluation.

Paper 3 & 4:

- The application of datums to inform surface topography measurements enables precise isolation and characterization of overmelted regions in LPBF parts.
- A single process informed ex-situ surface topography measurement can be used to characterize systemic overmelting in LPBF AM of IN625.
- A point-wise cross-over-time characterization and a critical cooling constant, t_{crit}, provide an interpretable way to locally link scan strategy data to undesirable LPBF process outcomes during fabrication.

7.2 Future Work

7.2.1 Future work based on paper 1,2

SQuID^{TM} provides a solution to the parameter selection problem for multi-class classification; however, a systematic and automated approach for determining optimal scales that enable classification is a logical next step to refine further and enhance the robustness of the proposed framework. A complimentary framework aimed at solving regression problems as opposed to classification would be another logical addition to SQuID^{TM}. Additional parameter development is also an area for future work. In terms of applications, general additive networks (GANs) have emerged as a viable solution to generate synthetic data to offset data imbalances and labor-intensive measurement processes [114]. The SQuID^{TM} framework and ISO parameters could be leveraged to optimize the loss function and ensure the generated synthetic data is quantitative and qualitatively similar to the reference surfaces.

7.2.2 Future work based on paper 3,4

Deriving t_{crit} values for different LPBF machines and alloys is a logical next step for the methodology presented in Chapter 5 and Chapter 6. Formulating an analytical dependence between local thermal and material characteristics, like temperature-dependent diffusivity, could enable t_{crit} to augment parameters like volumetric energy density that are not scan strategy aware. In Chapters 5 and 6, future work should incorporate high-fidelity thermal modeling to complement the experimental results. This would enable the simulation of the process conditions and determination of the relative influence of factors such as different process parameters, material, overhang conditions, dwell time, etc., on t_{crit}. Furthermore, developing a software application that enables users to quickly process and visualize scan strategy data for the entire build using the thresholded cross-over time maps could provide a route to identify defects and flaws that develop over multiple layers and inform corrective control strategies.

REFERENCES

[1] J. Berglund, R. Söderberg, K. Wärmefjord, R. Leach, and E. Morse, "Functional tolerancing of surface texture–a review of existing methods," *Procedia CIRP*, vol. 92, pp. 230–235, 2020.

[2] D. Whitehouse, "Surfacesâa link between manufacture and function," *Proceedings of the Institution of Mechanical Engineers*, vol. 192, no. 1, pp. 179–188, 1978.

[3] K. Tadist, S. Najah, N. S. Nikolov, F. Mrabti, and A. Zahi, "Feature selection methods and genomic big data: a systematic review," *Journal of Big Data*, vol. 6, no. 1, pp. 1–24, 2019.

[4] P. Wang, Y. Yang, and N. S. Moghaddam, "Process modeling in laser powder bed fusion towards defect detection and quality control via machine learning: The state-of-the-art and research challenges," *Journal of Manufacturing Processes*, vol. 73, pp. 961–984, 2022.

[5] B. Mullany, E. Savio, H. Haitjema, and R. Leach, "The implication and evaluation of geometrical imperfections on manufactured surfaces," *CIRP Annals*, vol. 71, no. 2, pp. 717–739, 2022.

[6] J. Redford and B. Mullany, "Construction of a multi-class discrimination matrix and systematic selection of areal texture parameters for quantitative surface and defect classification," *Journal of Manufacturing Systems*, vol. 71, pp. 131–143, 2023.

[7] J. Redford and B. Mullany, "Classification of visual smoothness standards using multi-scale areal texture parameters and low-magnification coherence scanning interferometry," *Materials*, vol. 17, no. 7, p. 1653, 2024.

[8] J. Redford, J. Fox, C. Evans, B. Mullany, A. Allen, and E. Morse, "Centerline-time functions and critical constants for predicting laser powder bed fusion melt pool distortion using one surface topography measurement," Proceedings of Joint Special Interest Group meeting between euspen and ASPE, 2023.

[9] L. Binder, "Der widerstand von kontakten," *Elektrotech. Maschinenbau*, vol. 30, pp. 781–782, 1912.

[10] T. D. Jacobs, L. Pastewka, and G. Editors, "Surface topography as a material parameter," *MRS bulletin*, vol. 47, no. 12, pp. 1205–1210, 2022.

