2017

Automated Design for Manufacturing and Supply Chain Using Geometric Data Mining and Machine Learning

Michael Jeffrey Daniel Hoefer
Iowa State University

Automated design for manufacturing and supply chain using geometric data mining and machine learning

by

Michael Jeffrey Daniel Hoefer

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Industrial and Manufacturing Systems Engineering

Program of Study Committee:
Matthew Frank, Major Professor
Caroline Krejci
Stephen Vardeman

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This work in this thesis would not have been possible without the support of my friends, family, colleagues, and mentors. First, I want to thank my advisor, Dr. Matthew Frank, for his exceptional guidance and mentoring throughout graduate school. Next, I would like to thank my graduate committee members, Dr. Caroline Krejci and Dr. Stephen Vardeman, for their valuable perspective and insights.

I want to thank Dr. Frank Peters, Dr. Michael Dorneich, Dr. John Jackman, and the rest of the IMSE faculty from whom I have learned much about research and engineering. I owe a large thank you to the IMSE staff, as well as the research group in the Rapid Manufacturing and Prototyping Lab for their help and contributions. Finally, I want to thank my parents, Jeff and Diane, siblings, Lea and Jake, and friends who have provided constant encouragement throughout my time in graduate school.

ABSTRACT

This thesis presents an automated method for assessing conceptual designs with respect to manufacturing and supply chain, using geometric data mining and machine learning algorithms. It is important for designers to understand how design decisions will impact downstream manufacturing and sourcing. Many critical decisions are made during conceptual design that impact production cost even before detailed design is finalized; however, the effects of these decisions are not known until later. Design for manufacturing and design for supply chain are methods that provide feedback to the user in a way that enables proactive design changes.

A conceptual design is largely defined by the geometry found in CAD files. In this work, feature-free geometric algorithms were used to extract meaningful manufacturability metrics from 3D models, which were classified as either castings or machined parts. The developed metrics serve as useful attributes for a machine learning model that can help select the manufacturing process of a conceptual design. A classification accuracy of 86% was achieved using a random forest algorithm, which is comparable to other approaches in the literature, while only using geometry as input. The work in this thesis provides methods for using geometry to evaluate a design for manufacturability and supply chain, enabling proactive design decisions early during new product development.

CHAPTER 1: INTRODUCTION

Research Motivation

Design decisions made during new product development significantly impact the downstream manufacturing systems and supply chains and therefore limit the profitability of the manufacturing systems that produce the designs. The complete details of a new product is provided in a Technical Data Package referred to as the TDP (Figure 1), which is defined as "a technical description of an item adequate for supporting an acquisition strategy, production, engineering, and logistics support [1]." New product development starts with conceptual design, where the general part geometry and schema of a design is determined. However, many details in the technical data package, such as quality assurance provisions and geometric dimensioning and tolerancing (GD&T) are still unknown. After conceptual design, detailed design seeks to fill out the TDP, resulting in all the information necessary to bring a design to fruition.

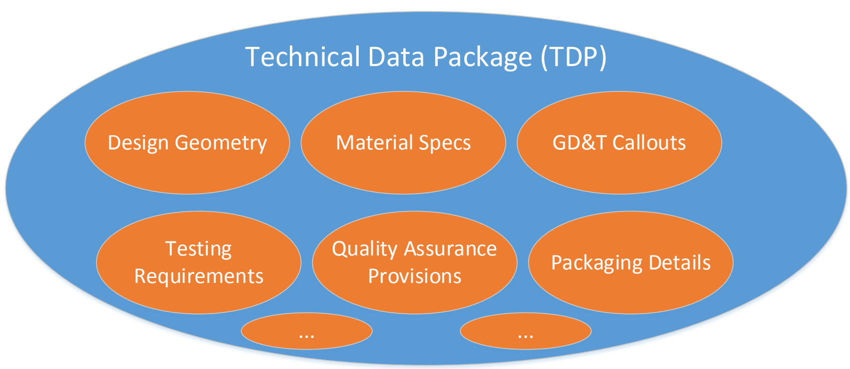

Figure 1. Design information included in the technical data package.

After the TDP is complete, firms will manufacture, distribute, and sell the product (Figure 2). It has been shown that only 20% of the avoidable cost of the product is due to decisions made by production engineering and 30% of the cost is due to detailed design. However, 50% of the avoidable cost in a product is due to design schemes, such as those determined during conceptual design [2]. The ability to predict how a design will impact manufacturing and the supply chain would enable proactive decisions early during new product development; however, it is not clearly understood how early design decisions impact downstream production activities [3].

Figure 2. Product development cycle.

Traditionally, the downstream activity of manufacturing alone was of primary concern. If there was difficulty in manufacturing, engineering change requests would be considered; however, engineering change requests are costly and can disrupt other parts of the manufacturing system. The practice of design for manufacture (DFM) and design for assembly (DFA) arose as a method of measuring the *manufacturability* of a design [4], defined as the ease at which a design can be produced using a given manufacturing process. DFM enabled designers to make proactive decisions to increase manufacturability, which became a consideration in addition to performance (Figure 3). The methods of DFM and DFA can be generalized to design for "X" (DFX), which also includes design for quality, reliability, maintenance, environment, and life cycle cost, to name a few [3].

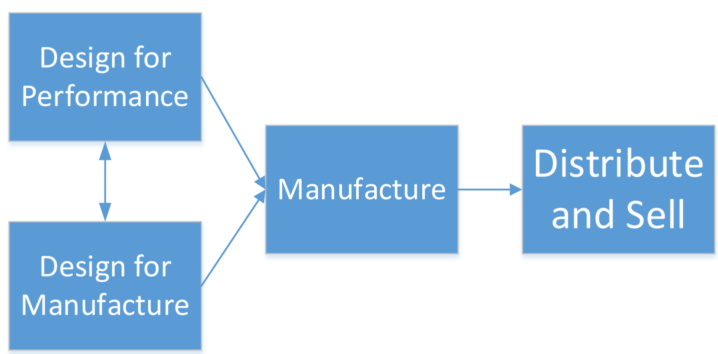

Figure 3. Product lifecycle including design for manufacture.

As product complexity rose, firms started to outsource fabrication to suppliers that specialize in certain manufacturing processes (Figure 4). Instead of manufacturing each individual part, firms source parts through complex global supply chains. Different designs yield different prices, lead times, and quality acceptance rates from suppliers in the supply chain. It is expected that some of these supply chain impacts are a result of the design of the product being sourced [5].

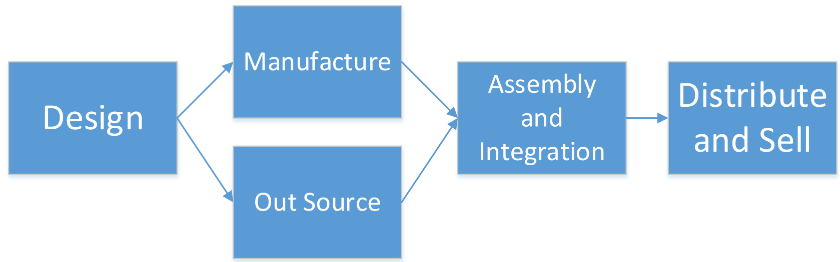

Figure 4. Product development cycle with outsourcing.

Supply chain management is now a more critical downstream activity. Design decisions affect the *sourceability* of a design (Figure 5), which is a general term that is defined as the ease at which a product can be sourced from a given supply chain, with respect to lead time, quality, cost, environmental impact, and more.

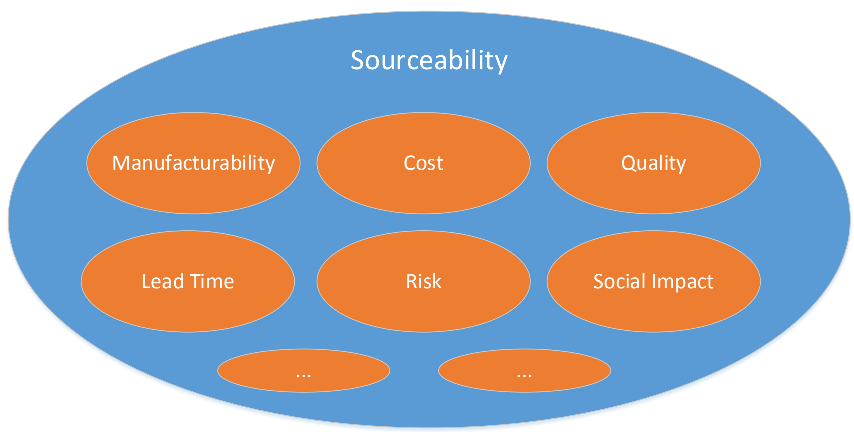

Figure 5. Elements of the sourceability of a design.

Design for supply chain (DFSC) is a relatively new method of measuring the sourceability of a design during product development, and providing feedback that enables proactive design decisions that improve aspects of supply chain management [6]. For example, if a design is identified early as requiring a forging manufacturing process, economic considerations around the forging industry may drive the design towards a different manufacturing process. A small supply base, long lead times, or poor quality records may all drive early conceptual design decisions towards another manufacturing process. Product performance, manufacturability, and supply chain management are all concerns the designer must consider when making decisions (Figure 6). It is important to note that sourceability is not independent from manufacturability. Indeed, the ease at which a part can be manufactured is just as important to the external suppliers as it is to firms that design and fabricate their own parts. However, manufacturability issues may hide in the form of increased prices, longer lead times, and quality defects that are passed on from the supplier to the buyer.

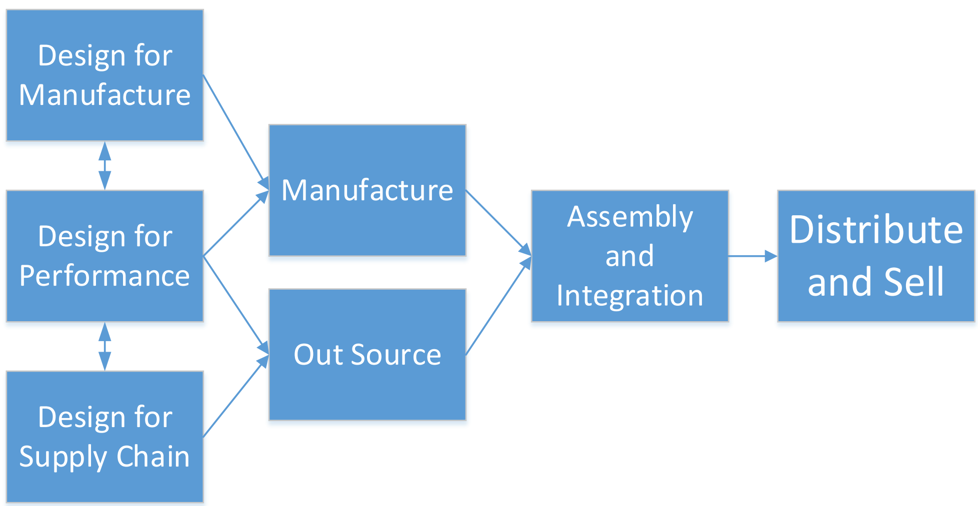

Figure 6. Product development cycle with design for supply chain.

The relationship between design and sourceability is complex, due to the many facets of a supply chain. There are many ways to measure sourceability [7], and there are many ways to characterize a design. Companies have enterprise databases containing information on both the TDP of designs (product data management, PDM) and on supply chain impact (enterprise resource planning, ERP). These databases contain many data points that can be fed to machine learning algorithms to provide designers with a better understanding of how decisions impact sourcing and manufacture of a design, enabling improved DFSC (Figure 7).

