



HAL
open science

A systematic review on data of additive manufacturing for machine learning application: the data quality, type, preprocessing, and management

Ying Zhang, Mutahar Safdar, Jiarui Xie, Jinghao Li, Manuel Sage, Yaoyao
Fiona Zhao

► To cite this version:

Ying Zhang, Mutahar Safdar, Jiarui Xie, Jinghao Li, Manuel Sage, et al.. A systematic review on data of additive manufacturing for machine learning application: the data quality, type, preprocessing, and management. *Journal of Intelligent Manufacturing*, 2022, 10.1007/s10845-022-02017-9 . hal-04096906

HAL Id: hal-04096906

<https://hal.science/hal-04096906v1>

Submitted on 14 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A systematic review on data of additive manufacturing for machine learning application: the data quality, type, preprocessing, and management

Ying Zhang¹, Mutahar Safdar¹, Jiarui Xie, Jinghao Li, Manuel Sage, Yaoyao Fiona Zhao*

Department of Mechanical Engineering,

McGill University, Montreal, Quebec H3A0G4

Email: ying.zhang8@mail.mcgill.ca, mutahar.safdar@mail.mcgill.ca, jiarui.xie@mail.mcgill.ca, jinghao.li@mail.mcgill.ca, manuel.sage@mail.mcgill.ca, yaoyao.zhao@mcgill.ca

Abstract

Additive manufacturing (AM) techniques are maturing and penetrating every aspect of the industry. With more and more design, process, structure, and property data collected, machine learning (ML) models are found to be useful to analyze the patterns in the data. The quality of datasets and the handling methods are important to the performance of these ML models. This work reviews recent publications on the topic, focusing on the data types along with the data handling methods and the implemented ML algorithms. The examples of ML applications on AM are then categorized based on the lifecycle stages, and research focuses. In terms of data management, the existing public database and data management methods are introduced. Finally, the limitations of the current data processing methods and suggestions are given.

1. Introduction

Additive manufacturing (AM, also known as 3D printing) refers to a process by which a 3D object is additively fabricated from a digital design, usually in a layer-by-layer fashion. Due to the nature of the AM process, which adds material only at the desired place, the lead-time and material waste are kept to a relatively low level. More importantly, since the tooling is no longer needed, AM unlocks a significant number of constraints for the designers to design products with complex geometry [1]. The technical features of AM make it suitable for industries with small batch sizes and complex part geometry, especially for hollow structures and curved surfaces. Thus, AM is becoming an important complement to the traditional manufacturing methods.

American Society for Testing and Materials (ASTM) F42 categorizes all the AM processes into seven classes, with several technologies under each category branded by different manufacturers [2, 3]. Compared to conventional manufacturing methods, there are more uncertainties in AM processes as they usually involve complex designs, phase transformation, and in-process control to achieve the desired microstructure and properties. In the meantime, the enormous data generated from design, process, structure, and property (PSP) linkage [4, 5] in AM techniques brings challenges to researchers, and the traditional trial and error methods are usually not efficient. Considering the melt pool size alone, it can change from hundreds of micrometers in powder bed fusion (PBF) to several millimeters in directed energy deposition (DED), and this may eventually result in very different microstructures and properties. The uncertainties in the PSP are restricting the development and application of AM techniques [6], and recognizing the interplay between PSP linkage in AM is crucial for quality control and the development of this technology. Modeling approaches and numerical simulations are the ideal tools to fill in the gap by saving time and experimental costs [7, 8]. However, establishing mechanistic models usually requires domain knowledge and model validation process before their reliable applications. In recent years, an

¹ Co-first authorship

* Corresponding author

increasing number of studies are looking into solutions by directly applying data-driven approaches. Machine learning (ML) continuously shows its advantages to deal with the uncertainties and multiple types of data from AM for various purposes. First, it avoids solving complex equations in many physical models. Second, it fills the gap where traditional theories are either non-existent or loss function due to their inability to solve application-specific problems. Finally, it is able to make a prediction for a long process range based on the well-tested and reliable ML algorithms encoded by open-source software [9].

When ML gets involved, datasets are an important matter to be considered as ML models are solely driven by data. Current AM datasets for ML are highly case-dependent due to a lack of existing standards. Therefore, in Section 2, this work first provides a comprehensive review of the state-of-art ML applications on AM in terms of the four data types (tabular, graphic, 3D, and spectrum) and their corresponding data handling procedures as well as the ML algorithms deployed. Then, based on the prediction target of ML models, the application area of ML on AM is summarized in Section 3. The existing public datasets are introduced in the following Section 4. Finally, some insights and suggestions for further developing trends are discussed in the end.

All the references come from Scopus including most of the journal papers with the keywords “machine learning” and “additive manufacturing” in the recent four years. 144 related journal articles have been reviewed. For convenience, the standards from ASTM to categorize the AM technology types are used and the list of abbreviations for AM aspect is listed in Table 1. The list of abbreviations for the most frequent ML terminology is listed in Table 2.

Table 1: The abbreviations of AM aspect used in this review

| Abbreviation | Definition |
|---------------------|----------------------------|
| AM | Additive manufacturing |
| ME | Material extrusion |
| PBF | Powder bed fusion |
| MJ | Material jetting |
| VP | Vat photopolymerization |
| DED | Directed energy deposition |

Table 2: The abbreviations of the most frequent ML terminology used in this review

| Abbreviation | Definition |
|---------------------|------------------------------|
| ML | Machine learning |
| NN | Neural network |
| MLP | Multilayer perception |
| ANN | Artificial neural network |
| CNN | Convolutional neural network |
| RNN | Recurrent neural network |
| k-NN | k-nearest neighbor |
| SVM | Support vector machine |
| SVR | Support vector regressor |
| RF | Random forest |
| GB | Gradient boosting |

| | |
|------|------------------------------|
| LDA | Linear discriminant analysis |
| DL | Deep learning |
| LR | Linear regression |
| GPR | Gaussian process regression |
| NB | Naïve Bayes |
| DT | Decision tree |
| LSTM | Long short-term memory |

In the literature, there are multiple survey papers on ML and AM. Reviews on ML applications in AM can be categorized into two groups: Specific and General. Specific reviews focus on either a subtopic in AM (process [10], material [11], monitoring [12, 13]) or in AI (model type [14, 15]). General reviews focus on the overall application of ML in AM [16-22]. However, those articles mainly focus on the opportunities of ML in AM. None of them provide a summary of the data perspective of AM for ML applications such as the most popular data type in AM, the typical data preprocessing methodology, data quality and quantity, and AM data management. Therefore, the primary objective of this article is to provide a comprehensive review on the data of AM for ML application. Specifically, this study contributes to:

- Introduce the most popular data types and data handling methods in AM studies.
- Summarize the data types, preprocessing methods, specific application, data quality, and selected ML algorithms for 100+ existing studies in terms of AM lifecycle stages of Design, Process, and Product.
- Present the existing public datasets for AM studies and introduce the existing AM database management system.
- Identify the research gaps, current limitations, and future research directions on AM data for ML applications.

2. Data types and data handling

This section discusses the original data types collected from the design, simulation, and manufacturing procedures of AM to construct ML models. Original data in this paper refers to the collected raw data before any further preparation or preprocessing. The data handling methods and popular categories of ML algorithms implemented for ML in AM are also illustrated in this section.

2.1. Overview

Research on ML and its applications utilizes many different terminologies for the work with data. These terms are often used interchangeably, complicating the transferability of implementations. For this paper, **data handling** is defined as an umbrella term comprising all work with data, from the collection of raw data to techniques solely applied to improve algorithm performance. In the field of ML-aided AM, data handling can be further divided into the four categories explained in the following:

Feature extraction: raw data acquired from sensors is usually in forms that are difficult to statistically analyze and build ML models upon. ML practitioners need to develop salient representations from the original data as the input features. Therefore, statistical and geometrical features are usually extracted from raw data to facilitate the subsequent analysis [23].

Feature selection and feature learning: after original features are derived from raw data, feature selection and feature learning can be performed to improve the performance of ML models. Feature selection evaluates and ranks the importance of features to choose a subset from the original features [24]. Feature learning, instead, generates new features that are better representations of the dataset [25].

Discretization: discretization describes the process of transforming continuous objects into discrete elements. Vision data such as images and 3D models are commonly discretized to generate features suitable for ML. The most popular discretization methods for AM are pixelization and voxelization.

Data preprocessing: data preprocessing summarizes techniques applied to improve the quality or suitability of the data for ML algorithms [26]. For AM, preprocessing can be further classified into image preprocessing and numerical preprocessing. Images are often processed with techniques such as gray-scaling and cropping to reduce computational complexity and enhance the quality of data [27]. Numerical preprocessing techniques such as normalization and removal of outliers support better performance and faster convergence for the training processes of ML [28].

2.2. Tabular data

Data stored in tabular forms is structured into rows and columns, shown in Figure 1. Each row represents an instance, acting as a training example for ML [29]. Each column stands for an attribute, also known as a feature in ML, to characterize the instances. The data in cells could be continuous such as numerical values, or discrete such as categorical or ordinal values. The cells under the same column must exhibit the same category of data. For AM, this data type is extensively used to investigate the correlations between build parameters and build quality of the print [30-32]. For example, Zhang et al. [30] construct an ML model that predicts the build quality with build parameters using tabular data. Each row records one experiment, consisting of build parameters as the input features and measurements of build quality as the target variables of the ML model. Figure 1 is an example of a tabular form, storing the experimental data of different process parameter settings of plastic 3D printing.

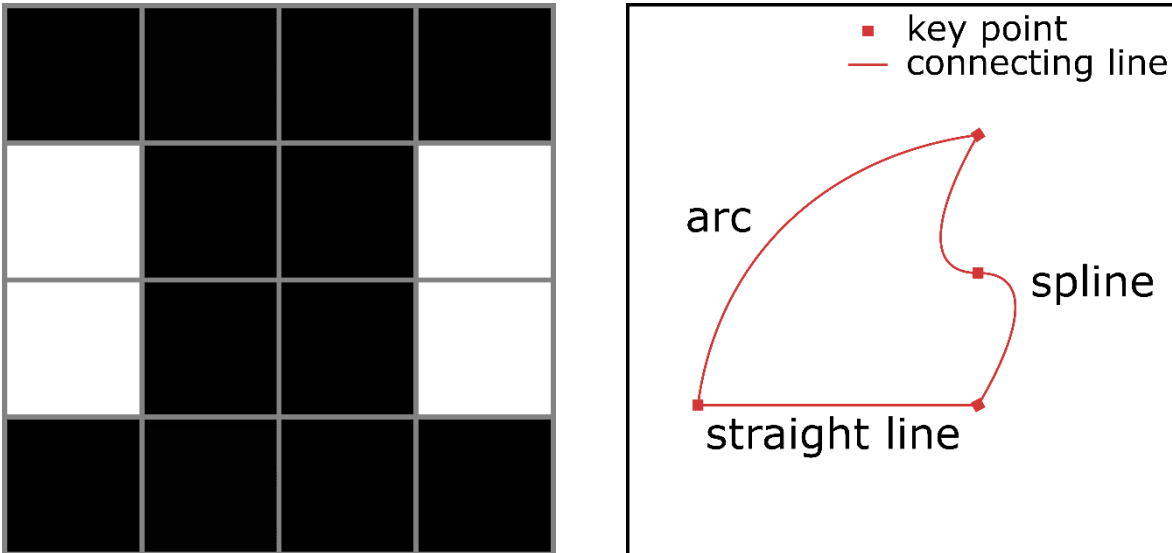
| Experiment ID | Layer thickness (mm) | Bed temperature (°C) | Nozzle temperature (°C) | Print speed (mm/min) | |
|---------------|----------------------|----------------------|-------------------------|----------------------|-------|
| 1 | 0.11 | 50 | 190 | 2000 | |
| 2 | 0.12 | 52 | 192 | 2100 | |
| 3 | 0.13 | 54 | 194 | 2200 | |
| 4 | 0.14 | 56 | 196 | 2300 | |
| 5 | 0.15 | 58 | 198 | 2400 | |
| | | | | | |

Figure 1: Format of tabular data: an example of tabular data form to store the experimental data of plastic 3D printing.