[11] T. Bartkowiak, K. Grochalski, B. Gapiński, and M. Wieczorowski, "Discrimination of surface topographies created by two-stage process by means of multiscale analysis," *Materials*, vol. 14, no. 22, p. 7044, 2021.

[12] ISO, "Geometrical product specifications (GPS) – Surface texture: Areal – Part 2: Terms, definitions and surface texture parameters," *ISO 25178-2*, 2012.

[13] R. Deltombe, K. Kubiak, and M. Bigerelle, "How to select the most relevant 3d roughness parameters of a surface," *Scanning: The Journal of Scanning Microscopies*, vol. 36, no. 1, pp. 150–160, 2014.

[14] D. J. Whitehouse, *Handbook of surface and nanometrology.* Taylor & Francis, 2002.

[15] K. Stout and E. Davis, "Surface topography of cylinder boresâthe relationship between manufacture, characterization and function," *Wear*, vol. 95, no. 2, pp. 111–125, 1984.

[16] P. Pawlus, R. Reizer, and M. Wieczorowski, "Functional importance of surface texture parameters," *Materials*, vol. 14, no. 18, p. 5326, 2021.

[17] P. Pawlus and R. Reizer, "Functional importance of honed cylinder liner surface texture: A review," *Tribology International*, vol. 167, p. 107409, 2022.

[18] A. Arslan, H. Masjuki, M. Kalam, M. Varman, R. Mufti, M. Mosarof, L. Khuong, and M. Quazi, "Surface texture manufacturing techniques and tribological effect of surface texturing on cutting tool performance: a review," *Critical Reviews in Solid State and Materials Sciences*, vol. 41, no. 6, pp. 447–481, 2016.

[19] P. Lu and R. J. Wood, "Tribological performance of surface texturing in mechanical applicationsâa review," *Surface Topography: Metrology and Properties*, vol. 8, no. 4, p. 043001, 2020.

[20] L. Wenbo, Z. Lijun, Z. Cunman, M. Dejian, and H. Pengfei, "The influence of surface topography on friction squeal-a review," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 236, no. 11, pp. 2067–2086, 2022.

[21] K. Chen, X. Yang, Y. Zhang, H. Yang, G. Lv, and Y. Gao, "Research progress of improving surface friction properties by surface texture technology," *The International Journal of Advanced Manufacturing Technology*, vol. 116, pp. 2797–2821, 2021.

[22] S. Chen, X. Liu, H. Luo, J. Yu, F. Chen, Y. Zhang, T. Ma, and X. Huang, "A state-of-the-art review of asphalt pavement surface texture and its measurement techniques," *Journal of Road Engineering*, vol. 2, no. 2, pp. 156–180, 2022.

[23] L. De Chiffre, P. Lonardo, H. Trumpold, D. Lucca, G. Goch, C. Brown, J. Raja, and H. N. Hansen, "Quantitative characterisation of surface texture," *CIRP Annals*, vol. 49, no. 2, pp. 635–652, 2000.

[24] R. J. Pike and T. R. Thomas, "Quantitative characterization of microtopography a bibliography of industrial surface metrology," tech. rep., US Dept. of the Interior, US Geological Survey,, 1998.

[25] A. Nsilani Kouediatouka, Q. Ma, Q. Liu, F. J. Mawignon, F. Rafique, and G. Dong, "Design methodology and application of surface texture: A review," *Coatings*, vol. 12, no. 7, p. 1015, 2022.

[26] N. Mahayuddin, J. A. Wahab, M. Salleh, S. F. Roduan, and H. K. Chen, "Surface texturing method and roughness effect on the substrate performance: A short review," *Jurnal Tribologi*, vol. 27, pp. 8–18, 2020.

[27] N. Naat, Y. Boutar, S. Naïmi, S. Mezlini, and L. F. M. Da Silva, "Effect of surface texture on the mechanical performance of bonded joints: a review," *The Journal of Adhesion*, vol. 99, no. 2, pp. 166–258, 2023.

[28] G. P. Petropoulos, C. N. Pandazaras, and J. P. Davim, "Surface texture characterization and evaluation related to machining," *Surface integrity in machining*, pp. 37–66, 2010.

[29] A. Rosenkranz, H. L. Costa, M. Z. Baykara, and A. Martini, "Synergetic effects of surface texturing and solid lubricants to tailor friction and wear–a review," *Tribology International*, vol. 155, p. 106792, 2021.

[30] K. J. Stout and L. Blunt, *Three dimensional surface topography*. Elsevier, 2000.