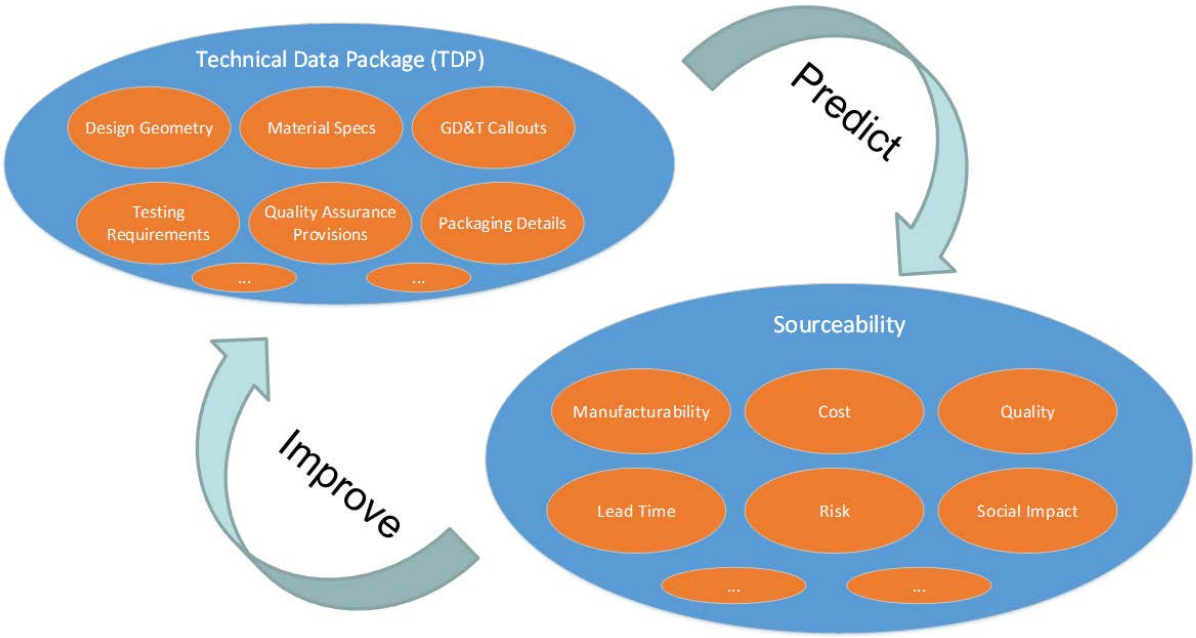

Figure 7. The connection between the TDP of a design and sourceability.

Both ERP and PDM databases contain a wide variety of information used in many business functions. While it is difficult to understand the ease at which a design can be produced through manufacturing or supply chain, data in ERP and PDM databases may provide useful information for designers. The problem is the lack of automated methods that allow designers to evaluate the supply chain and manufacturing impacts early during conceptual design. The objective of this thesis is to develop a data-driven method to automate design for manufacturing and supply chain. To achieve this goal, two sub-objectives need to be addressed. First, quantitative methods of assessing a design will be developed. This includes both measures of the geometry of a design, in addition to measures of supply chain suitability (sourceability). Second, the relationship between design and supply chain will be examined using statistical methods. Machine learning algorithms help provide an understanding of which design metrics have meaningful downstream impacts and serve as tools for evaluating new designs. By completing these objectives, this will lead to automated methods that enable optimal engineering designs with respect to supply chain and manufacturing.

Thesis Organization

Chapter two of this thesis consists of a literature review in the areas of design for manufacturing, design for supply chain, and geometric analysis related to manufacturing. Chapter three consists of a journal article presenting a method for automated manufacturing process selection, written by Michael Hoefer with guidance and revisions from Matthew Frank. Chapter four includes final conclusions and a discussion of future research activities.

REFERENCES

1. “Technical Data Package,” 2016, ACQuipedia, Defense Acquisition University.
2. Corbett, J., and Crookall, P., 1986, “Design for Economic Manufacture,” CIRP Annals - Manufacturing Technology, 35(1), pp. 93–97.
3. Kuo, T., Huang, S., and Zhang, H., 2001, “Design for Manufacture and Design for “X”: concepts, applications, and perspectives,” Computers & Industrial Engineering, 41(3), pp. 241–260.
4. Boothroyd, G., 1994, “Product Design for Manufacture and Assembly,” Computer-Aided Design, 26(7), pp. 505-520.
5. Cargille, B., and Fry, C., 2006, “Design for Supply Chain: Spreading the Word Across HP,” Supply Chain Management Review, 10(5), pp. 34-41.
6. Chiu, M., Gupta, S., and Okudan, G., 2009, “Integration of Supply Chain Decisions at the Conceptual Design Stage: A Repository Enabled Decision Tool,” Proceedings of the 2009 Industrial Engineering Research Conference, pp. 1512–1517.
7. Manohar, K., and Ishii, K., 2009, “Design for Supply Chain: Evaluation of Supply Chain Metrics,” Proceedings of the 2009 ASME International Mechanical Engineering Congress and Exposition, 4, pp. 203–211.

CHAPTER 2: LITERATURE REVIEW

This chapter contains a review of literature related to automated design for manufacture, design for supply chain, and geometric analysis.

Automated Design for Manufacture

Studies have shown that up to 80% of avoidable cost in a production system is due to decisions made during the design stage, and especially during conceptual design [1]. Once design decisions have been finalized, it is costly to retroactively change the design by using engineering change requests (ECR). In addition, ECRs can lead to unintended consequences in different parts of the product, as the decisions made for one part of the design are used as input for design of other parts. The field of design for manufacture (DFM) arose as a method for ensuring designs can be manufactured at a low cost. DFM generally consists of predicting the *manufacturability* of a design, which has been defined as the ease at which a part can be produced using a given manufacturing process. Using DFM feedback, designers seek to make design changes that improve the manufacturability, reducing downstream manufacturing cost and design changes.

There are two documented types of DFM analysis, plan-based and rule-based [2]. Plan-based methods first generate a process plan, and then evaluate the effectiveness of the generated plan. Rule-based methods, on the other hand, use rules to eliminate candidate manufacturing processes. An example of rule-based analysis is the fast-heuristics process filtering approach [3].

Many of the traditional DFM methods have focused on analyzing detailed designs, and tend to require a significant amount of user input. For example, Pro-DFM software uses various criteria and applies a penalty factor to a baseline cost, resulting in an estimate of product cost based on procurement, fabrication, and inventory cost [4]. The required inputs are likely unknown until after the detailed design stage. Similarly, the specific tolerances required by the ProMod software are also intended for detailed design [13]. However, by the time detailed design has begun, much of the general schema of the design has been determined. As shown in [1], the schema of the design can be accountable for up to 50% of the avoidable cost. Therefore, it is important to make decisions that improve the conceptual design before detailed design is finalized.

Performance requirements often command the attention of designers, driving a need to reduce the amount of time and human intervention required for DFM methods. There have been multiple attempts at automating DFM analysis. Many automated methods seek to directly analyze CAD models, without requiring a significant amount of user input. There are two main analysis approaches; feature-based and feature-free. Feature-based approaches seek to identify features from a model and perform analysis on those features, such as a plane or extrusion [5, 6]. While some methods automatically extract features from the CAD file [7, 8], others rely on user input to represent features [9]. Feature-free methods work directly on solid or surface based representations of the features. While features provide useful information, feature-free methods are able to handle any arbitrary geometry without the difficulty of feature identification. Prior feature-free DFM methods tend to focus on a single manufacturing process, such as machining [10 - 12]. The data-driven methods in this thesis are feature-free, and can be used for analysis in a variety of ways depending on the available data.

Design for Supply Chain

Global competition and rising product complexity has encouraged firms to specialize in certain manufacturing processes and core competencies. As a result, many firms outsource fabrication of piece parts to external suppliers. These firms focus on assembly and high level systems integration as their core competency. This necessitated the rise of complex, global supply chains to produce products like aircraft, automobiles, and consumer electronics. While firms that design and fabricate parts only need to consider the manufacturability of a design for production, firms that design and purchase parts need to consider the impact of the design on the supply chain. To this extent, it is important to understand the *sourceability* of a design, which has been defined as the ease at which a firm can procure a quality part in the desired quantity within the desired amount of time at a reasonable price [14]. By understanding how a design impacts downstream supply chain activities, designers can make proactive decisions to reduce cost, shorten lead time, and improve quality. The practice of design for supply chain (DFSC) is concerned with making these decisions to ensure the product is easily sourced.

While DSFC is a relatively new field, multiple companies have implemented DFSC practices and seen significant financial savings. Hewlett-Packard created a six-part DFSC toolkit, involving logistics enhancement, commonality and reuse, and postponed differentiation. Use of DFSC has provided an estimated savings of over $100 million as of 2006 [15]. A firm in the fashion industry also found success in DFSC by utilizing cross functional design teams that communicate across multiple facets of operations, resulting in designs that could be produced at a lower cost [16]. Despite the potential benefits, relatively few DFSC tools have been developed.

Recent methods focus on high level configurations of the bill of materials (BOM) of a product. In that sense, DFSC is applied to the assembly as whole, rather than looking at the geometry of individual piece parts. One tool focuses on design for assembly (DFA) and calculates a DFA index for each possible BOM from a variety of options. The highest scoring BOMs are then evaluated using a supply chain index [17]. This can help designers in selecting which part alternatives to include in an assembly. Another approach focused on the risk in the supply chain [18]. This study involved an industry survey to identify the most important risk factors, and the development of a mixed integer programming model to help select between different design alternatives.

Similar to manufacturability, there are multiple ways to measure the sourceability of a design. Multiple supply chain metrics have been defined in order to measure sourcing and procurement performance. Prior research has focused on metrics such as delivery, cost, inventory, and logistics, aligned with customer satisfaction. These metrics have been grouped as strategic, tactical, or operational [19]. Another study conducted an industry survey that identified lead time, quality, and social and environmental metrics as the most important for design for supply chain [20]. Different companies will benefit from focusing on metrics that are important to their specific product configuration and supply chain. For example, an aerospace firm that requires a highly specialized forging process may be concerned about supplier capacity metrics, to ensure suppliers will be able to meet production demand. On the other hand, a firm that specializes in consumer electronics may be more interested in environmental or social metrics, given the dependence of that industry on customer sentiment.

One important aspect in design for supply chain and design for manufacture is the manufacturing process used to produce the designed part. The geometry of a design will often dictate which manufacturing process can be used to produce the part. For example, a part with internal hollow cavities cannot be easily produced via machining, because the part will have surfaces that are inaccessible to a machine tool. In addition, some manufacturing processes are better suited to certain geometries due to economic or environmental concerns. For example, a part consisting of thin metal sections, such as a simple box, could be creating using machining. However, this would require a significant amount of material to be removed from a solid billet, resulting in costly machine time, tool wear, and material use. The part would likely be produced more effectively as a weldment, by fabricating individual plates and welding the pieces together at the end. For the same part, casting may be entirely infeasible due to the thin sections of the walls.

Understanding which process will be used to produce a part can provide insights into the possible cost, lead time, and quality the part will yield when it is fabricated. For example, parts that are cast generally have a poorer surface finish than those that are machined. Manufacturing process selection is a relatively well developed field that focuses on analyzing geometry and production requirements, among others, to select the most economical process for fabrication [21]. Simple methods of process selection involve picking a process from a grid based on production quantity and desired material. However, this method ignores the geometric constraints inherent to manufacturing processes.

While the geometry of a design can yield useful information, it can be difficult to extract data from the models. Geometric analysis is a field that focuses on collecting useful data from a 3D model. A significant amount of geometric analysis has been used for the purposes of clustering parts for group technology (GT). GT seeks to group similar parts for batch manufacturing, reducing the production cost of each piece. Automated methods have been developed to analyze the features of a STEP file and automatically assign an Optiz GT code for part retrieval and design reuse [22]. In addition, software has been written that can analyze an assembly based on mating geometries of piece parts [23]. Geometric analysis is often performed on surface based or solid models. One example is the use of curvature based measures to classify parts in the National Design Repository, using support vector machines (SVM) and k-nearest-neighbors (KNN) [24]. SVM and KNN are both methods for classification using machine learning. Other machine learning methods have been utilized for geometric data, including the use of learning logic [25]. The methods in this thesis utilize decision trees and random forest for classification of 3D models based on manufacturing constraints.