The most frequently used data handling method for tabular data is feature selection [33, 34]. Feature selection processes are easily implemented for tabular data because features are well defined at the beginning. The most relevant features for target prediction are selected to construct ML models to reduce the computational complexity while retaining acceptable accuracy. Feature selection also helps discover the most decisive build parameters and remove unnecessary sensors to reduce cost. Data preprocessing such as removal of outliers and normalization are also common practices to improve the predictive performance of ML models [35].

2.3. Graphics data

Graphics data refers to data forms that display or present images or drawings [36]. There are generally two types of graphics data: bitmap and vector data. Bitmap is a pixelized representation of images, where a picture is segregated by a uniform grid and each cell is numerically defined to represent its color. Figure 2(a) renders the shape of an I-beam cross-section using one of the bitmap data representations: binary image. Each block is a black or white pixel, digitally represented by 1 or 0, respectively. In-process monitoring and quality inspection on AM are usually performed with bitmap data acquired from digital cameras [37, 38], microscopes [39-41], X-ray sensors [42], and thermal cameras [43, 44]. For instance, layer-wise images are captured by digital cameras to monitor the conformance of geometry, or by microscope to inspect defects and microstructures. Commonly used bitmap formats are the joint photographic expert group (JPEG) and Portable graphics format (PNG). Vector data stores the key points of a drawing that can be restored with connecting lines and guidelines such as straight line, spline, and arc. Figure 2(b) is an example of a sketch reconstructed from vector data. The key points are connected with connecting lines following the guidelines such as straight line, spline, and arc. Vector data is frequently utilized to characterize the toolpaths and infill patterns of AM [45].



(a) (b)
Figure 2: Schematics of graphics data: (a) binary image – a type of bitmap data; and (b) a sketch restored by vector data

Pixelization is the most popular data handling method for ML in AM as most examples of image analysis in this field adopt bitmap data [46-50]. When an image is pixelized, each pixel is an original feature of the dataset. However, they might not be good representations for ML, and thus image preprocessing is commonly utilized to improve the datasets. Techniques such as cropping [46, 51], resizing [48], gray-scaling [41, 49], and binarization [52] are utilized to reduce the computational power required. Image augmentation [53] is also frequently conducted to generate more training examples using flipping [51], rotating [54], etc. Image analysis techniques including texture analysis [55] and edge detection [56] are implemented to extract geometrical features. An example of a monitoring system that employs denoising and edge detection is shown in Figure 3.

An image of the melt pool is captured in a pixelized format, then analyzed to generate its temperature contour map. Principal component analysis as a feature learning technique has been applied to graphics data to reduce the dimensionality of input variables of AM datasets [41, 51].

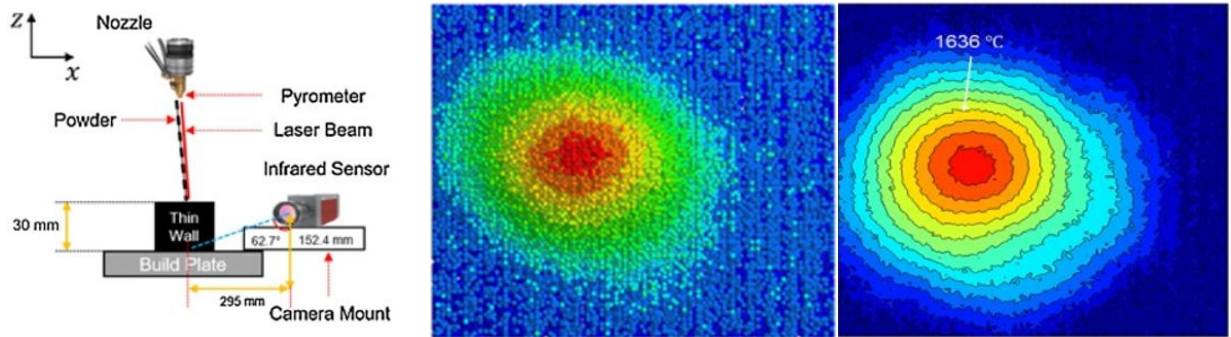


Figure 3: Example of denoising and edge detection: In-process melt pool monitoring system of the fusion zone measured by an infrared camera during DED Ti6Al4V (left), the image data gathered (middle), and the processing of denoising and edge detection (right). [57]

2.4. 3D data

3D data is a coordinate-based representation of 3D objects with specialized software and has wide applications in AM [58]. For example, the part definition of plastic 3D printing is saved in Computer-aided design (CAD) files such as STL. The CAD files are processed by slicing software that generates toolpaths to guide the extruder. Figure 4 demonstrates how 3D objects can be represented by 3D data using tessellation. Tessellation is one of the 3D data techniques that cover the surface of a 3D object with polygons to record its shape.[59]. In AM researches, 3D design files are utilized to train ML models that perform geometry detection [60], manufacturability analysis [60], and build quality prediction [61]. 3D data can also be constructed with stereo cameras or X-ray tomography to perform in-process defect detection [50, 62].

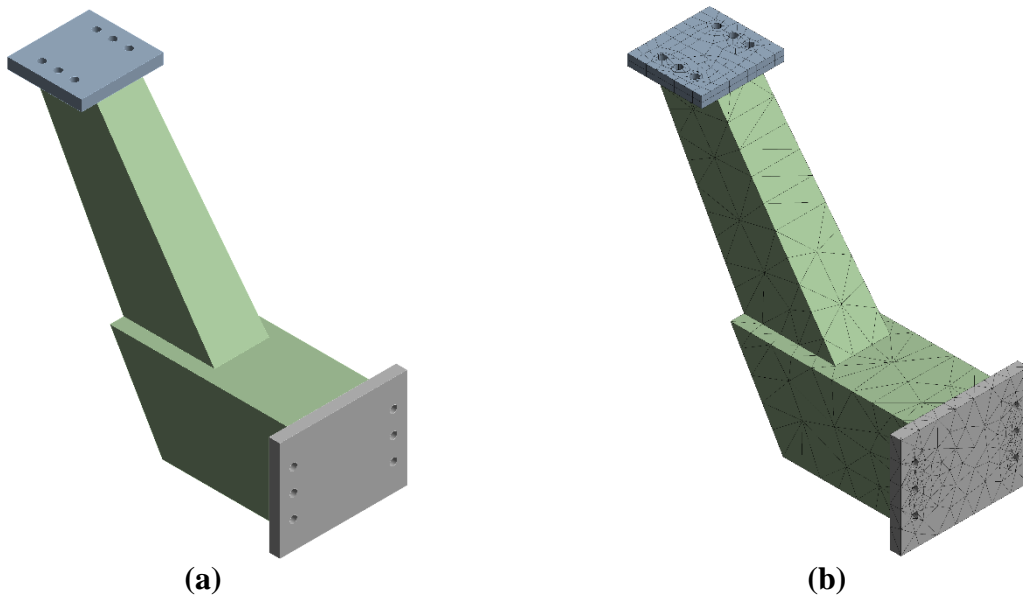


Figure 4: 3D object represented by 3D data using tessellation: (a) original 3D object; and (b) tessellated 3D object.

Before input into ML models, 3D data is processed into multi-view images, volumetric, point cloud, polygonal mesh, or primitive-based data, while point cloud and volumetric forms are most popular for AM. 3D models can also be voxelized into cubes to generate original features. Sparse representations are deployed to reduce the memory needed to process huge voxelized data [63]. Geometrical features such as relative distances extracted from point clouds can be the input features to train ML models [64]. Morphological and crystallographic features can also be extracted from volumetric data as input features to ML models [61]. Figure 5 demonstrates an example of handling methods for 3D data. The features such as volume, surface area, bounding box, number of components, and number of fasteners are extracted from the 3D model and then used as ML input features.

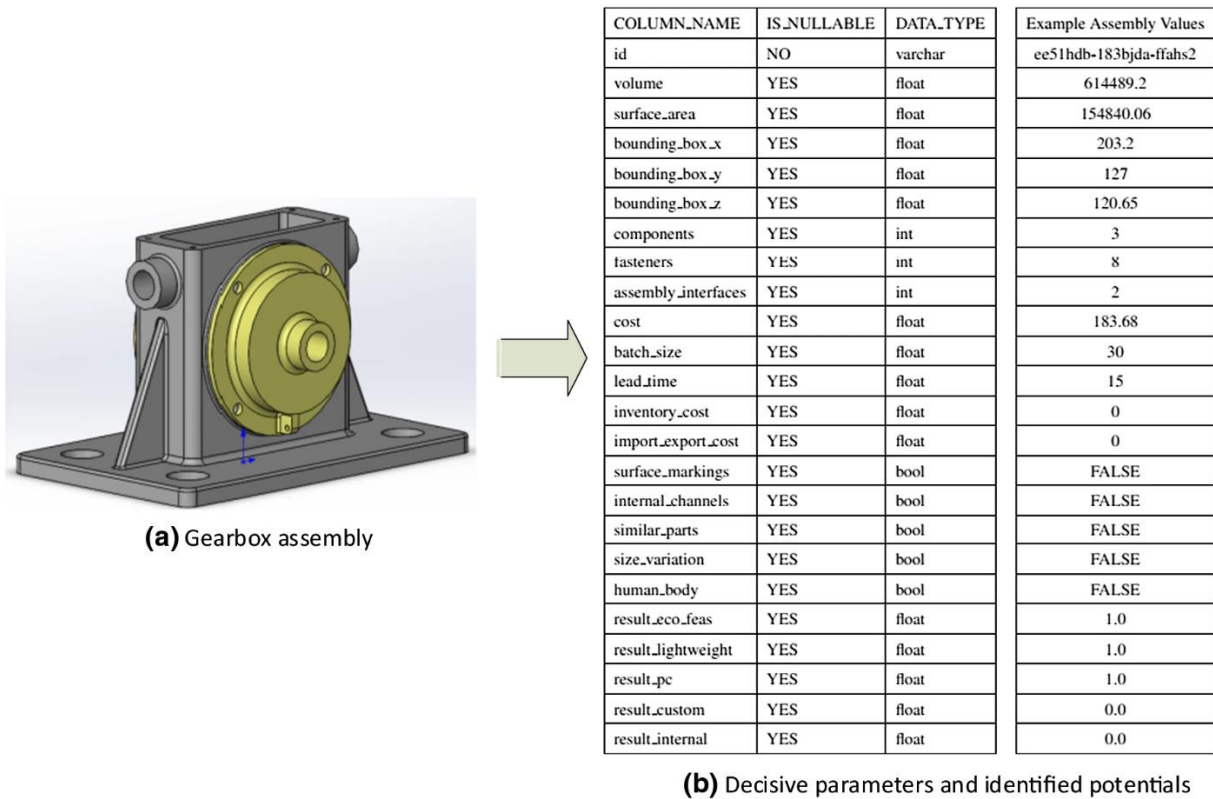


Figure 5: Example of feature extraction from 3D model [65]

2.5. Spectrum data

Spectrum data is discrete or continuous wavelengths measured from specific radiations. Figure 6 shows an example of vibration measurement, which belongs to spectrum data. Vibration data is usually first recorded in the time domain, then transferred to the frequency domain to reveal more information. Frequently measured radiations for AM are temperature [33], acoustic emission (AE) [66-68], photon [34, 69], and vibration. The corresponding sensors are typically thermocouples, AE sensors, photodiode sensing systems, and vibration sensors, respectively. Thermal sensors can help monitor melt pool states and reveal cracks and voids near the part surface by analyzing the thermal gradients and history of the target area. AE sensors can be attached to the build platform to detect any onset of irreversible deformations such as cracks, warpage, and delamination.

Vibration sensors can be installed on extruders to detect machine state errors such as filament runout, jamming, and breakage [70].

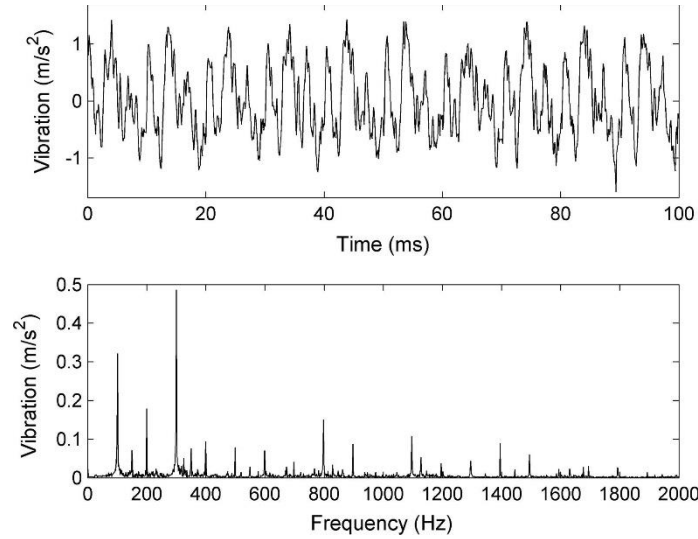


Figure 6: Measurement of vibration – one type of spectrum data. The plot on the top is the vibration amplitude in the time domain. The plot at the bottom is the vibration amplitude in the frequency domain (adapted from [71]).