[31] W. Zhou, Y. Cao, H. Zhao, Z. Li, P. Feng, and F. Feng, "Fractal analysis on surface topography of thin films: A review," *Fractal and Fractional*, vol. 6, no. 3, p. 135, 2022.

[32] I. Sherrington and E. Smith, "Parameters for characterizing the surface topography of engineering components," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 201, no. 4, pp. 297–306, 1987.

[33] X. Jiang, N. Senin, P. J. Scott, and F. Blateyron, "Feature-based characterisation of surface topography and its application," *Cirp Annals*, vol. 70, no. 2, pp. 681–702, 2021.

[34] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen, "From bow to cnn: Two decades of texture representation for texture classification," *International Journal of Computer Vision*, vol. 127, pp. 74–109, 2019.

[35] P. Cavalin and L. S. Oliveira, "A review of texture classification methods and databases," in *2017 30th SIBGRAPI Conference on graphics, patterns and images tutorials (SIBGRAPI-T)*, pp. 1–8, IEEE, 2017.

[36] M. Oechsle, L. Mescheder, M. Niemeyer, T. Strauss, and A. Geiger, "Texture fields: Learning texture representations in function space," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4531–4540, 2019.

[37] M. D. Riley, "The representation of image texture," 1981.

[38] S. S. Patil, A. Junnarkar, and D. Gore, "Study of texture representation techniques," *Int J Emerg Trends Technol Comput Sci*, vol. 3, no. 3, 2014.

[39] M. Haindl and J. Filip, *Visual texture: Accurate material appearance measurement, representation and modeling.* Springer Science & Business Media, 2013.

[40] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *ELCVIA: electronic letters on computer vision and image analysis*, pp. 1–22, 2008.

[41] L. Newton, *Development of methods for measuring and characterising the surface topography of additively manufactured parts.* PhD thesis, University of Nottingham, 2020.

[42] K. D. Akowua, *Integration of On-Machine Surface Metrology and Machining Process Prediction.* PhD thesis, University of Huddersfield, 2018.

[43] F. Blateyron, "The areal field parameters," in *Characterisation of areal surface texture*, pp. 15–43, Springer, 2013.

[44] L. A. Franco and A. Sinatora, "3d surface parameters (iso 25178-2): Actual meaning of spk and its relationship to vmp," *Precision Engineering*, vol. 40, pp. 106–111, 2015.

[45] L. D. Todhunter, R. Leach, S. D. Lawes, and F. Blateyron, "Industrial survey of iso surface texture parameters," *CIRP Journal of Manufacturing Science and Technology*, vol. 19, pp. 84–92, 2017.

[46] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[47] H. Wechsler, "Texture analysisâa survey," *Signal Processing*, vol. 2, no. 3, pp. 271–282, 1980.

[48] L. Van Gool, P. Dewaele, and A. Oosterlinck, "Texture analysis anno 1983," *Computer vision, graphics, and image processing*, vol. 29, no. 3, pp. 336–357, 1985.

[49] F. M. Vilnrotter, R. Nevatia, and K. E. Price, "Structural analysis of natural textures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 1, pp. 76–89, 1986.

[50] T. R. Reed and J. H. Dubuf, “A review of recent texture segmentation and feature extraction techniques,” *CVGIP: Image understanding*, vol. 57, no. 3, pp. 359–372, 1993.

[51] M. Tuceryan and A. K. Jain, “Texture analysis,” *Handbook of pattern recognition and computer vision*, pp. 235–276, 1993.

[52] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, “Textural features for image classification,” *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[53] B. B. Mandelbrot and B. B. Mandelbrot, *The fractal geometry of nature*, vol. 1. WH freeman New York, 1982.

[54] W. Hartmann and A. Loderer, “Automated extraction and assessment of functional features of areal measured microstructures using a segmentation-based evaluation method,” *Surface Topography: Metrology and Properties*, vol. 2, no. 4, p. 044001, 2014.

[55] L. Wang and W. Li, “Image texture analysis: a review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 509–525, 2010.

[56] T. Thomas, “Characterization of surface roughness,” *Precision Engineering*, vol. 3, no. 2, pp. 97–104, 1981.

[57] C. A. Brown, H. N. Hansen, X. J. Jiang, F. Blateyron, J. Berglund, N. Senin, T. Bartkowiak, B. Dixon, G. Le Goïc, Y. Quinsat, *et al.*, “Multiscale analyses and characterizations of surface topographies,” *CIRP annals*, vol. 67, no. 2, pp. 839–862, 2018.