REFERENCES

1. Corbett, J., and Crookall, P., 1986, “Design for Economic Manufacture,” CIRP Annals - Manufacturing Technology, 35(1), pp. 93–97.

2. Doll, W., Hong, P., and Nahm, A., 2010, “Antecedents and Outcomes of Manufacturability in Integrated Product Development,” International Journal of Operations and Production Management, 30(8), pp. 821–852.

3. Kim, W., and Simpson, T., 2013, “Toward Automated Design for Manufacturing Feedback,” IFIP Advances in Information and Communication Technology, 414, pp. 40–47.

4. Das, S., and Kanchanapiboon, A., 2011, “A Multi-Criteria Model for Evaluating Design for Manufacturability,” International Journal of Production Research 49(4), pp. 1197–1217.

5. Ramana, K., Singh, M., Gupta, A., Dey, R., Kapoor, A., and Rao, P., 2006, “A Manufacturability Advisor for Spun and Rollformed Sheet Metal Components,” International Journal of Advanced Manufacturing Technology, 28(3–4), pp. 249–254.

6. Jacob, D., Ramana, K., and Rao, P., 2004, “Automated Manufacturability Assessment of Rotational Parts by Grinding,” International Journal of Production Research, 42(3), pp. 505–519.

7. Jared, G., Limage, M., Sherrin, I., and Swift, K., 1994, “Geometric Reasoning and Design for Manufacture,” Computer-Aided Design, 26(7), pp. 528–536.

8. Madan, J., Rao, P., and Kundra, T., 2007, “Die-Casting Feature Recognition for Automated Parting Direction and Parting Line Determination,” Journal of Computing and Information Science in Engineering, 7(3), pp. 236.

9. Cutkosky, M., and Tenenbaum, J., 1992, “Toward a Framework for Concurrent Design,” International Journal of Systems Automation: Research and Applications, pp. 239-261.

10. Lee, R., Lin, Y., Tseng, M., and Wu, W., 2010, “Evaluation of Workpiece Orientation and Configuration of Multi-Axis Machine Tool Using Visibility Cone Analysis,” International Journal of Computer Integrated Manufacturing, 23(7), pp. 630–639.

11. Samarghandy, H., and Li, Y., 2013, “Detecting Re-design Area for Increasing Manufacturability of Drilling and Three-Axis Pocketing Operations,” International Journal of Advanced Manufacturing Technology, 69(1-4), pp. 337-349.

12. Barari, A., ElMaraghy, H., and ElMaraghy, W., 2009, “Design for Manufacturing of Sculptured Surfaces: A Computational Platform,” Journal of Computing and Information Science in Engineering, 9(June), 21006.

13. Nnaji, B., Liu, H., and Rembold, U., 1993, “A Product Modeller for Discrete Components,” International Journal of Production Research, 31(9), pp. 2017-2044.

14. Hoefer, M., Frank, M., and Dorneich, M., 2017, “Geometric Analysis to Automate Design for Supply Chain,” In Proceedings of the 2017 Industrial and Systems Engineering Research Conference. In Press.

15. Cargille, B., and Fry, C., 2006, “Design for Supply Chain: Spreading the Word Across HP,” Supply Chain Management Review, 10(5), pp. 34-36-41.

16. Khan, O., Christopher, M., and Creazza, A., 2012, “Aligning Product Design with the Supply Chain: A Case Study,” Supply Chain Management: An International Journal, 17(3), pp. 323–336.

17. Chiu, M., Gupta, S., and Okudan, G., 2009, “Integration of Supply Chain Decisions at the Conceptual Design Stage: A Repository Enabled Decision Tool,” Proceedings of the 2009 Industrial Engineering Research Conference, pp. 1512–1517.

18. Claypool, E., Norman, B., and Needy, K., 2010, “Identifying Important Risk Factors in Design for Supply Chain,” Proceedings of the 2010 Industrial Engineering Research Conference.

19. Gunasekaran, A., Patel, C., and Tirtiroglu, E. (2001). “Performance measures and metrics in a supply chain environment,” International Journal of Operations & Production Management, 21.

20. Manohar, K., and Ishii, K., 2009, "Design for supply chain: Evaluation of supply chain metrics," ASME International Mechanical Engineering Congress and Exposition, Proceedings, 4, pp. 203–211.

21. Swift, K., and Booker, J., 2013, Manufacturing Process Selection Handbook. Waltham, MA: Elsevier.

22. Zehtaban, L., Elazhary, O., and Roller, D., 2016, "A Framework for Similarity Recognition of CAD Models," Journal of Computational Design and Engineering, 3(3), pp. 274–285.

23. Lupinetti, K., Giannini, F., Monti, M., and Pernot, J. P., 2016, "Automatic Extraction of Assembly Component Relationships for Assembly Model Retrieval," Procedia CIRP, 50, pp. 472–477.

24. Ip, C. Y., and Regli, W. C., 2006, "A 3D Object Classifier for Discriminating Manufacturing Processes," Computers and Graphics, 30(6), pp. 903–916.

25. Al-Mubaid, H., Abouel Nasr, E., & Kamrani, A. (2008). Using data mining in the manufacturing systems for CAD model analysis and classification. International Journal of Agile Systems and Management, 3(1/2), pp. 147–162.

CHAPTER 3: AUTOMATED MANUFACTURING PROCESS SELECTION DURING CONCEPTUAL DESIGN

A paper submitted to the ASME Journal of Mechanical Design

Michael J. Hoefer and Matthew C. Frank

Abstract

This paper presents a method for automated manufacturing process selection during conceptual design. It is helpful to know which manufacturing processes can produce a design at an early stage, when the overall design can be changed for less cost. Early during new product development, geometric dimensions and tolerances may not yet be specified, but a general 3D model is often under development. Algorithms are presented to interrogate 3D models to calculate machining based manufacturability metrics. These algorithms are used on a dataset of 86 CAD models classified as machined or cast-then-machined. The metrics, such as visibility, reachability, and setup orientations, seek to characterize a part's manufacturability using machining domain knowledge. These metrics serve as inputs to machine learning models, which are used to classify parts by manufacturing process with 86% accuracy. Some of the incorrectly classified parts were instances that had robust designs capable of being manufactured using machining or casting. The results of the machine learning models indicate that the machining metrics can be used to provide process selection feedback during conceptual design.

1. Introduction

Increasing competition has put pressure on firms to reduce time to market and lower product cost. Understanding which manufacturing process will be used to produce a design is a critical step in the design process. Selecting an appropriate manufacturing process early during conceptual design results in parts that are more manufacturable [1]. Engineers are able to tailor a design towards a specific manufacturing process early on, which reduces manufacturing issues and downstream change requests. Traditionally, process selection has relied on human analysis and wisdom [2]. However, methods that rely on human intuition require prior training and are subject to error. It is necessary to develop systematic and objective methods for selecting a manufacturing process based on only a conceptual design.

Conceptual design is the first stage in new product development, and involves determining the general scheme of the solution [3]. It has been shown that up to 50% of the avoidable cost of a product is determined in the conceptual design stage [4]. Conceptual designs often include CAD drawings [3], but do not contain all the details necessary (technical data package) to produce the design. Conceptual designs are improved in an iterative process consisting of synthesis, analysis, and evaluation. Once a conceptual design is finalized, details are added until the schematics are ready for production. Detailed design then adds final details, such as those resulting from geometric dimensioning and tolerancing (GD&T). During detailed design, the manufacturing process has likely been selected, and designs are tailored for the specific process. Therefore, process selection is critical during conceptual design to avoid detailing a design for an inappropriate process.

This paper focuses on selecting between two common manufacturing processes, casting and machining. Apart from sheet metal forming, casting and machining processes are used to create the vast majority of metal production parts. The machining process has a high dimensional capability and leaves the material properties relatively unchanged [5]. However, since a machine tool must make contact with every surface of the finished geometry, the machining process is relatively inefficient for manufacturing large quantities of parts. The casting process is generally faster than machining (after tooling is created) and can be scaled to achieve production runs of large quantities. While there are a variety of casting processes with different capabilities, most tend to have lower dimensional accuracy and a rougher surface finish compared to machining. This leads to the use of casting to achieve near-net shape geometry for high quantity production runs, and the use of machining on critical features to meet the dimensional specifications [6]. These parts are deemed *cast-then-machined*.

Selecting between pure machining and a cast-then-machined approach involves multiple considerations. Production quantity and material both play a significant role in effective process selection [7]. For example, some materials are better suited for machining, while others are better for casting. Lead time may also be an important factor. Most all cast parts often require custom tooling (patterns) to be created before parts can be produced, whereas machining tends to require less custom fixturing, resulting in a shorter lead time if only one or a few parts are needed. While production requirements need to be considered, the geometry of the design often dictates which process will be most capable of creating the part due to manufacturing constraints. For example, a part with easily accessible flat surfaces

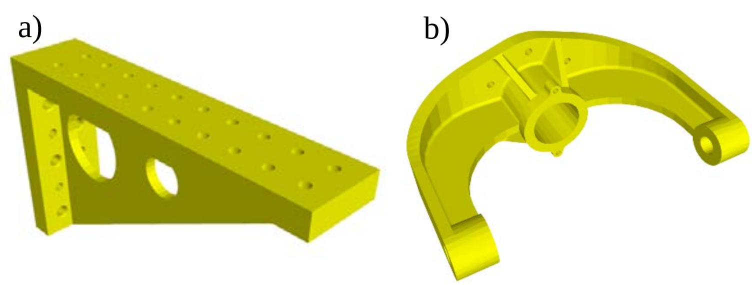

Figure 1. Example parts; a) A part with curved surfaces suitable for casting, b) A part with many flat surfaces, suitable for machining.

(Figure 1a) would be a stronger candidate for the machining process. On the other hand, geometry with many curved surfaces may lend itself for casting (Figure 1b) Casting tends to be better suited for designs that contain curved surfaces or other non-prismatic features.

This paper presents an automated method for assisting in manufacturing process selection between machined and cast-then-machined parts. Process selection using material and production quantity is a relatively well developed field of research. However, using data driven geometric analysis for process selection during conceptual design is an undeveloped research area. In this paper, geometric analysis is used to generate machining-focused manufacturability metrics that serve as useful measures for process selection. After selecting the most useful metrics, machine learning algorithms are used to create predictive models that aid in process selection during conceptual design.

2. Related Work

Simple methods of process selection involve picking a process from a grid based on production quantity and desired material [7]. However, several software-assisted methods have been developed for various aspects of design for manufacturing and process selection [8, 9]. Many involve methods with varying degrees of process planning or production rules [10, 11]. Most process selection efforts involve gathering a significant amount of information about the design, such as surface finish, tolerances, production rate, and time-to-market. The resulting tools rely on user input to provide process suggestions. MAMPS is a process selection support system that allows users to enter information such as part wall thickness, tolerances, and production volume and receive a compatibility score for three manufacturing processes [12]. PROSEL is system that aids in net-shape process selection from user input, and allows the user to select a general part shape complexity level for analysis [13]. A web-based advisory, system, WebMCSS, utilizes a database of process knowledge to provide information to users [14]. A faceted classification system was developed that allows designers to explore aspects of different processes [15]. The Manufacturing Advisory Surface (MAS) is a similar system that allows users to query processes with certain characteristics [16].

Other efforts seek to estimate the manufacturing cost as a basis for process selection [17 - 20]. Machine learning has been used to estimate the manufacturing cost of individual jet engine components, based on a combination of design, materials, and economics [21]. Task-based methods have been used for later stage detailed process selection [22], and specifically for aluminum castings [23]. These methods tend to require a significant amount of manual input from the designer or do not include the analysis of CAD models whatsoever.