Denoising and signal filtering are common techniques to clean spectrum data [41]. For many cases, spectrum data collected by the aforementioned sensors is times-series data with temporal relationships among data points, adding difficulties to data analysis. Statistical features extracted from both time and frequency domains are proven efficient to build data-driven tools with spectrum data [43, 72, 73]. Time-frequency analysis such as wavelet packet transform is implemented to extract statistical features from spectrum data [67]. LDA and principal component analysis as a feature learning method has been applied to AM spectrum datasets to reduce the input dimensionality [34, 74].

2.6. Machine learning techniques for additive manufacturing

This paper reviewed the reported ML techniques used in the ML in AM literature and plots the percentages by ML type in Figure 7. One differentiating factor of ML is the type of learning. Supervised learning algorithms are trained with labeled data and, during test time, seek to identify the correct label for a queried instance. With over 90% of the analyzed literature falling into this category, supervised learning is prevalent in the field. Less prominent are unsupervised approaches, intending to recognize patterns in (unlabeled) data, and reinforcement learning.

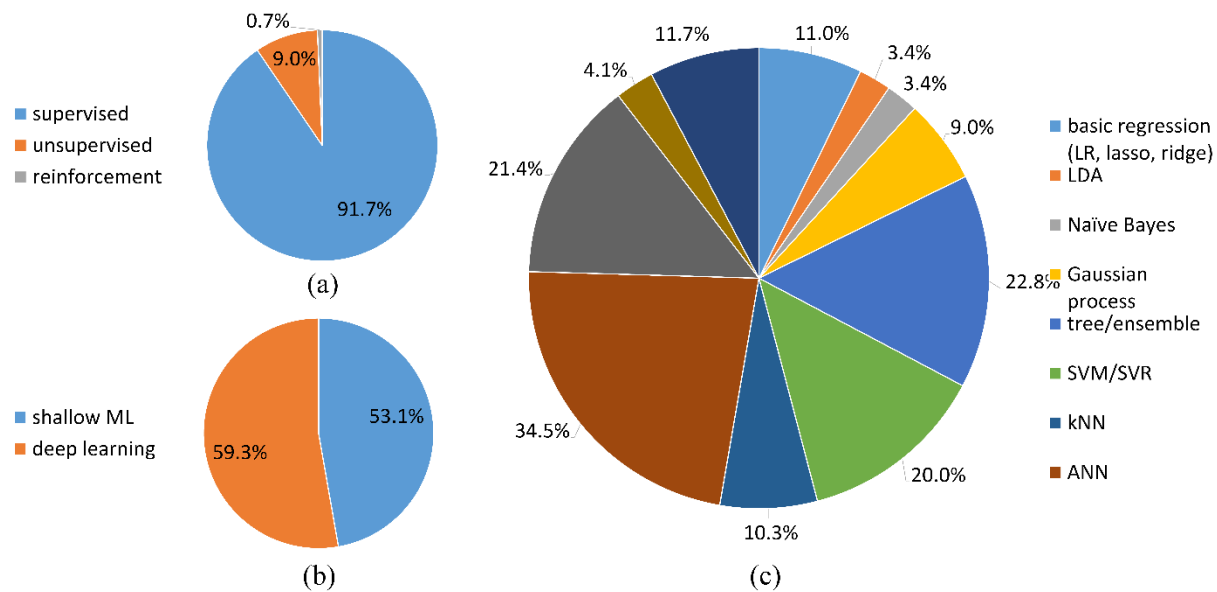


Figure 7: Popularity of ML techniques in the percentage of the 144 reviewed papers. Grouped by (a) types of learning, (b) types of ML architecture, and (c) algorithms or groups of algorithms. Note that charts do not sum to 100% as sever

Furthermore, ML algorithms can be associated with either shallow or deep learning. Deep learning algorithms are based on ANN and its adaptations such as CNNs or RNNs. In Figure 7, shallow ML describes those algorithms that do not fall into the category of deep learning. It can be seen that deep architectures are slightly more common in the field compared to shallow ones.

Figure 7c visualizes the popularity of certain algorithms or groups of algorithms with similar working principles. Basic regression algorithms comprise LR, ridge regression, and lasso regression and are implemented in 16 publications. Other conventional ML algorithms deployed in the field are NB, GPR, k-NN, and maximum margin algorithms such as SVMs and SVRs. Algorithms in the tree/ensemble category are based on decision trees and make use of either single trees or ensembles, for example in the case of RF and GB.

Table 3 displays the popularity of ML types introduced above for each of the four data types. For tabular data, it is observed that the proportions do not significantly differ from those reported overall (see Figure 7 or the first row of Table 3). Only within the group of deep learning algorithms, normal ANNs were more often preferred over CNNs and RNNs.

Table 3: Popularity of ML algorithms overall and by data type. Bold cells highlight ML categories or algorithms that are significantly more popular for one data type (>10 percentage points compared to average). Note that categories and rows do not sum to 100% as several review papers employed multiple techniques for one task.

| | supervised | unsupervised | shallow ML | deep learning | basic regression (linear, lasso, ridge) | Gaussian process | tree/ensemble | margin (SVM, SVR) | kNN | ANN | CNN | RNN | other |
|----------|--------------|--------------|---------------|---------------|--|------------------|---------------|-------------------|-------|--------------|--------------|------|-------|
| all | 91.7% | 9.0% | 53.1% | 59.3% | 11.0% | 9.0% | 22.8% | 20.0% | 10.3% | 34.5% | 21.4% | 4.1% | 11.7% |
| tabular | 95.0% | 5.0% | 56.7% | 58.3% | 11.7% | 13.3% | 26.7% | 20.0% | 6.7% | 48.3% | 8.3% | 1.7% | 11.7% |
| graphics | 90.9% | 9.1% | 49.1% | 63.6% | 7.3% | 5.5% | 23.6% | 18.2% | 12.7% | 25.5% | 36.4% | 7.3% | 7.3% |
| 3D | 92.0% | 8.0% | 48.0% | 76.0% | 12.0% | 8.0% | 16.0% | 16.0% | 16.0% | 40.0% | 32.0% | 4.0% | 20.0% |
| spectrum | 78.6% | 21.4% | 100.0% | 57.1% | 28.6% | 21.4% | 42.9% | 57.1% | 14.3% | 42.9% | 14.3% | 7.1% | 14.3% |

For graphics data, Similar to tabular data, solely the popularity of deep learning algorithms differs from the overall usage in the field. CNNs are significantly more frequently seen in publications with graphics data (36.4% compared to 21.4% overall), while ANNs are deployed less (25.5% compared to 34.5%). 3D datasets are less frequently combined with shallow ML models. Most of the analyzed literature for this data type (76%) utilized deep learning algorithms. Both ANNs and CNNs are more common with 3D datasets compared to the field.

Spectrum data shows various differences when focusing on ML models compared to the other three data types. Unsupervised approaches are more popular, deployed in 21.4% of the analyzed literature with this data type as opposed to 9% on average over all 4 data types. Furthermore, all publications with spectrum data utilized shallow ML algorithms. Particularly popular compared to the other data types are tree/ensemble methods and maximum margin methods. Similar to tabular data, within the group of deep learning algorithms, ANNs are more often preferred over CNNs.

2.7. Discussion on data types and data handling techniques

Table 4 is the pivot table indicating, for all four data types, the number of papers that have utilized a data handling technique belonging to the four categories introduced above. Tabular and graphics datasets have been extensively researched and utilized to construct ML models for AM, while 3D and spectrum datasets appear less frequently. Feature extraction techniques have been widely applied to the analysis of all data types, especially for spectrum data. Representations derived from feature extraction methods with domain knowledge are better input features than the original feature sets in many papers. Discretization is only applicable to bitmap and 3D models. Pixelization has been implemented in 84% of the reviewed papers with graphics data, while voxelization only appears in 32% of the reviewed paper with 3D data. Image preprocessing methods have been extensively utilized to improve the quality of graphics datasets. It is suspected by the authors that data preprocessing methods should be more frequently implemented than the observations of Table 4, as techniques such as normalization are generally implemented but barely mentioned. Feature selection techniques have been applied to more than 25% of the tabular and spectrum datasets reviewed. Tabular datasets have original features readily available for feature

selection to improve ML model performance. And many spectrum datasets need feature learning techniques to reduce their input dimensionality while retaining the most information.

Table 4: Frequencies of data handling methods utilized to process each data type of AM dataset. Color coding is applied to this table to indicate large numbers with red and small numbers with green. The total number of papers reviewed is stated at the right-most column, corresponding to each data type.

| | Feature extraction | Discretization | Data preprocessing | Feature selection and feature learning | Total |
|----------|--------------------|----------------|--------------------|--|-------|
| Tabular | 10 | 0 | 11 | 15 | 59 |
| Graphics | 18 | 46 | 31 | 7 | 55 |
| 3D | 12 | 8 | 9 | 3 | 25 |
| Spectrum | 14 | 0 | 2 | 6 | 20 |
| Sum | 54 | 54 | 53 | 31 | |

3. Targets/Applications

Targets represent outputs of ML models and specify the AM application for which an ML model is trained and deployed. ML applications in AM can be classified for lifecycle stages. In this section, these applications are broadly categorized into pre-processing, processing, and post-processing stages. The selection of an appropriate target within a lifecycle stage can expedite the modeling process and generate desired results. At the design stage, ML can be used for a range of tasks including but not limited to geometry prediction, design optimization, lattice design, and design classification. The main focus of ML applications in AM has undoubtedly been on the process stage as this can lead to corrective actions before a part is completely printed. At the process stage, ML can be used to predict optimal process parameters and identify defective process states. Apart from these applications, build and toolpath planning has been optimized according to different constraints through the use of ML models. Surrogate modeling of melt pools is gaining attraction to replace computationally expensive and time-intensive physical simulations of AM processes. In this regard, recent literature highlights the significance of effective surrogate modeling by combining ML models with physics-based models. The final category of ML applications in AM concerns product characteristics. Product characteristics have been divided into macro, micro, and mechanical properties (alongside other characteristics). Macro characteristics deal with the macrostructure and include dimensions, surface features, cross-sectional parameters, and visual defects. Micro characteristics include ML applications where microstructural defects are evaluated. A lot of attention is being given to control microstructural characteristics through ML models as this leads to overall control on product performance. Product properties are also predicted through ML models in some applications.

AM targets are linked with data types, handling techniques, algorithms, and instances in the form of tables. Trends of data handling techniques and ML algorithms within each data type were the focus of attention in the previous section. In this section, data types for specific applications are discussed in terms of trends and representative works. AM types are also mentioned in the tables

to indicate the frequency of ML-oriented research in each category. The number of instances indicates the comprehensiveness of each dataset and is a focus of discussion in the later sections of this survey. These tables have been arranged to depict common AM applications within each lifecycle stage. There is a total of seven such tables spanning design (Table 5), process parameters and process domains (Table 6), build and toolpath planning (Table 7), surrogate modeling of melt pool (Table 8), macro product characteristics (Table 9), micro product characteristics (Table 10), and mechanical properties with miscellaneous characteristics (Table 11).

3.1. Design characteristics

At the design stage, ML models can aid either in the material's design or structural design. Material design can be divided into homogenous and heterogeneous material design. Traditionally, design optimization techniques such as topology optimization and generative design have been the essential tools for structural design. Recently, ML models have been applied in conjunction with these optimization tools to improve their accuracy and reduce computational expense. Apart from optimizing structures, ML is being used to cope with AM constraints namely supports and overhangs. ML is also gaining attraction in the design of lattice, an active research field in design for additive manufacturing (DfAM). Prediction of geometric characteristics is found to be the focus of ML applications at this stage. ML applications in AM design are shown in Table 5.