[58] B.-G. Rosen, L. Eriksson, and M. Bergman, “Kansei, surfaces and perception engineering,” *Surface Topography: Metrology and Properties*, vol. 4, no. 3, p. 033001, 2016.

[59] Y. Liu, J. Li, K. Xu, T. Cheng, D. Zhao, W. Li, Q. Teng, and Q. Wei, “An optimized scanning strategy to mitigate excessive heat accumulation caused by short scanning lines in laser powder bed fusion process,” *Additive Manufacturing*, vol. 60, p. 103256, 2022.

[60] F. Azimi and B. Mullany, “Geometric surface feature detection using statistical based metrics,” *Precision Engineering*, vol. 60, pp. 602–609, 2019.

[61] M. Bigerelle, A. Van Gorp, and A. Iost, “Multiscale roughness analysis in injection-molding process,” *Polymer Engineering & Science*, vol. 48, no. 9, pp. 1725–1736, 2008.

[62] S. Jordan and C. Brown, “Comparing texture characterization parameters on their ability to differentiate ground polyethylene ski bases,” *Wear*, vol. 261, no. 3-4, pp. 398–409, 2006.

[63] D. Chrobak, M. Kołodzieczak, P. Kozlovska, A. Krzemińska, and T. Miller, "Leveraging random forest techniques for enhanced microbiological analysis: a machine learning approach to investigating microbial communities and their interactions," *Scientific Collection «InterConf+»*, no. 32 (151), pp. 386–398, 2023.

[64] J. D. Deng, C. Simmermacher, and S. Cranefield, "A study on feature analysis for musical instrument classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 2, pp. 429–438, 2008.

[65] M. Qadees and A. Hannan, "Cross comparison of covid-19 fake news detection machine learning models," in *2023 17th International Conference on Open Source Systems and Technologies (ICOSST)*, pp. 1–7, IEEE, 2023.

[66] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media, 2009.

[67] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[68] W. Kacalak, D. Lipiński, R. Różański, and G. M. Królczyk, "Assessment of the classification ability of parameters characterizing surface topography formed in manufacturing and operation processes," *Measurement*, vol. 170, p. 108715, 2021.

[69] Á. M. López, K. Marteleur, and A. Finne-Wistrand, "Monitoring and classification of polymeric surface features for enabling the adoption of polypropylene powder bed fusion as a standard tool for bioprocessing equipment production," *Additive Manufacturing*, vol. 72, p. 103632, 2023.

[70] K. Singh, N. Paliwal, and K. Kasamias, "Surface roughness characterization using representative elementary area (rea) analysis," *Scientific Reports*, vol. 14, no. 1, p. 1785, 2024.

[71] J. F. Senge, A. H. Astaraee, P. Dłotko, S. Bagherifard, and W. A. Bosbach, "Extending conventional surface roughness iso parameters using topological data analysis for shot peened surfaces," *Scientific Reports*, vol. 12, no. 1, p. 5538, 2022.

[72] I. Caravella, D. Cortis, L. Di Angelo, and D. Orlandi, "Experimental data collection of surface quality analysis of cucrzr specimens manufactured with slm technology: Analysis of the effects of process parameters," *Materials*, vol. 16, no. 1, p. 98, 2022.

[73] M. Bigerelle, T. Mathia, A. Iost, T. Correvits, and K. Anselme, "An expert system to characterize the surface morphological properties according to their functionalities," in *Journal of Physics: Conference Series*, vol. 311, p. 012010, IOP Publishing, 2011.

[74] E. M. Baker, “Signal detection theory analysis of quality control inspector performance,” *Journal of Quality Technology*, vol. 7, no. 2, pp. 62–71, 1975.

[75] W. P. Tanner Jr and J. A. Swets, “A decision-making theory of visual detection.,” *Psychological review*, vol. 61, no. 6, p. 401, 1954.

[76] L. A. Smith and J. W. Barany, “An elementary model of human performance on paced visual inspection tasks,” *AIIE Transactions*, vol. 2, no. 4, pp. 298–308, 1970.

[77] C. G. Drury, “Inspection of sheet materialsâmodel and data,” *Human Factors*, vol. 17, no. 3, pp. 257–265, 1975.