Additional utility in process selection can arise from direct analysis of part geometry. Physical parts have been measured for attributes, such as surface roughness, that were used to evaluate process chains involving additive manufacturing [24]. Other efforts focus on automated group technology (GT), which analyzes CAD geometry and finds natural grouping of parts [25]. STEP files can be automatically assigned an Optiz GT code, which involves traversing a decision tree to assign digits of the code [26]. GT is only one application of similarity assessment, which has been used for search, exploration, and retrieval of shapes during design [27, 28]. This has been attempted both for assemblies [29 - 31] and piece parts [32 - 36]. Other efforts focus on clustering CAD models based on features [37], or using hierarchical methods [38].

Most similar to the work of this paper are efforts to classify or evaluate parts in the National Design Repository. These efforts use general shape descriptors [39, 40], invariants [41], or scale-space decomposition [42]. The method in this paper uses domain knowledge of the machining process to generate slice-based and facet-based metrics, which contrasts with prior work that use general descriptors for classification. While previous efforts have used k-nearest neighbor (KNN), support vector machines [6], or learning logic [43], this paper uses decision trees and random forests.

3. Automated Process Selection

3.1 Solution Overview

The method presented in the paper consists of two main efforts. First, each part is characterized using three groups of metrics; aggregate geometry (such as volume and surface area), slice-based machining metrics, and facet-based orientation metrics. Second, these metrics, along with an assigned manufacturing process classification, are used as inputs to machine learning algorithms. The result of the machine learning algorithms is a model that predicts the classification of a new design, assisting in process selection. The process flow is shown in Figure 2. In this paper, the metrics are collected from a dataset of 86 parts from the National Design Repository [44] that are classified as either machined or cast-then-machined. Section 3.2 provides detailed descriptions of how the metrics are calculated, as well as the expected impact on process selection. Section 3.3 presents the machine learning algorithms used to generate the predictive model.

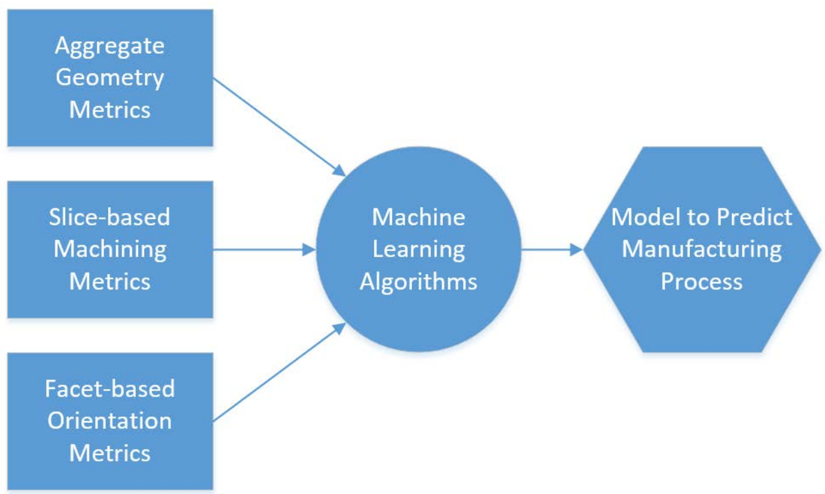

Figure 2. Composition of the model for predicting manufacturing process.

A CAD model is used as input when calculating the metrics in this work. There are two main categories of CAD file types: feature-based and feature-free. Feature-based models consist of discrete features controlled by parameters. Feature-based file formats tend to be proprietary in nature, and analysis is more complex as many types of features must be considered. Feature-free models, on the other hand, are surface-based representations consisting of polygons, such as triangles, or facets. The STL file format is a non-proprietary feature-free format that consists of a facet-based approximation of the surface of the geometry. The metrics presented in this paper are generated from algorithms that operate on STL files, which enables the algorithms to analyze any arbitrary geometry.

3.2 Metrics

The metrics used in this approach can be categorized into three groups. The first group consists of general measures of geometry, such as volume or surface area. The second group of metrics is based on manufacturing constraints of machining, using a slice-based method. The last group includes facet orientation and setup complexity metrics.

Before calculating the metrics, each CAD model was scaled such that the longest dimension along the X, Y, or Z primary axis was equal to 10 inches. This was to ensure that the size of the parts was relatively similar, and that differences in metric values were due to geometry rather than size. The first group of metrics to be discussed include general measures of geometry, such as volume or surface area and is presented in the following section on Aggregate Geometry Metrics.

3.2.1 Aggregate Geometry Metrics

Volume to Surface Area Ratio

The volume to surface area ratio is used as a measure of geometric complexity. A high ratio indicates a solid model with few features, while a low ratio indicates that the surface is complex relative to the volume and may contain thin sections. It is expected that machined parts will have a relatively low volume to surface area ratio, as significant portions of material are likely machined away from a block or cylinder of material.

Bounding Box Volume to Part Volume Ratio (Buy-to-Fly ratio)

The ratio of the volume of the bounding box to the volume of the part is colloquially known as the *buy-to-fly* ratio. This references the aerospace industry, in which a block of material is bought and the part is machined out and flown on an aircraft. The buy-to-fly ratio indicates how much material must be removed from a solid block of metal to create the part. It is expected that machined parts will have a higher buy-to-fly ratio, as there is significant cost associated with removing large volumes of material via machining. Designs with a low buy-to-fly ratio are expected to be classified as cast parts.

Bounding Box Surface Area to Part Surface Area Ratio

Similar to the buy-to-fly ratio, the *surface area ratio* is the ratio between the surface area of the bounding box of the part to the surface area of the part itself. Parts with many complex features will have a large increase in surface area compared to the rectangular prism of material from which the part would be machined.

Ratio of Longest to Shortest Dimension

The ratio between the longest and shortest dimension of a model is an indicator of how oblong the part is. Since cast parts require directional solidification to avoid voids in the final part, it is unlikely that extremely oblong parts will be classified as cast parts.

Facet Count to Surface Area Ratio

The facet count/volume ratio is another proxy for geometric complexity. When most commercially available CAD programs convert a model into an STL file, the parameters include a chordal deviation, which represents the permissible error from the true geometry. Flat surfaces, common in machined parts, can be perfectly represented with a low number of facets. Complex curved geometry, on the other hand, will require many facets to represent the true geometry and stay under the required deviation.

3.2.2 Slice-Based Machining Metrics

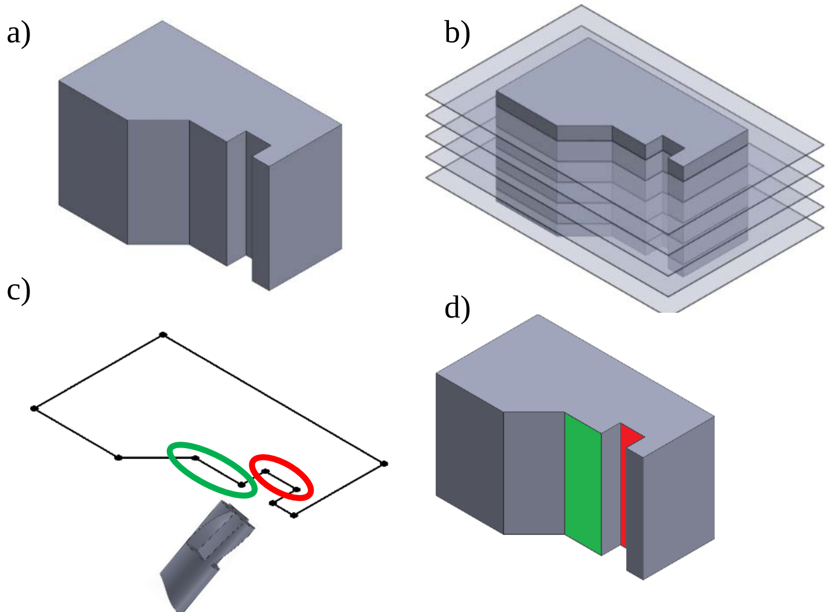

Figure 4. Slice based machining analysis; a) The original STL model, b) 2D slices generated from the model, c) Machining-based manufacturability analysis resulting in numeric results for each segment in a slice, d) Segment values are mapped back to the original surfaces

In processes such as additive manufacturing, process planning is simplified by slicing the 3D geometry of an STL file into a series of 2D slices. A similar method is used in this paper for calculating the visibility, reachability, and tool accessibility metrics. This method is derived from the analysis used in ANA, a system for automated manufacturability analysis [45]. First, a 3D model is sliced along each of the principle axes, resulting in three arrays of consistently spaced 2D slices (Figure 4b). Each slice consists of one or more closed polygonal chains of line segments. Manufacturability analysis is performed on each segment in a chain, resulting in numeric values for each segment (Figure 4c). Lastly, the segment values are mapped back to their original facets, resulting in a numeric score for each facet (Figure 4d). As each facet is assigned a single value based on a series of individual segments, facets with large areas may receive inaccurate scores. Therefore, each part is re-tessellated using the midpoint method of facet subdivision and a maximum facet edge length of 0.5 inches (12.5 mm).

Visibility

For a surface to be machined, a tool must at least have a direct line of sight to the surface in question. The visibility metric measures the range of angles from which a facet is visible with respect to the incident machine tool. If the surface is not in the direct line of sight from any external angle, that surface is not visible, and receives the lowest possible score of zero. To simplify the visibility calculation for each facet in a surface model, the slice based approach (Figure 5) is used to approximate the visibility range for each facet. Visibility for each segment is measured with respect to other segments in the same slice.

The original STL model is used to create an array of slices along each of the principle axes (Figure 5b). The visibility range is calculated for each segment with respect to its own chain using a convex hull visibility method [46].

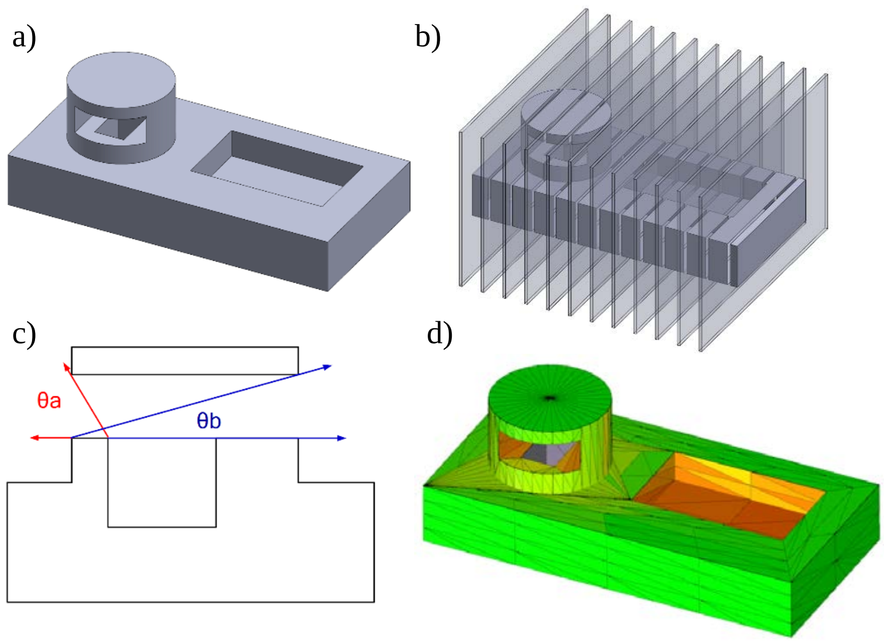

Figure 5. The slice based method for visibility analysis; a) The original STL file, b) The slices generated from one principle axis, c) The visibility calculations for a segment, d) The visibility scores mapped back to the original surface

The visibility range is calculated as the sum of angles from which the segment is visible with respect to the rest of the segments in the slice. In Figure 5c, the visibility score for the single segment is $\theta_a + \theta_b$. This total segment score is mapped back to the original facet (Figure 5d). In addition, the range of angles from which the segment is visible will be used in setup orientation calculations; (0 to θ_b), (θ_a to 180). As multiple segments are generated from a single facet, the worst-case visibility score of the segments is assigned as the visibility score for that facet along that particular axis, and the intersection of visibility ranges of the segments in a facet composes the visibility range for the entire facet. For each facet, the angle ranges from which the facet is visible around a certain axis of rotation are $(\theta_{na}, \theta_{nb})_1$, $(\theta_{na}, \theta_{nb})_2 \ldots$ $(\theta_{na}, \theta_{nb})_i$, where n represents the X, Y, or Z principle axis. The process is repeated for the remaining two principle axes, and the overall visibility score, *Visi*, for the facet is calculated as the sum of the visibility ranges for each principle axis, shown in Equation 1.