ML inputs in tabular form account for the majority of selected design targets making it the most common data type used at this stage. Tabular data can be extracted from a multitude of sources such as process parameters [75], lattice designs [76], spatial parameters [77], and simulations [78]. For instance, tabular data is used to explore a new meta-material concept that can adapt concerning different properties, base materials, length scales, and processes [78]. Tabular data of performance characteristics (stress-strain requirements) is mapped with the design parameters of an ankle brace [79]. Several ML models are employed to determine the dimensional features of printed parts by using tabular data of spatial parameters from STL and build orientation [77]. Apart from tabular data, graphic data is also used to predict design characteristics in AM. Graphic data can be captured through in-process vision-based sensors or microscopes and scans in the post-process stage. A composite material part's geometry and tool path are reversed engineered using CT scan images in an RNN [80]. Similarly, graphic data from lattice designs is also used to predict design-oriented targets [76, 81]. Cases, where 3D data is used to predict design characteristics, are found in the literature as well. Design files from the upstream section of product informatics are found to be the source of 3D data. These design files are either based on native CAD systems or reversed engineered through miscellaneous techniques. For example, image segmentation is performed using CNN to separate bone and background as a pre-process in medical AM [82]. In another work, 3D data from CAD models is used to identify parts eligible for AM [65].

There are several benefits of using ML at AM design phase. A major motivation for applying ML at the design stage is the fact that the part has not been printed which leads to empirical modeling of different design aspects in a direct (e.g., following the PSP chain) or indirect (e.g., using process or product data to model design characteristics) manner. This can result in significant cost and time savings. A key application in this regard is AM candidacy evaluation before a CAD file is sent downstream in the digital thread to incorporate build and machine information. The notion of success at the design phase can guide designers to design parts that are AM compatible for a specific application. Another form of designer guidance is through design rules represented in the

form of explicit AM knowledge. These rules can be formed using simple data-driven models. Finally, design parameters such as dimensions of features can be efficiently predicted using ML.

Table 5: ML applications at the design stage of AM

| Design Characteristics | | | | | | | |
|-------------------------------|------------|---|--|--|--|------------------|------------|
| Year | AM | Data | Data Handling | ML Algorithm | Application | Instances | Ref |
| 2021 | ME | Tabular data from parts and process | None mentioned | LR, NN | Dimensions (LWH) of CAD model | 450 | [75] |
| 2020 | ME | Tabular data of stress strain values | Get stress-strain response from simulation and use it in | ANN DL | Geometry parameters | 300 | [79] |
| 2020 | ME | CT-scan image | Pixelization | RNN | Direction of fibers | 78,373 | [80] |
| 2019 | ME | Tabular data from FEA | Sensitivity analysis to select the critical features | Bayesian machine learning | Design Classification | Not clear | [78] |
| 2020 | PBF | Tabular data of spatial parameters | Pearson correlation, normalization | LR, MLP, GB | Dimensional features (thickness, width and length) | 434 | [77] |
| 2021 | General AM | Tabular data Graphic data from lattice designs | None mentioned | CNN | Final design of lattice structure | 3,500+ | [76] |
| 2020 | General AM | Tabular data Graphic data of 2D lattice structures | Encode the structure then decode | DL | Design of micro lattices | 2,500 | [81] |
| 2020 | General AM | 3D data from CAD models | Parameter extraction using customized candidacy criteria | Boosted DT, NN, LR, decision forest regression, Bayesian linear regression | To identify parts are eligible for AM or not | 200 | [65] |
| 2018 | General AM | 3D data from bones, Graphic data from CT images | Normalization, patch selection | CNN | Bone or Background | 2,000,000 | [82] |

3.2. Process characteristics

Lack of repeatability has influenced process modeling in AM [83, 84]. Empirical (data-driven) and physical (physics-driven) modeling are two main approaches in this regard [85, 86]. Data-driven models provide a range of advantages over physics-based numerical or analytical modeling. There exist multiple bottlenecks which render physical modeling inefficient. The computational cost and time make the use of these models impractical. Another issue with these models can be overly simplistic assumptions without the important physical context required to solve complex AM processes. As a result, ML has emerged as a popular choice to understand AM processes. ML applications span a wide range of targets at the process stage, including process parameters, process domains, process planning, and melt pool modeling. These applications are linked with data types in the sections below.

The benefits of applying ML at the process stage can range from as simple as process state correction to more sophisticated applications such as operator guidance. Process window exploration is analogous to design space exploration where a multitude of factors can influence the final success of a process. The application of ML at this stage can help discover the patterns in the joint distribution of these factors. These patterns can be related to process parameters to serve as a benchmark for future processes. The ML-enabled shift from post-processing to in-process monitoring is probably the most significant advantage at this phase. This shortens the printing cycle while saving costs spent on product inspection and similar activities. The development of reliable AM data-driven models can replace computationally expensive process models. A well-developed and generalized ML model is efficient to use on the fly which suits its application during AM processes. These models can then be linked to operator-friendly GUIs (APIs for inter-application usage) to support actions such as parameter or path selection for specific applications.

3.2.1. Process parameters and process states

Most of the applications in this section are concerned with the process states and process parameters. A process state refers to the current state of a process which can be a custom-defined label. These targets can range from as simple as good/bad and acceptable/unacceptable to specific anomalies such as cyber-attacks or faulty conditions. There are numerous instances of using ML to model process anomalies [87-90] and conditions of interest in AM [52, 91]. Process parameters are of interest as ML models can help predict and optimize these with respect to various quality metrics. As a result, many researchers have focused on parameter prediction or optimization using ML techniques [92-94]. Process parameters related to deposition [95], material [40, 96], and energy source [39, 54, 97] have been a common target of ML models at this stage of AM lifecycle. These applications are highlighted in Table 6.

Data types used for process parameter and state prediction represent a relatively diverse set as compared to data types at the design stage. Tabular data is again found to be the most prevalent with 45% of the targets being predicted from this data type. Process information is found to constitute a significant portion of tabular data for these predictions. Process information for tables is extracted from process parameters [89, 92, 95], process conditions [52], and in-process images [97]. Tabular data is used to predict diverse parameter and state targets including deposition [95], critical velocity [98], cooling time [92], powder spreading quality [96], process conditions [52, 89], and key parameters of interest [94]. Graphic and spectrum data types have equal representation with each being used to predict 25% of AM targets at these stages. Digital cameras and microscopes are two main sources of graphic data in this regard. Graphic data in the form of images

is used to predict specific conditions [39, 52, 99] and detect process anomalies [87, 100]. Acoustic emission sensors appear to generate virtually all spectrum data for this stage. Process state prediction is found to be a prominent target of spectrum data. For example, the logistic regression model is developed from acoustic signals for identifying the potential of recreating G-codes through cyber-attacks [101]. ML applications where 3D data is used to predict process parameters are also found in the literature [54].

Table 6: ML applications at the process stage of AM – Process parameters and domains

| Process Characteristics - <i>Process parameters and process domains</i> | | | | | | | |
|--|-----------|---|--|--|-------------------------------------|------------------|------------|
| Year | AM | Data | Data Handling | ML Algorithm | Application | Instances | Ref |
| 2021 | ME | Tabular data on printing parameters | Feature Standardization | NN | Force in nozzle | 20,000 | [95] |
| 2019 | ME | Spectrum data from acceleration and temperature sensors | Instance Conversion, Attribute transformation and selection | SVM, NB, RF, k-NN | Presence or Absence of interference | 523,000 | [88] |
| 2019 | ME | Acoustic emission signal | Extract features from signals | Self-organizing map | Process failure | 213 | [66] |
| 2018 | ME | Acoustic emission from the process and tabular data of G-code | Signal filtering and feature extraction | SVM; LR | Process parameters | 442 | [101] |
| 2016 | ME | Heterogeneous sensor signals | Convert sensor signals to underdetermined linear system of equations | Online sparse estimation-based classification | Extrusion conditions | 2,000+ | [91] |
| 2021 | PBF | Microscope image | Segmentation | CNN | Laser polishing conditions | 432 | [39] |
| 2021 | PBF | Tabular data of material parameters | Feature filtering to eliminate irrelevant or redundant features | ANN | Critical velocity | Not clear | [98] |
| 2020 | PBF | Tabular data of process parameters | None mentioned | LR, stepwise linear regression, quadratic SVM, GPR, DT | Cooling time | 30 | [92] |
| 2020 | PBF | Digital image | Pixelization | CNN | Process anomaly detection | Not clear | [100] |

| | | | | | | | |
|------|-----|---|--|---|-------------------------------|--------|------|
| 2019 | PBF | Tabular data of powder spreading speed | Extract data from physical based model | NN | Powder Spreading Quality | 35 | [96] |
| 2019 | PBF | Greyscale image | Cropping, pixelization | CNN | Defective conditions | 1,200 | [99] |
| 2018 | PBF | Digital image and design file | Image preprocessing and pixelization. Extract geometry information from STL file. | CNN | Process anomaly detection | 10,071 | [87] |
| 2018 | PBF | Tabular data of process conditions, Graphic data of CT scans, and DSLR images | Binarization, complementation, noise reduction, spectral graph theory, multifractal analysis | SVM, LDA, k-NN, Ensemble (bagged trees), NN | Process conditions | 3,132 | [52] |
| 2017 | PBF | Microscope image | Adjusting contrast level, denoising, segmentation, background suppression, feature selection by K-means clustering, interest point localization, scale-invariant feature transform | SVM | Powder types | 282 | [40] |
| 2018 | DED | Tabular data for part geometries | None mentioned | ANN | Optimal processing parameters | 120 | [94] |
| 2020 | VP | Matrix and signal data | Principal component analysis, sensitivity analysis | Shallow NN | Energy consumption | 657 | [93] |
| 2020 | VP | 3D data from printed structures | Transformation (Microscope to binary) | NN | Digital masks with optimized | 900 | [54] |

| | | | | | | | |
|------|----|--|---|----|-----------------------------------|-------------|------|
| | | | image) and image rotation | | light exposure dose | | |
| 2019 | VP | Tabular data of process parameters | None mentioned | NN | Success or Failure wrt. Speed | 6,000 + 180 | [89] |
| 2018 | MJ | Tabular data of image extracted features | Extract features from real-time monitoring images | NN | Process parameter (voltage level) | 800 | [97] |

3.2.2. Build and toolpath planning

Deposition strategies in AM influence product structure and, subsequently, its properties and performance. ML models can ignore underlying physics and relate path strategies with structure, property, and performance characteristics. It helps avoid certain toolpaths that are prone to failure with respect to these characteristics before a part is completely printed. The reviewed references in build and toolpath planning are summarized in Table 7.

Among the available data types at this stage, 3D data is found to be the predominant input for build and toolpath predictions. This data type can be collected from different phases of AM processes such as design [102] and post-process [103]. In a relevant application, 3D data from sliced lattice models is used in an SVM model to predict optimal filling paths for lattice structures [104]. X-ray computed tomography (XCT) generates 3D volumes of parts that are sliced to 2D, cropped, and de-noised before being fed to a CNN for the prediction of build orientation [103]. Features of 3D junction geometries are employed in a NN to find optimal path length value to avoid material deficit [102]. Tabular data is another type that is used to predict build and tool path characteristics. The example applications that use tabular data have process parameters as their source. For example, tabular data of process parameters is also used to determine desired printing pattern [105]. In another example, a feed-forward NN improved the quality of the connection between two consecutively deposited paths using process parameters as inputs in tabular form [32].

Table 7: ML applications at the process stage of AM – Build and toolpath characteristics

| Process Characteristics - Build and toolpath characteristics | | | | | | | |
|---|-----------|--|---|---------------------|---------------------------------|------------------|------------|
| Year | AM | Data | Data Handling | ML Algorithm | Application | Instances | Ref |
| 2020 | ME | Tabular data for processing parameters | None mentioned | ANN | Connection status between paths | 400 | [32] |
| 2020 | PBF | The raw XCT-generated 3D volumes | Sliced to 2D, then cropped and finally denoised | 3D-ResNET | Build orientation | 192 | [103] |
| 2020 | DED | 3D data from junction geometries | Geometric features extraction | NN | Path length value to avoid | 63 | [102] |

| | | | | | | | |
|------|--------------------------|--|---|-----|--|-------|-------|
| | | | | | material deficit | | |
| 2021 | General AM | 3D data of polygons from sliced lattice models | Geometry-based feature value extraction | SVM | Double-spiral, Counter-parallel offset, or Straight skeleton | 1,500 | [104] |
| 2019 | General AM (robotic arm) | Tabular data of system parameters | None mentioned | NN | desired printed pattern | 200 | [105] |

3.2.3. Surrogate modeling of the melt pool

As a sub-category of data-driven approaches, ML models are perfect candidates to serve as surrogates of complex multi-phase and multi-physics process simulations in AM. There are multiple ways to develop these surrogate models. ML models can be trained completely on experimental data where no simulation results are needed. In some cases, process simulations are used to inform ML models partially or completely. There are also instances where physical knowledge is incorporated in ML models at the structural level i.e. a physics-based error function to learn model parameters [106]. Melt pool characteristics are of key interest in surrogate modeling. Accurate prediction of these characteristics can help pick an adaptive approach to process control. There has been extensive interest in predicting the thermal distribution of melt pools as this can be a good representative of future structures and properties [107-111]. Melt pool topography [31, 38, 112, 113] and other characteristics [48, 49] have been the target of some research works. The related research articles are briefed in Table 8.