[78] C. G. Drury and J. Addison, “An industrial study of the effects of feedback and fault density on inspection performance,” *Ergonomics*, vol. 16, no. 2, pp. 159–169, 1973.

[79] J. Sheehan and C. Drury, “The analysis of industrial inspection,” *Applied Ergonomics*, vol. 2, no. 2, pp. 74–78, 1971.

[80] P. Wallack and S. K. Adams, “The utility of signal-detection theory in the analysis of industrial inspector accuracy,” *AIIE Transactions*, vol. 1, no. 1, pp. 33–44, 1969.

[81] G. Smith, “Signal detection theory and industrial inspection,” in *Proceeding Of The Human Factor Society 16th Annual Meeting*, 1972.

[82] J. E. See, “Visual inspection: a review of the literature.,” 2012.

[83] A. Craig, “Human engineering: The control of vigilance,” *Sustained attention in human performance*, pp. 247–291, 1984.

[84] D. M. Green, J. A. Swets, *et al.*, *Signal detection theory and psychophysics*, vol. 1. Wiley New York, 1966.

[85] J. Berez and C. Saldaña, “Laser line profile scanning for powder bed topography measurement,” 2022.

[86] H. Yeung and B. Lane, “A residual heat compensation based scan strategy for powder bed fusion additive manufacturing,” *Manufacturing letters*, vol. 25, pp. 56–59, 2020.

[87] J. C. Fox, C. J. Evans, J. S. Weaver, and J. K. Redford, “Surface topography and melt pool behavior in rapid turnaround regions of laser powder bed fusion additive manufacturing of nickel superalloy 625,” *CIRP Annals*, vol. 72, no. 1, pp. 193–196, 2023.

[88] R. E. Ricker, J. C. Heigel, B. M. Lane, I. Zhirnov, and L. E. Levine, “Topographic measurement of individual laser tracks in alloy 625 bare plates,” *Integrating materials and manufacturing innovation*, vol. 8, pp. 521–536, 2019.

[89] T. Childs, C. Hauser, and M. Badrossamay, "Mapping and modelling single scan track formation in direct metal selective laser melting," *CIRP Annals*, vol. 53, no. 1, pp. 191–194, 2004.

[90] D. Simson and S. K. Subbu, "Effect of process parameters on surface integrity of lpbf ti6al4v," *Procedia CIRP*, vol. 108, pp. 716–721, 2022.

[91] C. Guo, Z. Xu, Y. Zhou, S. Shi, G. Li, H. Lu, Q. Zhu, and R. M. Ward, "Single-track investigation of in738lc superalloy fabricated by laser powder bed fusion: Track morphology, bead characteristics and part quality," *Journal of Materials Processing Technology*, vol. 290, p. 117000, 2021.

[92] S. Shrestha and K. Chou, "Single track scanning experiment in laser powder bed fusion process," *Procedia Manufacturing*, vol. 26, pp. 857–864, 2018.

[93] J. Yin, D. Wang, L. Yang, H. Wei, P. Dong, L. Ke, G. Wang, H. Zhu, and X. Zeng, "Correlation between forming quality and spatter dynamics in laser powder bed fusion," *Additive Manufacturing*, vol. 31, p. 100958, 2020.

[94] S. Zhang, S. Shrestha, and K. Chou, "On mesoscopic surface formation in metal laser powder-bed fusion process," in *TMS 2021 150th Annual Meeting & Exhibition Supplemental Proceedings*, pp. 149–161, Springer, 2021.

[95] Z. Feng, G. Wang, Z. Hao, Y. Wang, H. Tan, W. Fan, M. Dang, S. Zhang, Y. Chen, Y. Peng, *et al.*, "Influence of scale effect on surface morphology in laser powder bed fusion technology," *Virtual and Physical Prototyping*, vol. 19, no. 1, p. e2336157, 2024.

[96] A. Townsend, N. Senin, L. Blunt, R. Leach, and J. Taylor, "Surface texture metrology for metal additive manufacturing: a review," *Precision Engineering*, vol. 46, pp. 34–47, 2016.

[97] J. C. Fox, S. P. Moylan, B. M. Lane, B. Whitenton, *et al.*, "Preliminary study toward surface texture as a process signature in laser powder bed fusion additive manufacturing," in *2016 Summer Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing*, 2016.

[98] J. C. Fox, S. P. Moylan, and B. M. Lane, "Effect of process parameters on the surface roughness of overhanging structures in laser powder bed fusion additive manufacturing," *Procedia Cirp*, vol. 45, pp. 131–134, 2016.