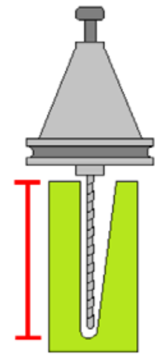

Figure 6. A feature that requires a long tool for machining.

$$Visi = \sum_{1}^{i}(\theta_{xbi} - \theta_{xai}) + \sum_{1}^{i}(\theta_{ybi} - \theta_{yai}) + \sum_{1}^{i}(\theta_{zbi} - \theta_{zai}) \qquad (1)$$

The highest possible visibility value for a segment with respect to a single axis is 180 degrees. Therefore, the highest possible *Visi* score is 540 degrees, which would represent a facet on the convex hull of the 3D model.

While visibility is necessary for machining, models must also have high visibility for metal casting. The casting process involves linear separation of geometry both by removing the mold from the pattern and removing the part from the mold. While some casting processes, such as investment casting, can handle complex internal geometry, it is likely that cast parts will also have high visibility scores; in particular along the parting directions.

Reachability

Machining a surface with a long tool can result in tool deflection and can cause dimensional accuracy issues and poor surface finishes [47]. It is therefore useful to characterize a surface's *reachability*, which represents the length of tool required to machine the surface. The reachability length is defined as the shortest visible distance from the surface to the edge of the part for a given machining angle. The reachability score can take the value of zero and up. A reachability length of zero indicates the facet is on the bounding box of the part. A reachability score of infinity means the facet is not visible from any angle, and therefore is not reachable with any tool length. Parts that contain deep features, such as pockets or tall sections (Figure 6) will have some surface area with poor reachability (long required tool depth).

Calculating reachability is again approximated using 2D slices of a surface model (Figure 7) [48]. The reachability distance is calculated for each segment by measuring the distance from each point on a segment to the line perpendicular to the machining angle that first touches the convex hull of the slice (Figure 7c). The reachability depth R_j for segment j is selected as the longest depth of point R_{ji} from a particular orientation (Equation 2).

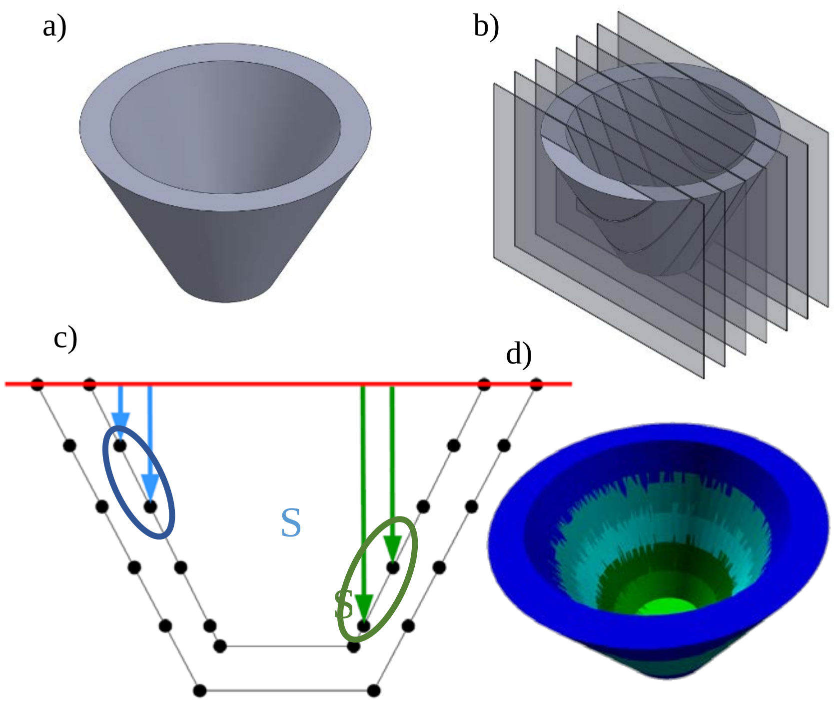

Figure 7. The slice-based method for reachability analysis; a) The original STL file, b) Slices generated from one principle axis, c) Reachability calculation for a single slice and single angle, d) Reachability scores are mapped back to the original facet.

$$R_j = \max_{i \in 1,2} R_{ji} \tag{2}$$

The longest depth across all of a facet's segments is assigned as the reachability depth for that particular facet for a particular angle of approach. The shortest depth across all angles is mapped back to the original model (Figure 7d).

It is expected that machined parts will generally have good (low) required reachability depths to avoid tool deflection. Therefore, parts requiring long machine tools will likely be classified as cast parts.

Tool Accessibility

Parts that contain small features or sharp corners may not be completely accessible by a machine tool without a collision, regardless of visibility or reachability. Tool accessibility takes into account the diameter of the cutting tool when evaluating a surface's machinability. Tool accessibility is approximated using 2D slices of a surface model (Figure 8a and 8b) using the C-Space machinability analysis for 3-axis flat end milling [49]. Within a slice, the machinability of individual points along a segment is analyzed using the concept of tool space (TS) and obstacle space (OS). Tool space is defined as "the aggregate of all feasible cutter locations to cut a point *p* from an orientation α [49]." The obstacle space for obstacle *i* (Ob_i) is the region a tool cannot enter without gouging the obstacle. Obstacles can exist on the same slice as the segment in question (Figure 8c), or they can exist on slices adjacent to the slice containing the segment in question (Figure 9). Obstacles on adjacent slices are considered to be to

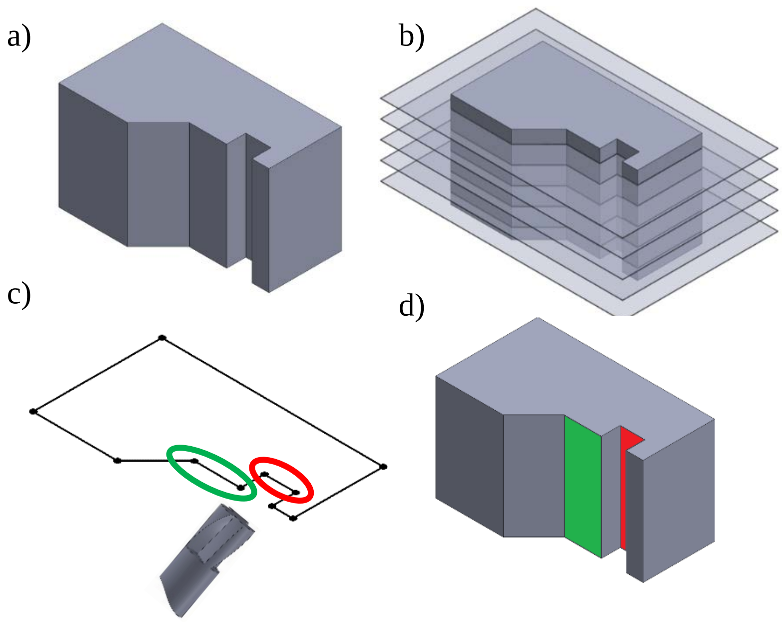

Figure 8. Slice based tool accessibility analysis; a) The original STL model, b) 2D slices generated from the model, c) Tool accessibility analysis on a single slice, d) Segment values are mapped back to the original surfaces

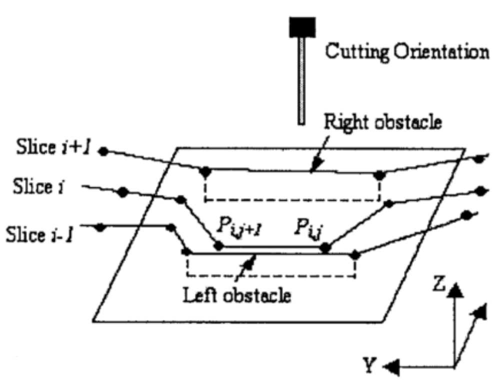

Figure 9.Tool space and obstacle space for a single segment consisting of points P_{ij} and P_{ij+1}. Source: [49]

the "left" (L_m) or the "right" (R_n), when traversing along the polygon chain, of the slice in question. For perpendicular machining (end milling), tool space for a particular orientation, α, is calculated by subtracting the obstacle space (left, right, and same slice) from the maximum possible tool space (MTS), as given in Equation 3.

$$TS = MTS - \sum_{m} OS(L_m, \alpha) - \sum_{n} OS(L_n, \alpha) - OS(i, \alpha) \qquad (3)$$

Tool space is calculated for each segment in each slice, and if the tool space is not empty with other segments, the segment is considered to be accessible from that particular machining orientation (Figure 8c). The accessibility is calculated for multiple setup orientations and a discrete number of tool diameters ranging from .125 inches up to 1 inch, in increments of .125. The worst case diameter across all segments in a facet is mapped back to the original surface for each machining orientation (Figure 8d). Finally, the largest tool diameter across all angles of approach is chosen as the "tool accessibility" metric for each facet.

While difficult-to-access features may be a challenge for casting processes, it is likely that parts with poor accessibility will not be classified as machined. While small holes will have low values for tool accessibility, the surfaces that comprise the holes will likely be a small percentage of the surface area of the model, resulting in a relatively low impact on the weighted metrics for tool accessibility.

Tool Length to Diameter Ratio

The ratio between the tool length and diameter has been shown to have a significant impact on surface roughness of the part [47]. The reachability depth metric serves as a surrogate for tool length, and the tool accessibility diameter metric serves as a surrogate for tool diameter. Therefore, a feasible *tool length to diameter ratio* is calculated for each facet by dividing the reachability depth by the tool diameter. For this study, the longest tool length is 10 inches and the smallest tool diameter is 0.125 inches, meaning the largest possible value for length to diameter ratio is 80.

Number of Axes and Number of Rotations

A significant cost factor in machining is the number of physical setups and orientations that are required to machine a part (Figure 10). In general, the goal is to limit the number of setups to as few as possible. With the increase of four and five axis machining, parts with complex setup requirements may not need to be manually re-aligned, but there is an increased burden on the CNC programmer to avoid tool collisions. Using the visibility ranges calculated for each facet in the *Visibility* section, a greedy heuristic algorithm based on surface area is used to solve the set covering problem to estimate the minimum number of setups required to machine the entire surface of the part [46].

Figure 10. The main block model which requires many setups to machine every facet.