Tabular, graphic, and 3D data types are seen as potential inputs for ML-based surrogate modeling. Tabular data comprising process [31], material [106], geometry [108], and temperature parameters [114] is widely used to model melt pool characteristics. It roughly accounts for 60% of all data types employed for surrogate modeling. Finite element (FE) generated tabular data is a clear addition to existing trends at this stage [107, 109, 111, 112]. Specific applications of tabular data include melt pool geometry and thermal distribution prediction. The thermal history of a metal additive manufacturing (MAM) process is computed using an unsupervised clustering technique applied to input geometry and scan parameters in tabular form [114]. Graphic data from digital cameras and microscopes is found to be the second-best choice for melt pool modeling. Melt pool is monitored through these sensors and the resulting images/videos are used in ML models to predict characteristics of interest. Digital images from simulations are used in a CNN to predict melting conditions in a PBF process [48]. 3D data is also used in ML models at this stage of AM process flow. Features of 3D data are employed in deep models to predict thermal field of a wire DED process [110].

Table 8: ML applications at the process stage of AM – Surrogate modeling of the melt pool

Process Characteristics - Surrogate modeling of the melt pool

| Year | AM | Data | Data Handling | ML Algorithm | Application | Instances | Ref |
|------|------------|--|--|-----------------------|--|-----------|-------|
| 2020 | ME | Tabular data of part geometry | Extract features from tabular data (relative distances from the cooling surfaces, from the heat sources, and a set of deposition times influencing the thermal behavior) | ANN | Thermal distribution | 11 | [108] |
| 2021 | PBF | Tabular data of FE results | Extract features from tabular data ([Node, X, Y, Z, Temperature]) | GPR | Thermal distribution | Not clear | [107] |
| 2021 | PBF | Tabular data for build parameters and microscope image | Pixelization, gray-scale, binary filter, connected component labeling algorithm, extraction of material descriptors, label encoding, normalization, cropping | RF, deep NN, SVR, GB | Creep rate | 512 | [49] |
| 2021 | PBF | Tabular data of FE results | Extract features from the tool path | RNN | Thermal distribution | 340 | [111] |
| 2021 | PBF or DED | Tabular data of process, material, and geometry | None mentioned | NN | Temperature, Melt pool dynamics, and dimensions, Cooling rates | Not clear | [106] |
| 2020 | PBF | Tabular data of FE results | Generate a matrix to store the laser scanning pattern from the tabular data | RNN+ANN | Thermal distribution | 100 | [109] |
| 2020 | PBF | Digital image of melt pool | None mentioned | Polynomial regression | Melt pool area | 20,902 | [38] |
| 2020 | PBF | Tabular data of simulations | Extract features from tabular data | GPR | Melt pool geometry | 200 | [112] |
| 2020 | PBF | Digital image from the simulation | Resizing, pixelization, normalization | CNN | Melting conditions | 1,412 | [48] |

| | | | | | | | |
|------|-----|--|--|-----------------------|---|-----------|-------|
| 2020 | PBF | Tabular data for thermal history | Dimensionality reduction | k-means clustering | Zoned maps of thermal history | Not clear | [114] |
| 2019 | PBF | Tabular data for processing parameters and material properties | Sensitivity analysis | RF, SVM | Melt pool geometry | 472 | [31] |
| 2018 | PBF | Graphic data of in-situ videos from the melt pool | Center Cropping | CNN | Track width, Track width standard deviation, and Track continuity | 870 | [113] |
| 2021 | DED | 3D Matrix, Model, representation | Extract features from numerical simulation | RNN, deep NN, and CNN | Thermal field | 15,000+ | [110] |

3.3. Product characteristics

The last category of ML applications in AM is related to product characteristics. There are numerous parameters of interest that relate to printed parts. These can be classified with respect to product structure and properties. In AM, quality and business constraints require parts to be checked at both macro and micro levels. Macro-level deals with geometrical and visual aspects, whereas micro-level deals with anomalies and defects in the microstructure of printed parts. ML models are also employed to predict products' properties and other aspects (cost, time, life, etc.). These applications are discussed in detail in the subsequent sections.

ML applications concerning AM product characteristics offer unique benefits as well. Data-driven models for macro and micro characteristics can replace labor-intensive tasks such as measurements, characterizations, and microstructural evaluations. Given the complex nature of such characteristics (e.g., the microstructure of composite material systems), reliable physics-driven models are often impossible to develop. ML models on the other hand can extract key input-output relations in the context of given applications. This can relieve practitioners from expensive alternatives to deduct such complex characteristics. This applies to both macro (e.g., deviations in the printed parts) and micro (e.g., defects in the microstructure) characteristics of AM products. Similar to design and process spaces, property spaces can be established to guide designers and operators. Regions of desired properties can be linked to either design or process parameters. All of this is possible from base ML models correctly capturing the geometry, microstructural, and property traits of AM printed parts.

3.3.1. Macro level

The majority of ML applications in AM deal with product characteristics at the macro or micro level. Macro-level targets usually concern with the visual characteristics of printed products and

are often the first category of quality metrics against which products are checked. Geometric dimensions and visual defects are two main targets in this regard. These can relate to the macro characteristics of a single path, single layer, or multiple layers. As a result, several works focused on predicting geometry deviations in AM printed parts [115-119]. Inspecting visual defects has been a common way to judge product quality and is used in several ML models with tabular [63], graphic [120, 121], spectrum [74] data as inputs. Regression-based ML models are a popular choice to determine the exact geometry of products in AM [122-124]. There are cases where the domain or expert-defined labels are used to make decisions on the macro-level quality of AM parts [125, 126]. The reviewed references for macro structure characteristics are outlined in Table 9.

Graphic, tabular, and 3D data types have more or less similar proportions for macro level targets with each representing 36%, 30%, and 30% of datasets respectively. The majority of graphic data come from digital cameras that capture images at different stages of AM process [115, 119, 120, 127]. Microscopes [41] and infrared cameras [128] are also employed to capture AM process. These images are then used to predict a range of macro characteristics such as dimensional variations and visual defects. Digital images of the process are employed in different ML models to detect visual defects in a ME process [120]. Simulation and camera images are used to detect real-time cyber-attacks resulting in malicious defects using ML models [129]. Tabular data from design [122] and process [123] stages is used to model macro structure anomalies. Tabular data alongside a design file is used in a CNN to predict visual flaws of ME printed parts [90]. 3D data from design files and point clouds is also used to predict macro structure targets [118, 130-132]. In one application, spectrum data of acoustic signals is used to predict geometric defects in PBF printed parts [74].

Table 9: ML applications at the product stage of AM – Macro structure characteristics

| Product Characteristics - Macro Structure Characteristics | | | | | | | |
|--|-----------|------------------------------------|---|--|---|------------------|------------|
| Year | AM | Data | Data Handling | ML Algorithm | Application | Instances | Ref |
| 2021 | ME | Digital image of layer | Pixelization | CNN | Over extrusion and under extrusion (macro) | 1,400 | [127] |
| 2021 | ME | Tabular data of process parameters | Randomization | GPR, SVM | Geometric deviation | 288 | [117] |
| 2021 | ME | Digital image of layer | Image preprocessing and filtering | NN, GB, SVM, cluster charting | Visual defects during process (Macro scale) | 6,000 | [120] |
| 2021 | ME | Design file | Generate 3D point cloud based on design file and preprocessing to | Bagging of Trees, GB, RF, k-NN, and Linear SVM | Geometrical defect detection | 50 | [130] |

| | | | | | | | |
|------|----|---|---|--|---|-----------|-------|
| | | | extract critical features | | | | |
| 2021 | ME | Digital image from coaxial camera | Editing and normalizing the data and then pixelization | CNN | Geometrical anomalies | Not clear | [119] |
| 2020 | ME | Digital image cross-sectional images | Image processing to extract geometry data | ANN | Geometry deformation | 270 | [115] |
| 2020 | ME | Vector data in tabular format | Extract critical features based on domain knowledge | Multi-linear regression, k-NN, SVM, RF, NN, and DL | Printability in shape | 145 | [133] |
| 2020 | ME | Graphic data from GoPro camera images | Transformations (resize and reshape) | CNN | Part Success or Failure based on different macro-structural defects | 200 | [125] |
| 2020 | ME | Digital Image of printed part | Pixelization, grey scaling | CNN | Warping defect | 674 | [37] |
| 2019 | ME | Tabular data and design file | The features are generated from the STL file, toolpath locations and parameter settings | CNN + GB | Visual flaws | 144 | [90] |
| 2018 | ME | 3D point cloud | Downsampling | Self-organizing map | Geometric accuracy | 12 | [118] |
| 2018 | ME | 3D data from as-designed shapes | None mentioned | GPR | Shape variations | Not clear | [131] |
| 2017 | ME | Graphic data from simulations and camera-based images | Image segmentation, feature extraction | k-NN, RF | Malicious defect from cyber attack | 3,887 | [129] |

| | | | | | | | |
|------|-----|---|--|---|--|-------|-------|
| 2021 | PBF | Tabular data of part location and surface orientation | None mentioned | DT | Surface height | 648 | [122] |
| 2021 | PBF | Tabular data of microstructure | Min-max normalization | Self-organizing map | Elongation | 360 | [134] |
| 2021 | PBF | Design file and tabular data | Voxelization and hot encoding for tabular data | CNN+ANN | Geometry (shape and dimensions) Functional (visual defects) characteristics | 245 | [60] |
| 2021 | PBF | Video of process | Get images from Video and pixelization | CNN | Visual powder bed defects | 8,514 | [121] |
| 2021 | PBF | Design file and tabular data | Convert the design file to the sparse matrix | Sparse CNN + ANN | Visual defects | 245 | [63] |
| 2021 | PBF | Spectrum data of acoustic signals | Projecting the original data matrix into a lower-dimensional space | LDA | Geometry defects (missing struts) in lattice structures | 210 | [74] |
| 2021 | PBF | 3D CAD models | Voxelization, feature learning | Auto encoder-generative adversarial network | Cell structure manufacturability | 50 | [132] |
| 2020 | PBF | Tabular data for processing parameters and tool path, and 3D model from simulations | Voxelization | k-NN, CNN | Geometry | 9,000 | [123] |
| 2020 | PBF | Thermal images of melt pool | Resizing and pixelization | Depthwise-separable CNN | Delamination and splatter | 4,314 | [128] |

| | | | | | | | |
|------|------------|---|--|--|---|-----------|-------|
| 2020 | PBF | Melt pool images | The real-time function of sensor data | K-Clustering | Defect detection from melt pool size monitoring | 2,000+ | [135] |
| 2020 | PBF | Tabular data of rule metrics | The features are calculated by the defined design rules | Singular-value decomposition and Euclidean distance measurement techniques | Manufacturability in shape | Not clear | [136] |
| 2020 | PBF | 3D Matrix | Feature extraction | SVM, RF | Visual defects | 1,250 | [137] |
| 2018 | PBF | Microscopic images of surface | Pixelization, denoising, gray-scaling, brightness threshold, statistical features, PCA | RF | Detect fused track boundaries | Not clear | [41] |
| 2018 | PBF | Tabular data of process parameters | Critical feature selection | LDA | Cracking caused by delamination | 47 | [138] |
| 2018 | PBF | Tabular data of part and process parameters | Elbow method and K means clusters | RF | Good or Defective part based on inspection of defects | 1,000 | [126] |
| 2021 | DED | Tabular data of weld bead | Signal denoising filter, curve fitting process | SVM | Bead profile | Not clear | [124] |
| 2020 | DED | 3D point cloud data, Tabular data of extracted features | Filtering, segmentation, surface-to-point distance calculation, point clustering | SVM, KNN, GPR, DT, NB, ANN, RF and AdaBoost | Surface defect identification | Not clear | [139] |
| 2020 | General AM | Thermal measurement signals in graphic form | Feature selection | LR, GPR, RF, ANN, SVM | Defect geometries | 600 | [33] |
| 2020 | General AM | 3D data for lattice infill pattern | Surface 3D coordinates extraction | NN | Symmetrical deviation | 4,800 | [116] |

| | | | | | | | |
|------|------------|----------------|--------------------------------|---------------------------|-----------------------|----|-------|
| | | | | | surface coordinates | | |
| 2017 | General AM | 3D point cloud | Spectral graph theory approach | k-NN, NN, NB, SVM, and DT | Dimensional Variation | 12 | [140] |

3.3.2. Micro level

Porosity, lack of fusion, micro-cracks, and balling are some of the most frequent defects in AM printed parts' microstructure. Porosity has been a common target of ML models in this category. In this regard, ML models have been employed for a range of tasks including pore detection [34, 42, 44, 68, 141-144], pore classification [50, 57, 67, 145, 146], and pore size prediction [93]. Some ML applications also deal with the classification of specific microstructure types [51]. Lack of fusion and balling defects have also been considered as targets in several ML models. The references are summarized in Table 10.