[99] M. Shange, I. Yadroitsava, A. du Plessis, and I. Yadroitsev, "Roughness and near-surface porosity of unsupported overhangs produced by high-speed laser powder bed fusion," *3D Printing and Additive Manufacturing*, vol. 9, no. 4, pp. 288–300, 2022.

[100] J. Rogers, J. Elambasseril, C. Wallbrink, B. Krieg, M. Qian, M. Brandt, and M. Leary, “The impact of surface orientation on surface roughness and fatigue life of laser-based powder bed fusion ti-6al-4v,” *Additive Manufacturing*, p. 104149, 2024.

[101] J. Elambasseril, J. Rogers, C. Wallbrink, D. Munk, M. Leary, and M. Qian, “Laser powder bed fusion additive manufacturing (lpbf-am): the influence of design features and lpbf variables on surface topography and effect on fatigue properties,” *Critical reviews in solid state and materials sciences*, vol. 48, no. 1, pp. 132–168, 2023.

[102] S. Rott, A. Ladewig, K. Friedberger, J. Casper, M. Full, and J. H. Schleifenbaum, “Surface roughness in laser powder bed fusion–interdependency of surface orientation and laser incidence,” *Additive Manufacturing*, vol. 36, p. 101437, 2020.

[103] T. Markovits and L. F. Varga, “Investigating the surface roughness of 3d printed metal parts in case of thin 20 μm build layer thickness,” *Journal of Materials Research*, pp. 1–11, 2023.

[104] A. Charles, M. Bayat, A. Elkaseer, L. Thijs, J. H. Hattel, and S. Scholz, “Elucidation of dross formation in laser powder bed fusion at down-facing surfaces: Phenomenon-oriented multiphysics simulation and experimental validation,” *Additive Manufacturing*, vol. 50, p. 102551, 2022.

[105] B.-Q. Li, Z. Li, P. Bai, B. Liu, and Z. Kuai, “Research on surface roughness of alsi10mg parts fabricated by laser powder bed fusion,” *Metals*, vol. 8, no. 7, p. 524, 2018.

[106] M. Yonehara, C. Kato, T. Ikeshoji, K. Takeshita, and H. Kyogoku, “Correlation between surface texture and internal defects in laser powder-bed fusion additive manufacturing. sci rep uk, 11: 22874,” 2021.

[107] J. Gockel, L. Sheridan, B. Koerper, and B. Whip, “The influence of additive manufacturing processing parameters on surface roughness and fatigue life,” *International Journal of Fatigue*, vol. 124, pp. 380–388, 2019.

[108] A. Thompson, *Surface texture measurement of metal additively manufactured parts by X-ray computed tomography.* PhD thesis, University of Nottingham, 2019.

[109] N. Senin, A. Thompson, and R. Leach, “Feature-based characterisation of signature topography in laser powder bed fusion of metals,” *Measurement Science and Technology*, vol. 29, no. 4, p. 045009, 2018.

[110] T. Özel, A. Altay, B. Kaftanoğlu, R. Leach, N. Senin, and A. Donmez, “Focus variation measurement and prediction of surface texture parameters using machine learning in laser powder bed fusion,” *Journal of manufacturing science and engineering*, vol. 142, no. 1, p. 011008, 2020.

[111] J. Lee, M. S. Hossain, M. Taheri, A. Jameel, M. Lakshmipathy, and H. Taheri, "Characterization of surface topography features for the effect of process parameters and their correlation to quality monitoring in metal additive manufacturing," *Metrology*, vol. 2, no. 1, pp. 73–83, 2022.

[112] J. C. Fox, A. Sood, R. Isaacs, P. Brackman, B. Mullany, E. Morse, A. Allen, E. C. Santos, and C. Evans, "Surface feature characteristics of laser powder bed fusion of nickel super alloy 625 bulk regions," *Procedia CIRP*, vol. 108, pp. 531–536, 2022.

[113] J. Fox, C. Evans, A. Sood, R. Isaacs, B. Mullany, A. Allen, and E. Morse, "Weld track distortion in laser powder bed fusion of nickel superalloy 625," 2022.

[114] J. Eastwood, L. Newton, R. Leach, and S. Piano, "Generation and categorisation of surface texture data using a modified progressively growing adversarial network," *Precision Engineering*, vol. 74, pp. 1–11, 2022.