For each facet, the angle ranges from which the facet is visible around a certain axis of rotation are $(\theta_{ca},\theta_{cb})_1$, $(\theta_{ca},\theta_{cb})_2 \ldots (\theta_{ca},\theta_{cb})_i$, where c represents either the X, Y, or Z principle axis. The array of visibility ranges for each facet (Figure 11a) is covered by the array of angles $(\theta_{c1},\theta_{c2},\ldots,\theta_{ck})$ from each axis ($Axis_c$) of rotation (Figure 11b), such that every facet is visible from at least one angle selected in the axis and angle array.

a)

$$\begin{array}{c} Facet_1 \\ Facet_2 \\ \ldots \\ Facet_n \end{array} \begin{bmatrix} \left[(\theta_{xa},\theta_{xb})_1 \ldots (\theta_{xa},\theta_{xb})_i\right] & \left[(\theta_{ya},\theta_{yb})_1 \ldots (\theta_{ya},\theta_{yb})_i\right] & \left[(\theta_{za},\theta_{zb})_1 \ldots (\theta_{za},\theta_{zb})_i\right] \\ \left[(\theta_{xa},\theta_{xb})_1 \ldots (\theta_{xa},\theta_{xb})_i\right] & \left[(\theta_{ya},\theta_{yb})_1 \ldots (\theta_{ya},\theta_{yb})_i\right] & \left[(\theta_{za},\theta_{zb})_1 \ldots (\theta_{za},\theta_{zb})_i\right] \\ \ldots & \ldots & \ldots \\ \left[(\theta_{xa},\theta_{xb})_1 \ldots (\theta_{xa},\theta_{xb})_i\right] & \left[(\theta_{ya},\theta_{yb})_1 \ldots (\theta_{ya},\theta_{yb})_i\right] & \left[(\theta_{za},\theta_{zb})_1 \ldots (\theta_{za},\theta_{zb})_i\right] \end{bmatrix}$$

b)

$$\begin{array}{c} Axis_x \\ Axis_y \\ Axis_z \end{array} \begin{bmatrix} \theta_{x1},\theta_{x2},\ldots,\theta_{xk} \\ \theta_{y1},\theta_{y2},\ldots,\theta_{yk} \\ \theta_{z1},\theta_{z2},\ldots,\theta_{zk} \end{bmatrix}$$

Figure 11. Visibility set cover problem; a) The array of n facets containing the visible angles for each axis of rotation. b) The completed set cover of selected axes and angles.

The number of axes required and number of rotations (angles) for each axis are captured as metrics for the model. Additional required visibility orientations can be costly for both machined and cast parts, as cast parts require directional separation of the part from the mold. However, the chosen angles for casting may not align with the three principle axes, given the variety of curved surfaces and complex features.

3.2.3 Facet-Based Orientation Metrics

Angle Between Facet and Machine Tool

Previous calculations assign each facet to an axis of rotation, and an angle from that axis that results in the highest scoring tool accessibility (largest tool diameter). Each facet on a surface has a unit normal vector, which is perpendicular to the facet and faces away from the solid model (Figure 12). The *tool accessibility orientation angle* is calculated for each facet. A preferred facet orientation for machining would allow for either end milling or face milling. Face milling would require the angle between the facet normal and the machine tool to be zero, while end milling requires an angle of 90 degrees. Deviations from zero or 90 may require ball milling to shape the surface in traditional three axis milling, resulting in additional cost. For that reason, the angles are transformed into *Angle*$_t$ (Equation 4) so that deviations from 0 or 90 degrees are penalized;

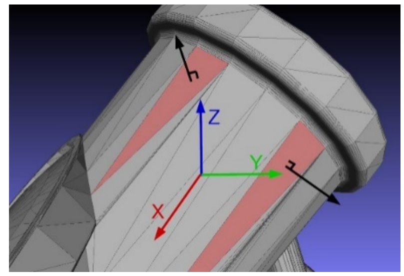

Figure 12. A tessellated model indicating the unit normal vectors.

$$Angle_t = |45 - Angle| \tag{4}$$

where *Angle*$_t$ is the transformed angle ranging from zero to 45 degrees. A value of 45 indicates the facet is aligned with the machine tool such that end milling or face milling is possible. It is therefore expected that machined parts will have more facets with angles closer to 45 degrees, as opposed to cast parts, which are more likely to have curved surfaces that would require ball milling. In addition, the deviation is likely larger for cast parts, given how curved surfaces have a wide degree of variability in facet orientation. Machined parts often consist of flat planar surfaces, which will lower the standard deviation for machined parts.

Deviation Angle

As many features of machined parts are aligned with the principle Cartesian axes, the deviation of the facet normal from the axes is another useful measure. Surfaces ideal for machining will have an angle of 0, 90, or 180 degrees with respect to one of the principle axes. Values of 0 or 180 would indicate the surface is perpendicular to common machine tool setups, leading towards face or slab milling. A value of 90 degrees indicates the surface is parallel to common machine tool setups, which is preferred for end milling. Deviations from these three angles indicate the facet would require costly ball milling from standard machining orientations. Similar to the angle between facet and machine tool, the *deviation angle* is normalized to the range of (0, 45) degrees using Equation 5, and the maximum of the three axes is selected as the deviation angle for the facet.

$$DeviationAngle = \max_{n \in x,y,z} \left| \left| A_{fn} - 90 \right| - 45 \right| \tag{5}$$

In Equation 5 A_{fn} is the angle between the facet normal and the n principle axis, n being X, Y, or Z. The deviation angle metric helps to characterize the facet's orientation with respect to standard machining orientations, which will likely help discriminate between cast and machined models.

3.2.4 Metrics Overview

Table 1 provides an overview of the metrics presented in this section. Metrics with a per-facet frequency will be calculated as the surface area weighted mean, standard deviation, and quantiles.

Table 1. Metrics calculated for each model.

Type	Metric	Units	Range	Frequency
Aggregate Geometry Metrics	Volume to Surface Area	Inches	0 – infinity	Per model
	Buy-to-Fly Ratio	Unit-less	1 – infinity	Per model
	Surface Area Ratio	Unit-less	1 – infinity	Per Model
	Side length ratio (longest/shortest)	Unit-less	0 - 1	Per model
	Facet Count to Surface Area	Facets/Square Inch	0 - infinity	Per model
Slice-Based Machining Metrics	Visibility Score	Degrees	0-540	Per facet
	Reachability Depth	Inches	0-Infinity	Per facet
	Maximum Tool Diameter	Inches	0-1	Per facet
	Tool Length/Diameter	Unit-less	1-80	Per facet
	Required Number Axes	Count	1-Infinity	Per model
	Required Number Rotations	Count	2-Infinity	Per model
Facet-Based Orientation Metrics	Tool Accessibility Orientation Angle	Degrees	0 to 45	Per facet
	Deviation Angle	Degrees	0 to 45	Per facet

Algorithms implemented in C++ were used to analyze the geometry of the 86 models classified by manufacturing process (49 machined, 37 cast) in the National Design Repository.

R scripts were used for statistical analysis. The per-facet metrics result in a distribution of scores for each model. These distributions are summarized using the weighted mean, variance, and 0^{th} (minimum), 25^{th}, 50^{th} (median), 75^{th}, and 100^{th} (maximum) percentiles. In calculating these summary statistics, each facet's value is weighted by its surface area to accommodate variation in facet size. Prior work has shown a statistical difference in many of these metrics between the machined and cast group, using an unpaired t-test [50].

3.3 Machine Learning for Process Selection

Once the machining and geometry metrics are compiled for each model, they are used as inputs to multiple machine learning algorithms. Previous work in classifying parts by manufacturing process have used the k-nearest-neighbor (KNN) and support vector machines (SVM) algorithms for classification [6]. This study also uses KNN, but investigates decision trees and random forests in predicting manufacturing process.

Estimated accuracies are provided for each machine learning method, measured by splitting the dataset into a training group and testing group, or in the case of random forest, using the out-of-bag estimation error. An analysis of a decision tree is provided to determine if the branching decisions are congruous with real manufacturing constraints. Models that are incorrectly analyzed were visually inspected to gain potential insights. The following sub-sections detail the motivation for using each machine learning algorithm.

3.3.1. K-Nearest Neighbor

The k-nearest neighbor (KNN) classification method is based off the KNN clustering method. To predict the classification of a new model, the KNN classification algorithm determines the similarity of the new model to all the existing models. The K most similar models are deemed the "neighbors" of the new model, and the most common classification of the neighbors is selected for the prediction of the new model. The attributes were standardized to lessen the effect of attributes with large values or skewed distributions. A random 20% sample was set aside as the test set, and the remaining 80% served as the “neighbors.” The “class” R package was used for KNN classification [51]. The KNN method was chosen because cast and machined parts may tend to be designed similarly, and using measures of similarity to other parts will likely result in effective classification.

3.3.3. Decision Trees

Decision trees are a collection of hierarchical Boolean decision nodes that form a tree for predicting the classification of new instances. Each node contains an attribute and a value with which the data is “split” by. The root node attribute is selected for the ability to the best ability to split the dataset. The “rpart” R package was used for decision tree classification, which evaluates a split based on the altered priors method [52]. An independent accuracy estimation of an individual decision tree requires a split between the training and test dataset. A random sampling of 20% of the data points were set aside for the accuracy evaluation. Leaf nodes are removed (pruned) to avoid overfitting the tree to the training set. Decision trees are transparent and can be understood by looking at the nodes in the tree, and may provide insight into how parts are classified. In addition, the hierarchical classification process used by decision trees is similar to the process used when assigning group technology classifications [26], and may be suitable for mimicking how a human would perform process selection.

3.3.4. Random Forest

The random forest is an ensemble method involves the creation of many decision trees (a forest). The randomForest package [53] was used to generate the random forest model, based on Breiman's implementation [54]. Each tree is constructed using a random sampling (with replacement) of the available instances. A random subset (size mtry, set to four) of attributes is evaluated for each split in the tree based on gini impurity. The number of trees grown (ntrees) was set to 2000. Once the forest is constructed, new models are run through each tree in the forest and the most commonly predicted category is selected for the model. To estimate the accuracy of the random forest method, the accuracy of each tree is evaluated for the instances that were not used in generating that specific tree; this is considered the out-of-bag error. Random forests provide an importance ranking of the attributes based on the decrease in accuracy when each specific attribute is randomly permuted. One motivation for using random forest is its robustness with respect to correlated variables. The machining based metrics are not completely independent, as surfaces that are easy to machine will score well for visibility, reachability, and tool accessibility. In addition, individual trees in the forest will serve as "experts" for a subset of parts and attributes, simulating a group of manufacturing engineers with different expertise voting on which manufacturing process to use.

4. Results and Discussion

The accuracy of each machine learning algorithm is shown in Table 2.

Table 2. Model accuracies.

Algorithm	Accuracy
KNN	55%
Decision Tree	68%
Random Forest	86%

As shown in Table 2, the overall highest accuracy was the random forest method using an ensemble of 2000 trees. The KNN classifier did not achieve accuracy much greater than 50%, which would be the expected accuracy of a random classifier. This is congruent with previous attempts of classifying this dataset using KNN with curvature descriptors [6]. A single decision tree achieved an accuracy of 68%, but this number varied significantly depending on the training and test data split. The random forest method, which uses an ensemble of decision trees, created a model with an expected accuracy of 86%. The ten most important variables in the model are shown in Figure 13. The importance of each variable was calculated by evaluating the decrease in out-of-bag accuracy when that particular variable was randomly permuted during prediction.

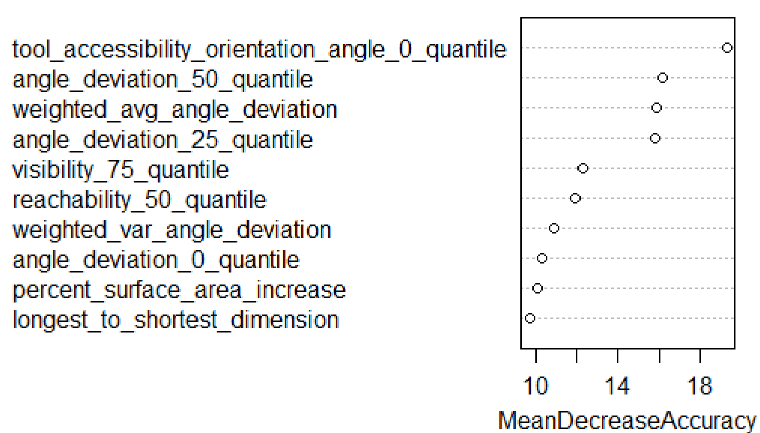

Figure 13. Variable importance plot for the random forest classification model.

The most important variable was the minimum tool accessibility orientation angle of a part. The distribution, plotted by manufacturing process, is shown in Figure 14. The range in tool accessibility orientation angle is from 0 – 45 degrees, with a lower value indicating an unusual machining orientation allows for the largest tool diameter. Parts classified as castings appear to have a lower worst-case machining angle than machined parts. This may be due to complex curved features that have a non-standard machining angle. Machined parts, on the other hand, generally do not have surfaces with extremely low scoring machining orientations.