Datasets for predicting micro characteristics fall in all four general types introduced in this survey. Graphic data stands out as a clear choice when it comes to predicting micro characteristics of AM parts. 59% of the available data in this section belongs to the graphic category. The sources of graphic data are found to be diverse as well. Digital cameras [147, 148], thermal cameras [42, 44], and microscopes [145, 149] are used for in-process graphic data generation to predict micro characteristics. This data type is mainly used to predict microstructural defects. Graphic data of melt pool images is used to classify balling, keyholing, porosity, under-melting, and desirable conditions in an SVM classifier [150]. Layer-wise images of a PBF process are cropped and used in a deep learning model to distinguish the lack of fusion defects from standard cases [151]. 3D, spectrum and tabular data types are found to be relatively less common and account for 10%, 14%, and 17% of data share respectively. XCT-based 3D data and microscopic images are used to cluster different types of pores [50]. Spectrum data from acoustic signals is used in a NN to predict porosity-based quality (poor, medium, high) of PBF printed parts [67]. A few other applications of spectrum data also employ acoustic sensor data in ML models and are listed in Table 10. Tabular data from diverse sources can be found to predict porosity [143, 152] and grain growth in AM [153].

Table 10: ML applications at the product stage of AM – Micro structure characteristics

| Product Characteristics - Micro Structure Characteristics | | | | | | | |
|--|-----------|--|---|---------------------|---|------------------|------------|
| Year | AM | Data | Data Handling | ML Algorithm | Application | Instances | Ref |
| 2021 | ME | 3D data of as-designed and as-built models | Visual extraction of geometric features | NN, k-NN | Flow resistivity, Porosity, Tortuosity, Thermal length, Viscus length, and Permeability | 500 | [64] |

| | | | | | | | |
|------|-----|--|--|-----------------------------|----------------------------------|-----------|-------|
| 2020 | ME | Tabular data of process parameters | Normalization | NN | Bond quality, porosity | 1,525 | [152] |
| 2021 | PBF | Microscope image of bead cross section | Image preprocessing to eliminate the noise and highlight defects | RF | Ex-situ porosity | 195 | [149] |
| 2021 | PBF | Graphic data from layer-wise images | Image Cropping | CNN, NN | Lack of fusion or Normal | Not clear | [151] |
| 2021 | PBF | Layer-wise part data in tabular form | Extract physical effects from machine setting parameters | LR, GPR, SVR | porosity | 549 | [143] |
| 2020 | PBF | Infrared image thermal profile, X-ray image melt pool dynamics, and porosity | Time window, denoising, manual selection of statistical features | LR, RF, GPR, GB | Porosity | 15 | [42] |
| 2020 | PBF | Infrared image of surface temperature | Denoising | K-means | Porosity | Not clear | [44] |
| 2020 | PBF | Photodiode sensing system signal | Graph Fourier Transform features, time window, sensitivity analysis, PCA | k-NN, SVM, DT, LDA, NN, SVM | Porosity | 440 | [34] |
| 2020 | PBF | 3D X-ray tomography and microscope image of parts | Voxelization/pixelization, segmentation, statistical features, scaling | K-means | Pore types | 3,142 | [50] |
| 2020 | PBF | 3D X-ray tomography of parts | Voxelization, cropping, normalization | CNN | Porosity segmentation | 1,100 | [62] |
| 2020 | PBF | Digital image of build plate | Get texture from images (The quantified powder bed | NB classification | Microstructural defects - online | 205 | [147] |

| | | | | | | | |
|------|-----------|--|---|------------------|--|-----------|-------|
| | | | 3D-DIC data, alongside ex situ identification of physical defect locations) | | | | |
| 2020 | PBF | Graphic data from electron-optical images | Miscellaneous image processing techniques | CNN | Good, porous, or bulging | 16,000 | [154] |
| 2020 | PBF | Graphic data of optical tomography images | Cropping and background removal | k-means and k-NN | Drift (melt pool hotspots) or No-drift at the layer-level | 240 | [155] |
| 2020 | Laser PBF | IR images of surface temperature distribution | Signal denoising filter | Neural Learning | Pore detection | 100 | [144] |
| 2020 | PBF | Pyrometer signal from build chamber, XCT of part | Image preprocessing | K-d tree | Porosity | 10,597 | [141] |
| 2020 | PBF | Tabular data of energy and temperature | N/A | NN | Nanoparticle sizes | Not clear | [93] |
| 2020 | Laser PBF | Layer-wise Images | Texture analysis | hybrid deep NN | Porosity defect | 3,000+ | [55] |
| 2019 | DED | Digital camera and spectrometer for melt pool monitoring | Graph Kronecker product analysis | SVM | The average length of lack-of-fusion defects | Not clear | [148] |
| 2019 | PBF | Graphic data of melt pool images | Bag of Words with SIFT features | SVM | Balling, Severe Keyholing, Keyhole Porosity, Under-Melting, or Desirable | 24,385 | [150] |

| | | | | | | | |
|------|-----------|---|--|-----------------------------------|---|-----------|-------|
| 2019 | PBF | Acoustic emission signal | Extract wavelet spectrograms from signals | Reinforcement learning | Porosity | 180 | [68] |
| 2019 | Laser PBF | Acoustic emission signal | Spectral Convolution Neural Network | DL | Real-time Printing Quality based on porosity (3 levels) | Not clear | [146] |
| 2018 | PBF | Digital Images of powder bed | Apply filters to images, Selection of training images | Bag-of-keypoints (unsupervised) | Anomaly detection (micro) | 2,400+ | [156] |
| 2018 | PBF | Spectrum data of acoustic signals | Wavelet packet transform | NN | Porosity-based quality (Poor, Medium, High) | 600 | [67] |
| 2020 | DED | Tabular data of thermal gradient, crystal orientation, and Marangoni effect | Crystal orientation reconstruction and into a 2D vector | ANN | Regression of the competitive grain growth behavior | 50 | [153] |
| 2020 | DED | Digital image of the microstructure | Cropping, random horizontal, vertical flipping, PCA | CNN, SVM | Microstructures classification | 1,801 | [51] |
| 2020 | DED | Pyrometer images of melt pool | Pixelization | CNN | Porosity | 1,557 | [142] |
| 2020 | DED | Microscopic image of part | Image denoising, smoothing, and unblurring to highlight the areas with pores | RF | Porosity classification | 6,000+ | [145] |
| 2018 | DED | Thermal images of melt pool | Functional principal component analysis | KNN | Porosity classification | 60 | [57] |
| 2020 | BJ | CT images of defect data | Reconstruction and feature extraction | Gaussian mixture model clustering | Pore evolution | Not clear | [157] |

3.3.3. Mechanical properties and other characteristics

Properties lie at the end of the process-structure-property chain of AM process flow. The future performance of products is based on the underlying properties. Mechanical properties have been a target of immense interest for ML models in AM [47, 56, 158-161]. The representative examples are tensile strength [43, 45, 61, 69, 162], elongation [161], hardness [163], fatigue [30, 164, 165], and surface roughness [73, 166-168]. Some works also focus on ML-based modeling of residual stress [169, 170] and density [72, 168, 171] in printed parts. Quality metrics are based on specific properties and serve as labels in ML models [172, 173]. Finally, miscellaneous applications where business-related characteristics of printing cost and time are computed using ML modes can be found in AM literature [174]. Table 11 summarizes the existing articles that fall in this category.

ML-based prediction of mechanical properties has a clear winner in tabular data with 61% share of all datasets. Tabular data of material [175], deposition [45], process [162], fatigue [164], and geometry [160] parameters is frequently used to model diverse mechanical properties. Process parameters have a major share among all sources of tabular data at this stage [159, 165, 167, 168, 172, 176, 177]. For instance, process parameters alongside mechanical properties in tabular form are used in an adaptive neuro-fuzzy inference system to estimate the fatigue life of AM printed metal parts [30]. 17% of the available data falls in the graphic category making it an attractive choice after tabular data. Similar to tabular data, graphic data has been used to predict a range of mechanical properties. Digital cameras are a common source of graphic data at this stage [56, 178, 179]. Spectrum and 3D data types are found to have an equal contribution of 11%. Spectrum data from optical signals was used to classify density levels (low, medium, and high) of PBF printed parts in an SVM classifier [72]. 3D data of microstructures is used in a CNN model to predict the effective yield strength of parts manufactured by MAM processes [61]. Spectrum and 3D data types are also employed to determine several other properties as referenced in Table 11.

Table 11: ML applications at product stage of AM – Mechanical properties and other characteristics

| Product Characteristics - Mechanical Properties and other characteristics | | | | | | | |
|--|-----------|--|--|---------------------|---|------------------|------------|
| Year | AM | Data | Data Handling | ML Algorithm | Application | Instances | Ref |
| 2021 | ME | Graphic data for X-sectional images | Sobel and Canny edge finding algorithm, Hough line finding algorithm | NN | Mechanical compression curve values | 250 | [56] |
| 2021 | ME | Tabular data of 3D printed carbon fiber composites | None mentioned | GPR | Mechanical properties | 30 | [161] |
| 2020 | ME | Tabular data of three material parameters | Normalization | NN | Force displacement curve (FDC) error difference | 20,000 | [175] |

| | | | | | | | |
|------------|-----|---|--|---|--|-----------|-------|
| what a2020 | ME | Tabular data for infill patterns | None mentioned | ensemble learning with k-NN, SVM, etc. | Flexural strength | 162 | [45] |
| 2020 | ME | Tabular data of 3D printing parameters | None mentioned | Lasso, XGBoost, and SVR | Tensile strength | 192 | [162] |
| 2019 | ME | Thermocouple and accelerometer signals | Statistical features | RF, SVR, ridge, lasso | Surface roughness | 27 | [73] |
| 2019 | ME | Tabular data of process parameters | Critical feature selection | Knowledge-based ANN | Part quality | Not clear | [176] |
| 2019 | ME | Sensor signal of temperature and vibrations | Feature extraction and feature selection using random forest | Classification and regression trees, Random vector functional link network, Ridge regression, SVR, RF, AdaBoost | Surface roughness | 81 | [166] |
| 2019 | ME | IR sensor, thermocouple, and accelerometer signals, and tabular data for processing parameters, mechanical properties and part geometries | Statistical features (mean, std dev and RMS) | LSTM+ANN | Tensile strength | 144 | [43] |
| 2019 | ME | 3D Design file | Voxelization | CNN | Part mass, support mass and build time | 72,000 | [180] |
| 2021 | PBF | Tabular data of process parameters | Standard Scaler method | ANN | Density ratio and surface roughness. | 2,048 | [177] |
| 2019 | PBF | Tabular data of build orientation | Feature filter and selection | GB, DT, AdaBoost Regressor | Mechanical properties | 434 | [158] |