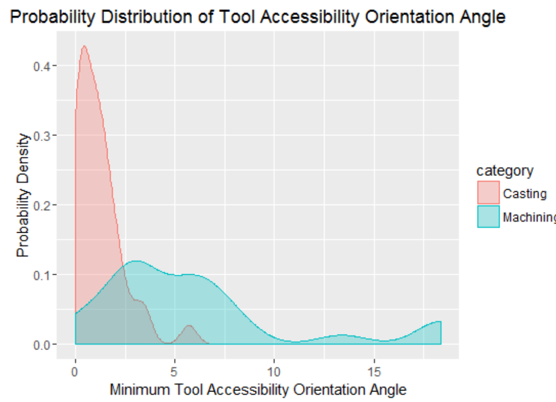

Figure 14. Probability distribution for minimum tool accessibility orientation angle.

The three variables with the next highest importance measures are all derived from the Angle Deviation metric. Higher values for these metrics indicate surfaces that are aligned with traditional orthogonal machining setup orientations. The probability distribution between machined and cast parts is noticeably different (Figure 15). This can be interpreted to suggest that many machined parts have over half of their surface area directly aligned with one of the three principle axes. This is congruent with the idea that machined parts are designed using right angles with respect to the Cartesian coordinate system.

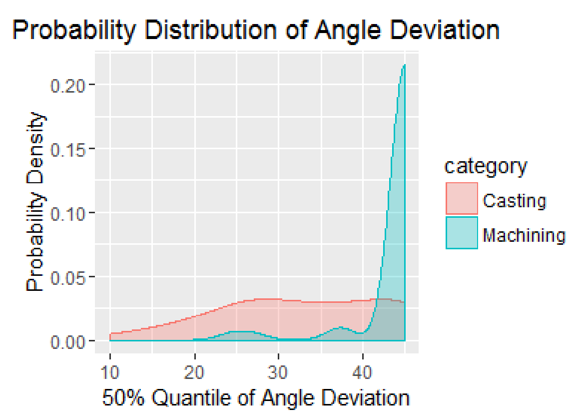

Figure 15. Probability distribution for median angle deviation.

The 75% quantile of visibility was the fifth most important variable. As seen in the histogram (Figure 16), a large percentage of machining models have a 75th visibility percentile at the maximum value of 540 degrees. A value of 540 for the 75th percentile means that at least 25% of the surface area of the part is on the 3D convex hull of the part. Figure 17a shows an example machined part with a large amount of surface area having a visibility score of 540 (completely shaded green). Figure 17b, on the other hand, shows a cast part where much of the surface area scores lower than 540, shaded from yellow to red based on visibility score. Machined parts tend to have large flat surfaces that serve as datums and aid in fixturing, which results in a considerable portion of the surface area having "desirable" visibility.

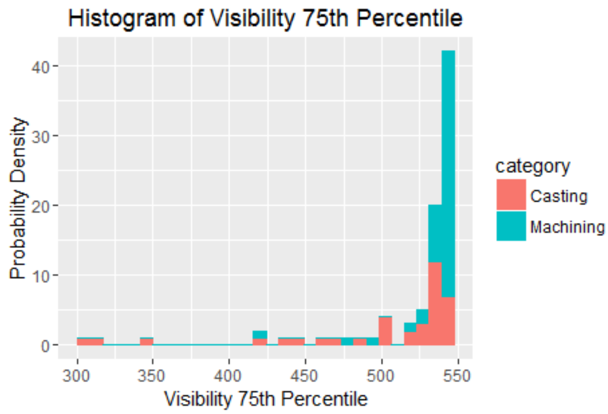

Figure 16. Histogram for 75th percentile of visibility.

a) b)

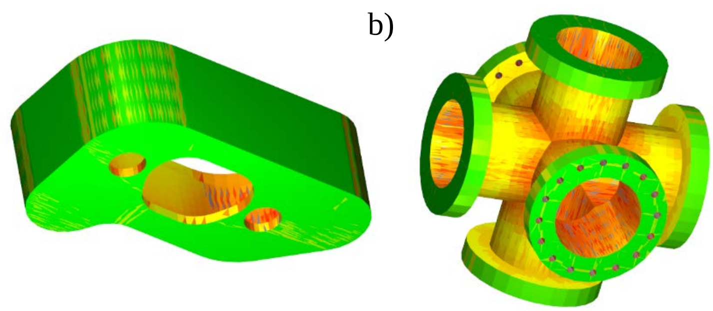

Figure 17. Visibility map where highly visible surfaces are shaded green and less visible surfaces are shaded red; a) a machined part ("part 10"), b) a cast part ("cross").

The median reachability depth was another important predictor in the random forest classifier, likely for the same reasons that the 75th percentile visibility was an important metric. Surfaces that have an ideal visibility score of 540, by definition, are on the three dimensional convex hull of the part, which means those surfaces must also have an ideal reachability depth of zero inches. The distribution of median reachability depth (Figure 18) indicates that most machined parts have a significant amount of their surface area with a reachability depth of zero inches. In summary, the attributes driving the accuracy of the random forest model appear to be associated with the flat planar surfaces commonly found in machined parts.

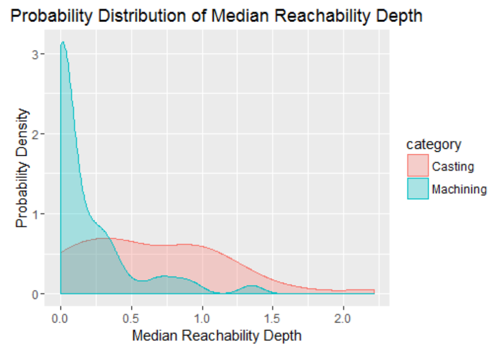

Figure 18. Probability distribution of median reachability depth

5. Conclusions and Future Work

While the accuracy of the random forest method was comparable to similarly published classifiers, there were a handful of misclassified models as measured using the out-of-bag predictions. A few of the casting models were classified as machined models. Glass 1 (Figure 19a) and Glass 2 (Figure 19b) were two cast-then-machined models that were incorrectly classified as machined parts by the random forest model. It is apparent that these parts have significant flat planar surfaces found in many of the

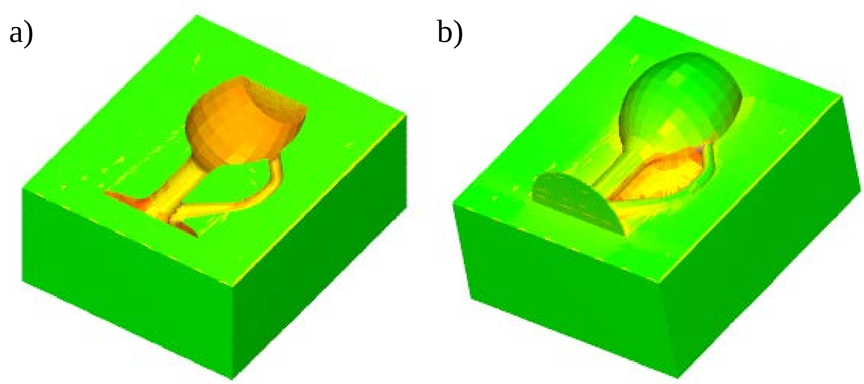

Figure 19. Example casting parts misclassified as machined parts; a) Glass-1, b) Glass-2.

machined parts, which resulted in a high scoring visibility, reachability, tool accessibility, and orientation metrics. These metrics likely "tricked" a majority of decision trees into believing the parts were indeed machined. The parts appear to be a mold and/or pattern for casting a goblet. It is unlikely that the mold/pattern itself would be cast. This also brings into question the integrity of the original dataset. Publications presenting the dataset do not thoroughly explain the process of how the manufacturing classifications were assigned, and in future work, an expert evaluation may be necessary to validate the assigned classifications. A potential improvement to the dataset would be to isolate the geometry of the cup, which would be a suitable candidate for casting.

Some machined parts were misclassified as castings. Assembly Five (Figure 20) consists of a significant amount of curved surfaces that resulted in lower facet orientation scores, which resulted in the confusion by the classifier. MyCami2 (Figure 21), on the other hand, was composed of many flat surfaces. However, the 45 degree angle in the part resulted in poor facet orientations with respect to standard orthogonal setup orienations, contributing towards being misclassified as a casting. In both of these cases, it is possible that both machining and casting would be a suitable near-net shape process to create the design. The curvature of the parts would result in directional solidification necessary for casting, and the flat geometry would also be suitable

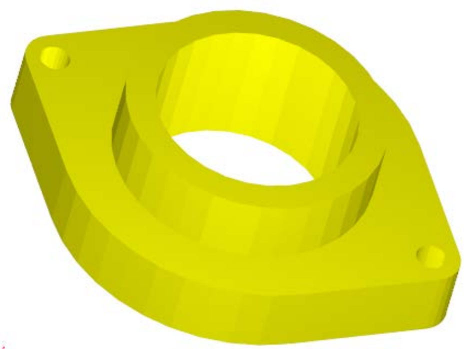

Figure 20. Assembly Five, a machined part misclassified as a casting.

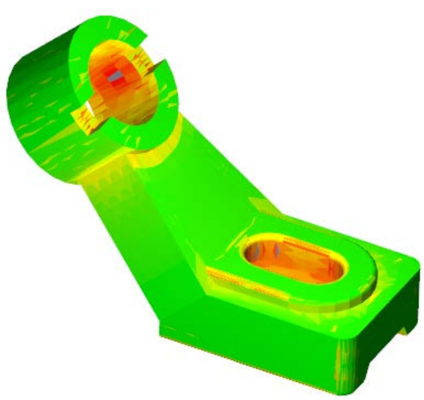

Figure 21. The MyCami2 machined part misclassified as a casting.

for machining. A third “either cast or machined” classification would aid in identifying robust designs that can be manufacturing using either process. In addition, future efforts could work towards providing an overall measure of manufacturability with respect to manufacturing process, rather than the simple binary classification used in this paper. Parts with robust designs will possibly score well for both machining and casting processes, and the decision to choose casting or machining would be a result of production requirements, rather than geometry.

Future work will involve using these methods on expanded datasets that include more production information, beyond the manufacturing process. For example, relating the manufacturability metrics to cost or lead time would provide designers useful feedback early in conceptual development. This work focused on process selection using conceptual design geometry. As prior methods have noted the important of production quantity and material, it is likely that integrating the geometry of the conceptual design with these production requirements will provide improved assistance in process selection.

The work presented in this paper indicates that slice-based and facet-based metrics built from machining domain knowledge can serve as useful predictors for process selection for CAD models creating during conceptual design. A variety of metrics were presented in three categories: aggregate geometry, slice-based machining metrics, and facet-based orientation metrics. Multiple classification algorithms were used to train a predictive model, including k-nearest neighbors, decision trees, and random forest. Using the random forest algorithm, an out-of-bag accuracy of 86% was achieved. The most important geometric indicators measured by the random forest were measures of facet orientation both with respect to a machine tool, and to the principle axes. This is the first known method to use a collection of manufacturing based metrics and machine learning to automatically classify a part by process. The use of these metrics and methods will assist in process selection during conceptual design, without requiring significant user input or expert knowledge.

REFERENCES

1. Esawi, A. and Ashby, M., 2000, "The development and use of a software tool for selecting manufacturing processes at the early stages of design," Journal of Integrated Design & Process Science, 4(2), pp. 27–43.

2. Lovatt, A. and Shercliff, H., 1998, "Manufacturing process selection in engineering design. Part 1: the role of process selection," Materials & Design, 19(5–6), pp. 205–215.

3. Pahl, G., Beitz, W., Feldhusen, J., and Grote, K., 2007, *Engineering design: a systematic approach*. Springer-Verlag London.

4. Corbett, J., and Crookall, P.R., 1986, "Design for Economic Manufacture," CIRP Annals - Manufacturing Technology, 35(1), pp. 93–97.