| | | | | | | | |
|------|-----------|--|-----------------------------------|--|---|-----------|-------|
| 2021 | PBF | Tabular data of AM fatigue parameters | None mentioned | NN, RF | Fatigue life cycles | 3*500 | [164] |
| 2021 | PBF | Tabular data of process parameters | Feature selection and filtering | GB | Magnetic characteristics (iron loss and permeability) | 780 | [159] |
| 2021 | PBF | Tabular data of process parameters | Chi-merge and Statistical ANOVA | NB | Density and Hardness | 181 | [171] |
| 2021 | PBF | Tabular data of process parameters | None mentioned | ANN, RF, SVM | Fatigue life | 600 | [165] |
| 2021 | PBF | Tabular data of process parameters | None mentioned | Multivariate Gaussian process | Density and surface roughness | 15 | [168] |
| 2021 | PBF | Tabular data of production parameters | Feature scaling | LR, SVM, DT, RF, XGBoost, MLP, k-NN, LDA | Quality repeatability based on mechanical properties | 251 | [181] |
| 2020 | PBF | Spectrum data from optical signals | Statistical features extraction | SVM, MLP | Density classes (low medium high) | 4,402 | [72] |
| 2020 | Laser PBF | Vector (tabular format) | Scaling | Hybrid Bayesian network | Part quality characteristics | 300+ | [182] |
| 2020 | PBF | 3D data of lattice structures | Image windowing, image transforms | CNN | Deformation work | 3*4,095 | [183] |
| 2020 | PBF | Tabular data for part geometries | None mentioned | ANN | Mechanical properties | 15 | [160] |
| 2020 | Laser PBF | Tabular data of process parameters | Kernel Density Estimation | Stochastic optimization | Failure wrt ultimate strength, yield strength, and elongation | Not clear | [172] |
| 2020 | Laser PBF | Vector and images of thermal histories | Dimensionality Reduction | SVR | Printing quality based on energy efficiency and shape factor | 100 | [173] |

| | | | | | | | |
|------|------------|--|--|---------------------------------------|------------------------------------|-----------|-------|
| 2020 | Laser PBF | Videos (graphical format) of met pool | None mentioned | MLR, k-NN, SVM, RF, NN and DL | Quality assurance | 1,000+ | [178] |
| 2020 | PBF | Tabular data of process parameters | Normalization | GPR | Relative density | 82 | [184] |
| 2020 | PBF or DED | 3D data of microstructure | Morphological and crystallographic feature extraction, feature standardization | CNN | Effective Yield Strength | 7,680 | [61] |
| 2019 | PBF | Tabular data for processing parameters and mechanical properties | None mentioned | Adaptive neuro-fuzzy inference system | Fatigue life | 139 | [30] |
| 2019 | PBF | Photodiode sensing system signal | Downsampling, singular value decomposition | Gaussian mixture model | Ultimate tensile strength | 49 | [69] |
| 2021 | DED | Tabular data of FE results | Extract features from tabular data (thermal history and spatial coordinates of the elements/nodes) | ANN | Residual stress distribution | Not clear | [169] |
| 2021 | DED | Tabular data of process parameters | None mentioned | Adaptive neuro-fuzzy inference system | Surface roughness | 27 | [167] |
| 2021 | DED | Tabular data of process parameters | ANOVA, correlation matrix | LR, quadratic polynomial regression | Hardness | Not clear | [163] |
| 2020 | DED | Tabular data of properties and process variables | Feature importance and hierarchical influence | RF, NN | Residual stresses and Delamination | 243 | [170] |
| 2020 | VP | Videos of part fabrication | Get images from Video | CNN+LSTM | Part quality | 1,832 | [179] |

| | | | | | | | |
|------|------------|----------------------------------|--|-----|-----------------------|--------|-------|
| 2018 | General AM | Microscope image from simulation | Segmentation, pixelization | CNN | Mechanical properties | 10,000 | [47] |
| 2018 | General AM | 3D data of CAD models | Extraction of geometric/non-geometric features from G-code | LR | Printing cost | 135 | [174] |

4. Public AM databases

This section briefly introduces the major existing public databases containing datasets for AM studies.

- National Institute of Standards and Technology (NIST)

The engineering laboratory of NIST offers a database system named the additive manufacturing materials database (AMMD). It is built to provide open data access and material data sharing to the AM community. Their datasets include testing data, machine data, part design data, build part process data, and material data. Though their database is designed for all seven typical AM processes, only PBF data is collected at the current stage [185]. [This database can be used to applications such as prediction of mechanical performance, material selections in AM, and real-time quality monitoring.](#)

- Mendeley Data

Mendeley Data is a cloud-based repository where users can store and search research data. It is hosted by Elsevier. As Elsevier is a publishing company specializing in science and technology, they collect all the uploaded datasets for the published research articles from their flagship journals. Users can also store their datasets by uploading files of any type. By researching “additive manufacturing” in their search tool and narrowing down the data types to “dataset”, 2020 datasets are found at the current stage. Users can easily download the datasets and export the citations as well [186].

- DOE Data Explorer

DOE Data Explorer is supported by the U.S. Department of Energy and the Office of Scientific and Technical Information. It is a search tool for finding their supported projects and included dataset records. By searching “additive manufacturing”, 67 datasets can be found. All the links to download datasets are provided as well [187].

- DATA.GOV

DATA.GOV is a searching tool founded by the U.S. government to support finding data and research. Users can narrow down the target data via the search engine. At the time of this writing, 930 datasets can be found under the term ‘additive manufacturing. However, some of them are solely brief introductions to the research project. Users should filter the search and find the potential useful information and datasets for their research [188].

- DataCite

DataCite is a non-profit organization that provides persistent identifiers (DOIs) for research data and other research output. Their members include data centers, libraries, research universities, and governments over the world. They also offer the service to look for reliable data shared by the research community and reuse the data for other research studies. By searching “additive manufacturing” and filter the works to “dataset” type, 225 downloadable datasets are found [189].

- **IEEEDataPort**

IEEEDataPort is developed by IEEE. This platform provides free uploads of datasets and free downloads of the open-access datasets shared by IEEE members and users. There are over 1,500 datasets and most of them are in the artificial intelligence, machine learning, computer vision, and image processing aspects. By searching “additive manufacturing” in their dataset library, only one dataset is available. However, more datasets are expected in the coming future [190].

5. Existing data management approaches in AM

There are countless AM studies in the literature, and they have been continuously published every year. This trend has gained momentum in recent years resulting in enormous data for ML applications in AM. As a result, the management, storage, and accessibility of this data have become a challenge.

For most data-search engines, they use keywords as the first step to filter the search. After that, the most common filter options are data types including datasets, documents, images, videos, etc., and sources including universities over the world, journals, etc. Finally, users should use their knowledge to open the results in order and read the descriptions to determine whether the selected result is suitable for their needs. As most data-search engines include datasets from various research areas, they cannot offer specific filter options for AM. In this case, finding the target data is still difficult for AM researchers.

There are also some systematic studies to establish data management strategies specifically for AM. One of the representative researches is from Liu et al [154]. They proposed a cloud-based digital twin-enabled data management framework for MAM. It contains AM data in different product lifecycle stages including product design, quality measurement, process planning, manufacturing, and post-processing. Data items of each lifecycle stage are listed with sub-categories, common measurement methods, and data format. Lu et al. [191] provided a similar approach on the collaborative AM data management system. Their data management system aims to establish the correlations between processes, materials, and parts. A web interface is available to store, explore and download data which has been introduced in Section 4.1. In their database, material data, machine data, design data, process data, and test data are collected. Both approaches provide a well-organized data structure to store and manage data. However, none of them is sufficient in data searching and sharing. They are more appropriate for institutes, industries, or organizations to internally store and manage data. When sharing to the public, datasets are expected to be easy to access. In their current version of the data management system, it is either hard for external users to understand the system or hard to find suitable filters to search datasets of interest. The existing effort on this topic is very limited, which motivates more future interests and developments in AM data management systems.

6. Limitations and challenges

The above sections presented the existing ML applications in AM with details of their data types, data handling approach, and applications. It can be seen that the current data in ML-assisted AM studies still has some problems including quantity and diversity of datasets, the guarantee of the manual label accuracy, reproducibility and standardization of data-driven research, and the need for a simple, easy-access, and systematic database. The following section will provide a detailed discussion on the [limitations and challenges](#).

6.1. Quantity and diversity of datasets

Diversity and the data size are extremely important characteristics in a well-developed dataset that can directly affect the quality of the dataset and the performance of the ML model. Figure 8 shows the frequency of the number of instances used in the research articles reviewed in this paper. Most of the existing research falls in the group of 100-1,000 instances, and only 13 out of 144 research articles include 10,000+ instances in their datasets. Determination of whether the size of the dataset is large enough depends on the applications and the complexity of the ML model. However, the evaluation for data size and diversity is absent in most studies, and the datasets for some studies are limited in quantity and diversity.

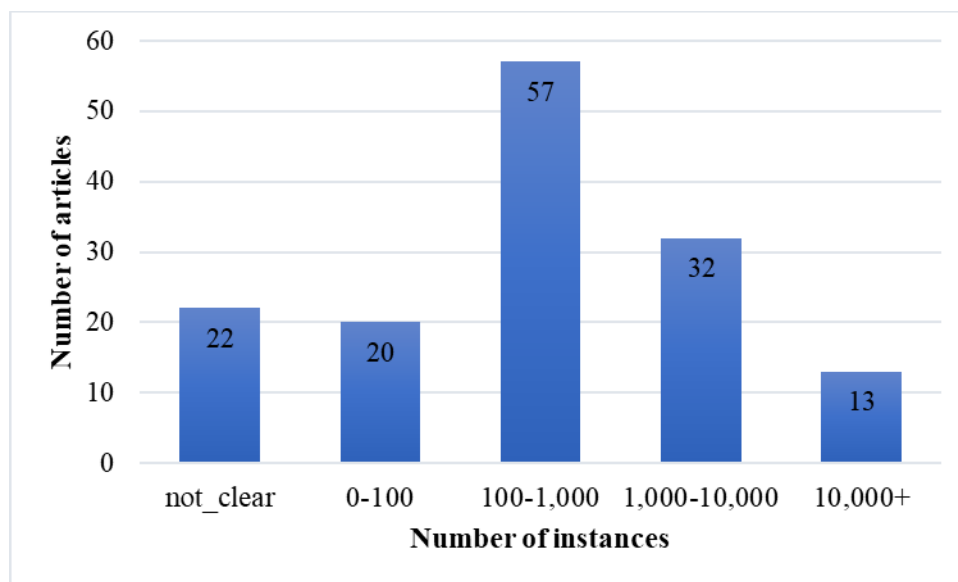


Figure 8: Frequency of the number of instances used in the reviewed research articles

The ideal method to identify the size of data is to generate a learning curve for the model performance on datasets [192]. The required number of data size can be obtained when the learning curve reaches the saturation point. To make it simple, there are some common rules from the ML community to identify the ideal size of the dataset. These rules are generally a factor of certain characteristics of the prediction problem. For example, some researchers indicate that the data size needs to be at least 50 to 1000 times the number of prediction classes [193]. Another rule states that the data size needs to be at least 10 to 100 times the number of the features [194, 195]. The most common method is to include at least 10 times the number of weights in the network if neural network models are used [196, 197]. However, a later study [192] states that the factor of 10 is insufficient, and they conclude that the data size needs to be at least 27 to 31 times the number of weights in the network. Even though the data size may also vary on the different applications, those common rules can provide a general idea of how many samples are enough for their studies.

6.2. Guarantee of the manual label accuracy

Another limitation of the current ML-assisted AM studies is to ensure the accuracy of the label/ground truth and how to increase the ease of the manual labeling process. At the current stage, most existing studies utilize the supervised learning approach which requires labeling. The performance of ML models is always restricted by the available data. As introduced in Section 3, the most common ML applications in AM include design-related, process-related, and product-related applications. The label/ground truth can also be categorized into experimental results, computational results, and manual labeling. In terms of label accuracy, the experimental results such as mechanical properties and computational results such as thermal distribution obtained from the finite element method (FEM) model are more promising. Manual labeling such as the location of pores, visual defects determination, mark of the failure area, etc. is less reliable. Mislabeling errors may occur during the labeling process. Most of the existing studies reviewed in this paper assume manually labeled data to be authentic. There is no such annotation tool developed for labeling AM data and ensure the quality of the manually labeled data. Some of them may acquire manual votes from several AM experts. Moreover, the manual labeling process is extremely time-consuming.