5. Wetzel, S., 2014, "When to Cast, When to Machine." Metal Casting Design and Purchasing. Sept/Oct.

6. Ip, C., and Regli, W., 2006, "A 3D object classifier for discriminating manufacturing processes," Computers and Graphics (Pergamon), 30(6), pp. 903–916.

7. Swift, K., and Booker, J., 2013, *Manufacturing Process Selection Handbook*. Waltham, MA: Elsevier.

8. Gupta, S., Regli, W., Das, D., and Nau, D., 1997, "Automated Manufacturability Analysis: A Survey," Research in Engineering Design, 9, pp. 168–190.

9. Esawi, A., and Ashby, M., 2000, "The development and use of a software tool for selecting manufacturing processes at the early stages of design," Journal of Integrated Design & Process Science, 4(2), pp. 27–43.

10. Hummel, K., 1989, "Coupling rule-based and object-oriented programming for the classification of machined features," In Proceedings of the ASME Computers in Engineering Conference pp. 409–418.

11. Nau, D., 1987, "Automated process planning using hierarchical abstraction," TI Technical Journal, Winter, pp. 39–46.

12. Giachetti, R., 1998, "A decision support system for material and manufacturing process selection," Journal of Intelligent Manufacturing, 9, pp. 265–276.

13. Yurdakul, M., Arslan, E., Ic, Y., and Tuerkbas, O., 2014, "A decision support system for selection of net-shape primary manufacturing process," International Journal of Production Research, 52(5), pp. 1528–1541.

14. Zha, X., 2005, "A web-based advisory system for process and material selection in concurrent product design for a manufacturing environment," International Journal of Advanced Manufacturing Technology, 25(3–4), pp. 233–243.

15. Giess, M., McMahon, C., Booker, J., and Stewart, D., 2009, "Application of faceted classification in the support of manufacturing process selection," Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 223(6), pp. 597–608.

16. Smith, C., Wright, P., and Séquin, C., 2003, "The Manufacturing Advisory Service: Web-based process and material selection," International Journal of Computer Integrated Manufacturing, 16(6), pp. 373–381.

17. Djassemi, M., 2009, "A computer-based economic analysis for manufacturing process selection," International Journal of Agile Manufacturing, 11(1), pp. 11–18.

18. Esawi, A., and Ashby, M., 2003, "Cost estimates to guide pre-selection of processes," Materials and Design, 24(8), pp. 605–616.

19. Lee, C., 1992, "A knowledge-based systems approach for manufacturing process selection in design," Dissertation.

20. Allen, A., and Swift, K., 1990, "Manufacturing Process Selection and Costing," Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 204(143), pp. 143–148.

21. Loyer, J., Henriques, E., Fontul, M., and Wiseall, S., 2016, "Comparison of Machine Learning methods applied to the estimation of manufacturing cost of jet engine components," International Journal of Production Economics, 178, pp. 109–119.

22. Lovatt, A., and Shercliff, H., 1998, "Manufacturing process selection in engineering design. Part 2: a methodology for creating task-based process selection procedures," Materials & Design, 19(5–6), pp. 217–230.

23. Lovatt, A., Bassetti, D., Shercliff, H., and Bréchet, Y., 2000, "Process and alloy selection for aluminium casting," International Journal of Cast Metals Research, 12(4), pp. 211–225.

24. Thompson, M., Stolfi, A., and Mischkot, M., 2016, "Process chain modeling and selection in an additive manufacturing context," CIRP Journal of Manufacturing Science and Technology, 12, pp. 25–34.

25. Musti, S., 1988, "Coding From a Three-Dimensional CAD Database," Transactions of the ASME, 110(August).

26. Zehtaban, L., Elazhary, O., and Roller, D., 2016, "A framework for similarity recognition of CAD models," Journal of Computational Design and Engineering, 3(3), pp. 274–285.

27. Biasotti, S., Cerri, A., Bronstein, A., and Bronstein, M., 2016, "Recent Trends, Applications, and Perspectives in 3D Shape Similarity Assessment," Computer Graphics Forum, 35(6), pp. 87–119.

28. Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., and Ramani, K., 2005, "Three-dimensional shape searching: state-of-the-art review and future trends," Computer-Aided Design, 37(5), pp. 509–530.

29. Chen, X., Gao, S., Guo, S., and Bai, J., 2012, "A flexible assembly retrieval approach for model reuse," CAD Computer Aided Design, 44(6), pp. 554–574.

30. Deshmukh, A., Banerjee, A., Gupta, S., and Sriram, R., 2008, "Content-based assembly search: A step towards assembly reuse," CAD Computer Aided Design, 40(2), pp. 244–261.

31. Deshmukh, A., Gupta, S., Karnik, M., and Sriram, R., 2005, "A system for performing content-based searches on a database of mechanical assemblies," Design Engineering, Parts A and B, 2005(January), pp. 411–423.

32. Liu, Z., Bu, S., Zhou, K., Gao, S., Han, J., and Wu, J., 2013, "A survey on partial retrieval of 3D shapes," Journal of Computer Science and Technology, 28(5), pp. 836–851.

33. Bai, J., Luo, H., & Qin, F., 2016, "Design pattern modeling and extraction for CAD models," Advances in Engineering Software, 93, pp. 30–43.

34. Kim, D., 2004, "Interactive 3-D Shape Retrieval System Using the Attributed Relational Graph," IEEE.

35. Gao, W., Gao, S., Liu, Y., Bai, J., and Hu, B., 2006, "Multiresolutional similarity assessment and retrieval of solid models based on DBMS," CAD Computer Aided Design, 38(9), pp. 985–1001.

36. Pu, J., Kalyanaraman, Y., Jayanti, S., Ramani, K., and Pizlo, Z., 2007, "Navigation and discovery in 3D CAD repositories," IEEE Computer Graphics and Applications, 27(4), pp. 38–47.

37. Qin, F., 2014, "A deep learning approach to the classification of 3D CAD Models," Journal of Zhejiang University Science C. Computers and Electronics, pp. 91–106.

38. Chakraborty, T., 2005, "Shape-based clustering of enterprise CAD databases," Computer-Aided Design and Applications, 2(1–4), pp. 145–154.

39. Jayanti, S., Kalyanaraman, Y., and Ramani, K., 2009, "Shape-based clustering for 3D CAD objects: A comparative study of effectiveness," CAD Computer Aided Design, 41(12), pp. 999–1007.

40. Bespalov, D., Ip, C., Regli, W., and Shaffer, J., 2005, "Benchmarking CAD search techniques," Proceedings of the 2005 ACM Symposium on Solid and Physical Modeling - SPM, 1(212), pp. 275–286.

41. Peabody, M. and Regli, W., 2001, "Clustering Techniques for Databases of CAD Models," Technical Report DU-MCS-01-01.

42. Bespalov, D., Shokoufandeh, A., Regli, W., and Sun, W., 2003, "Scale-space representation of 3D models and topological matching," Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications, 208.

43. Al-Mubaid, H., Abouel Nasr, E., and Kamrani, A., 2008, "Using data mining in the manufacturing systems for CAD model analysis and classification," International Journal of Agile Systems and Management, 3(1/2), pp. 147–162.

44. National Design Repository. http://www.designrepository.org, 2004.

45. Hoefer, M., Chen, N., and Frank, M., 2017, "Automated Manufacturability Analysis for Conceptual Design in New Product Development," In Proceedings of the 2017 Industrial and Systems Engineering Research Conference. In Press.

46. Frank, M., Wysk, R., and Joshi, S., 2006, "Determining Setup Orientations From the Visibility of Slice Geometry for Rapid Computer Numerically Controlled Machining," Journal of Manufacturing Science and Engineering, 128(1), pp. 228–238.

47. Arabia, S., 2008, "The relationship between tool length/diameter radio and surface roughness in end milling applications," In International Congress on Sound and Vibration, pp. 1382–1389.

48. Chay, J., Jackman, J., Frank, M., Peters, F., 2017, "A New Metric For Evaluating Machinability of a Design," In Proceedings of the 2017 Industrial and Systems Engineering Research Conference. In Press.

49. Li, Y., and Frank, M., 2006, "Machinability Analysis for 3-Axis Flat End Milling," Journal of Manufacturing Science and Engineering, 128 (May 2006), p. 454.

50. Hoefer, M., Frank, M., and Dorneich, M., 2017, "Geometric Analysis to Automate Design for Supply Chain," In Proceedings of the 2017 Industrial and Systems Engineering Research Conference. In Press.

51. Venables, W. and Ripley, B., 2002, Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

52. Therneau, T., Atkinson, B., and Ripley, B., 2015, "rpart: Recursive Partitioning and Regression Trees." R package version 4.1-10.

53. Liaw, A., and Wiener, M., 2002, “Classification and Regression by randomForest,” R News, 2(3), pp. 18–22.

54. Breiman, L., 2001, “Random forests,” Machine learning, 45(1), pp. 5-32.

CHAPTER 4: GENERAL CONCLUSIONS AND FUTURE WORK

Conclusions

The contribution of this thesis is the development of an automated method that characterizes a conceptual design's geometry and uses that information to help select a suitable manufacturing process. To understand the relationship between a design and manufacturing process, algorithms analyze a 3D model to calculate geometry metrics that are associated with the machining process. For example, a visibility score is calculated that measures what percentage of a model's surface is visible from a machine tool. The machining metrics are used as inputs to a series of machine learning classification algorithms, including k-nearest neighbor (KNN), decision trees, and random forest. The accuracy of the machine learning models was measured using an independent test set of data, or in the case of random forest, the average out-of-bag (OOB) classification error. The algorithms were executed using machining metrics alongside traditional geometry measures such as volume to surface area, and "buy-to-fly" ratio. Included in the results is a presentation of which geometry metrics were most useful at classifying a part with respect to a manufacturing process.

An accuracy of up to 86% was observed with a random forest model. It appears the significant percentage of flat surface area in machined parts is a driving factor in the classification models, as the orientation of the individual facets was the most important attribute. There were, however, some misclassified models. For these models, some of the designs scored well on the machining metrics but were classified as cast parts. This classification error could be the result of a robust design, meaning the design was suitable for machining or casting, and/or other factors beyond geometry helped influence the original classification.

The success of the proposed machining metrics at predicting manufacturing process selection suggests that the metrics could provide information for predicting overall supply chain impact of a design. DFM based metrics, along with other design requirements, can serve as useful inputs to a machine learning model that helps predict supply chain impacts, such as cost, quality, and manufacturing process. As firms continue to collect more data about designs, manufacturing, and operations, automated knowledge discovery methods are required to enable data driven decisions during conceptual design. This thesis has presented a new automated method for design for supply chain, which requires characterizing geometry found in CAD files and using machine learning to understand how geometry affects sourceability. Machining metrics are introduced that can be used to effectively discriminate parts by manufacturing process.

Future Work

These metrics and methods serve as a groundwork for which future automated design for "X" systems can be created. Future systems will be able to include multiple other aspects of the product lifecycle, including maintainability, sustainability, safety, and quality. It will be important to experiment with these methods using more complete datasets. For example, these geometry metrics could be useful predictors of manufacturing cost or lead time when integrated with PDM and ERP systems. It is also critical to provide this information to the designer at an early stage. To accomplish this, tools will need to be created that provide real-time feedback during the iterative process of conceptual design. If designers can receive feedback about how their design affects downstream activities like manufacturing and supply chain, they can make proactive decisions that seek to optimize more than just product performance.

Another future research area is the use of machine learning for automated design for manufacturing and sustainability feedback. Data can be captured concerning which surfaces are the most difficult to process or result in high levels of defects, and then relationships with facet-based metrics can be discovered. This would allow for effective DFM feedback early on during conceptual design. In addition, life cycle assessments will yield data about various products and their impact on the environment. Similar machine learning methods can be used to compare the geometry of new designs with previous designs, to estimate the potential environmental impact of a design under development.