There are some annotation tools available in the literature for computer vision and natural language processing [198-200]. The annotation tool can help users on labeling maps, attributes, classes, etc. For example, for images and videos, the annotation tool can provide functions to easily mark the area with different shapes for segmentation tasks. The users can even do real-time labeling of the segmentation target and justify their strategy based on the performance of ML models. This is much more efficient than the traditional trial-and-error methods by applying different filters. The annotation tool can also help to convert audio to text automatically. Another advantage of the annotation tool is the control of quality. It can offer the function for users to provide feedback on the accuracy of the label. ML studies in AM can take the benefits from the existing annotation tool, and a specific annotation tool for AM can be developed in future work as well.

6.3. Reproducibility and standardization of data-driven research in AM

As a major principle of the scientific method, it is important to keep the reproducibility of the research. Reproducibility generally refers to obtaining consistent results when the study is replicated by using the same input data, methodology, codes, and experimental conditions [201]. Recently, with more and more ML studies conducted in AM applications, reproducibility becomes a challenge due to a lack of shared datasets. Less than half of the articles shared their datasets unless making the request personally. Unlike other ML applications, there are no standard datasets that can be used in the literature such as MNIST (handwritten digit database) [202], IMDB datasets (50k movie reviews), MIMIC (datasets for computational physiology) [203], ShapeNet (3D model repository) [204], etc. Moreover, AM is a relatively complex application for ML. Even for the same AM technology, the different machine has various printing performances. Using the same brand of machine and material is always recommended for fabrication. AM techniques have not been standardized maturely which leads to difficulties in building a general database or dataset for AM.

AM has various research areas including design, process, and manufacturing. It even contains seven different types of technology. It is challenging to build a dataset including all the critical characteristics for AM. However, building a general dataset for a specific application such as visual defects, porosity, thermal distribution, etc. for a specific AM technology such as LPBF and

ME is possible. As summarized in Table 5 to Table 11, the most popular AM techniques are LPBF and ME. A small number of studies focus on DED and VP. With the guarantee of diversity, quantity, non-duplication, and accuracy, multiple stand-alone datasets can be combined together to establish a standard and rich dataset for use of multiple research studies.









6.4. Simple, easy-access, and systematic database for AM

There are still few well-developed AM databases and there is no well-known or commonly used dataset in AM studies. A small portion of AM datasets is duplicated which results in a waste of time. Most of them are private and hard to access. The existing databases are either not designed for AM or more suitable for an organization to manage its internal data. There is no simple data port designed for sharing and accessing AM data publicly. A database for AM is required, and it is expected to be simple, easy-access, and systematic. Therefore, datasets from different studies can be collected. Researchers can save time on collecting data and data sharing can encourage the connection and collaboration between researchers. Moreover, some small datasets can be combined together to generate a larger and richer dataset which can be beneficial for all the AM researchers.

Hence, based on the best understanding of the authors, a potential simple data port is proposed here and ready for data uploading and query. This data port is expected to be web-based and shared with the public. Everyone is welcome to provide their open data or download and reuse the data. For each dataset, the donator needs to fill five required fields and five optional fields. The required fields include AM technique type, raw input data type, application/targets, whether the data is labeled or not, and the zip file for data. The optional fields include Raw output data type, reference source, contact information, machine type, and material type. Machine type and material type include the brand and series information for selected machines and materials. This information can be viewed in the “more details” panel. The preliminary design for the AM data port is shown in Figure 9. On the left of the page, users have the option to filter the database to what they are looking for based on ‘AM technique type’, ‘applications/target’, ‘raw data type’, and ‘labeled or not?’. This simple and informative AM data port is aimed to increase data sharing in the AM community and accelerate the ease of data gathering as well. This data port will not recommend any data handling process or ML algorithm to users. Raw datasets are provided, and users have unlimited freedom to process the data. The data is expected to be used in various research.

Welcome to AM data port

Home Share your data

| No | AM technique type | Application/Target | Raw data type | Labelled or not? | |
|----|-------------------|---------------------------------|---------------|------------------|--|
| 1 | PBF | product-related; visual defects | graphic | Yes | More details  |
| 2 | PBF | product-related; visual defects | graphic | Yes | More details  |
| 3 | PBF | product-related; visual defects | graphic | Yes | More details  |
| 4 | PBF | product-related; visual defects | graphic | Yes | More details  |
| 5 | PBF | product-related; visual defects | graphic | Yes | More details  |
| 6 | PBF | product-related; visual defects | graphic | Yes | More details  |
| 7 | PBF | product-related; visual defects | graphic | Yes | More details  |
| 8 | PBF | product-related; visual defects | graphic | Yes | More details  |

AM technique type

PBF
 ME
 BJ
 Photopolymerization
 MJ
 SL
 DED

Application/Target

Design-related
 Process-related
 Product-related

Raw data type

Tabular
 Graphics
 3D data
 Spectrum data

Labelled or not?

Yes
 No




Figure 9: Preliminary design for the AM data port

7. Conclusions and future perspectives

This paper presents a comprehensive review on ML data processing and management for AM research and applications. Based on the reviewed papers that are published in the recent four years, the utilized data handling methods are summarized for four major data types: tabular data, graphic data, 3D data, and spectrum data. The major handling methods include feature extraction, discretization, data processing, feature selection, and feature learning. It has been noticed that the ML approaches have been applied to various AM applications. At the current status, most of the ML applications in AM focus on product characteristics such as printability, porosity, and surface roughness. The existing studies have demonstrated promising performances. Moreover, the ML approaches are already shown to be suitable and valid for investigating the PSP linkage in AM and the potentials in design for AM. However, there are still some challenges when applying ML in practice including the quantity and diversity of the dataset, the accuracy of the manual labeling, reproducibility, and standardization of data-driven research, and the limitation of the existing databases and management systems. **These challenges need future investigations and could motivate some potential research directions including:**

1. Most of the existing studies use a single type of data as their input to a single ML model. Multiple types of data can be combined together to develop a more comprehensive hybrid ML model.
2. A systematic AM database is critical in ML-assisted AM studies as well as the investigation on the role of AM in industry 4.0. However, there are few studies related to AM database in the literature. A database framework to organize and store AM data including both structured data and unstructured data can be a potential research direction.

3. Drawing analogies from the field of AI, knowledge transfer can be an important next step in data-driven AM. Recently, some researchers have applied the idea of transfer learning to the models developed in the same study [205]. Managing data systematically will support data-based transfer learning from existing ML applications to new ML applications in AM. This can be an important research direction.
4. 91.7% of reviewed articles select supervised machine learning as their approach. 9% of them use unsupervised learning and only 0.7% select reinforcement learning. Unsupervised learning has the advantage of solving problems by learning the data and classifying it without any labels. Also, it can be really helpful in finding patterns in data. Reinforcement learning learns by the modeling self by making and correcting mistakes. It has the potential to solve very complex problems and create a perfect model for a particular problem. More studies on the potential of applying unsupervised learning and reinforcement learning in AM can be conducted in the future.
5. Only a small portion of existing studies focus on the design characteristics of AM. More studies on how ML can help on design for AM can be conducted. For example, ML can help with design idea generation based on the functional needs of the product. Also, ML can help in generating a design more suitable for AM processes with the consideration of functions and cost.

Acknowledgment

Financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development (CRD) CRDPJ 520348-17 for Ying Zhang is acknowledged with gratitude.

Financial supports from the National Research Council of Canada NRC INT-015-1 and McGill Engineering Doctoral Award (MEDA) grant for Mutahar Safdar are acknowledged with gratitude.

Financial supports from MITACs Advanced Manufacturing Automation, Digitization and Optimization (AMADO) grant and McGill Graduate Excellence Fellowship Award for Jiarui Xie are acknowledged with gratitude.

Financial support from the NSERC CRD CRDPJ 479630-15 for Jinghao Li is acknowledged with gratitude. Jinghao Li also received partial funding from the NSERC Collaborative Research, Training Experience (CREATE) Program Grant 449343, MEDA grant, and China Scholarship Council (201706460027).

Financial supports from MITACs AMADO grant and MEDA grant for Manuel Sage are acknowledged with gratitude.

References

1. Kumke, M., et al., *Methods and tools for identifying and leveraging additive manufacturing design potentials*. International Journal on Interactive Design and Manufacturing (IJIDeM), 2018. **12**(2): p. 481-493.
2. Wilson, J.M., et al., *Remanufacturing of turbine blades by laser direct deposition with its energy and environmental impact analysis*. Journal of Cleaner Production, 2014. **80**: p. 170-178.

3. Dutta, B. and F.H.S. Froes, *The additive manufacturing (AM) of titanium alloys*, in *Titanium powder metallurgy*. 2015, Elsevier. p. 447-468.
4. Kumke, M., et al., *Methods and tools for identifying and leveraging additive manufacturing design potentials*. International Journal on Interactive Design and Manufacturing (IJIDeM), 2018. **12**(2): p. 481-493.
5. DebRoy, T., et al., *Additive manufacturing of metallic components—process, structure and properties*. Progress in Materials Science, 2018. **92**: p. 112-224.
6. Greitemeier, D., et al., *Uncertainty of Additive Manufactured Ti-6Al-4V: Chemistry, Microstructure and Mechanical Properties*. Applied Mechanics and Materials, 2015. **807**: p. 169-180.
7. Li, J., et al., *Solidification Microstructure Simulation of Ti-6Al-4V in Metal Additive Manufacturing: A Review*. Additive Manufacturing, 2019: p. 100989.
8. Markl, M. and C. Körner, *Multiscale Modeling of Powder Bed-Based Additive Manufacturing*. Annual Review of Materials Research, 2016. **46**(1): p. 93-123.
9. DebRoy, T., et al., *Metallurgy, mechanistic models and machine learning in metal printing*. Nature Reviews Materials, 2021. **6**(1): p. 48-68.
10. Huang, D.J. and H. Li, *Review of machine learning applications in powder bed fusion technology for part production*. 2018.
11. Johnson, N., et al., *Machine Learning for Materials Developments in Metals Additive Manufacturing*. 2020.
12. Zhu, K., J.Y.H. Fuh, and X.J.I.A.T.o.M. Lin, *Metal-Based Additive Manufacturing Condition Monitoring: A Review on Machine Learning Based Approaches*. 2021.
13. Mahmoud, D., et al., *Applications of Machine Learning in Process Monitoring and Controls of L-PBF Additive Manufacturing: A Review*. 2021. **11**(24): p. 11910.
14. Qi, X., et al., *Applying neural-network-based machine learning to additive manufacturing: current applications, challenges, and future perspectives*. 2019. **5**(4): p. 721-729.
15. Joshi, M., et al. *Applications of Supervised Machine Learning Algorithms in Additive Manufacturing: A Review*. in *2019 International Solid Freeform Fabrication Symposium*. 2019. University of Texas at Austin.
16. Tian, C., et al., *Data-Driven Approaches Toward Smarter Additive Manufacturing*. 2021: p. 2100014.
17. Goh, G.D., S.L. Sing, and W.Y.J.A.I.R. Yeong, *A review on machine learning in 3D printing: applications, potential, and challenges*. 2021. **54**(1): p. 63-94.
18. Wang, C., et al., *Machine learning in additive manufacturing: State-of-the-art and perspectives*. 2020: p. 101538.
19. Jin, Z., et al., *Machine learning for advanced additive manufacturing*. 2020. **3**(5): p. 1541-1556.
20. Meng, L., et al., *Machine learning in additive manufacturing: A review*. 2020. **72**(6): p. 2363-2377.
21. Razvi, S.S., et al. *A review of machine learning applications in additive manufacturing*. in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. 2019. American Society of Mechanical Engineers.
22. Yang, J., et al. *Survey on artificial intelligence for additive manufacturing*. in *2017 23rd International Conference on Automation and Computing (ICAC)*. 2017. IEEE.
23. Guyon, I., et al., *Feature extraction: foundations and applications*. Vol. 207. 2008: Springer.

24. Li, J., et al., *Feature selection: A data perspective*. ACM Computing Surveys (CSUR), 2017. **50**(6): p. 1-45.
25. Bengio, Y., A. Courville, and P. Vincent, *Representation learning: A review and new perspectives*. IEEE transactions on pattern analysis and machine intelligence, 2013. **35**(8): p. 1798-1828.
26. Alasadi, S.A. and W.S. Bhaya, *Review of data preprocessing techniques in data mining*. Journal of Engineering and Applied Sciences, 2017. **12**(16): p. 4102-4107.
27. Chaki, J. and N. Dey, *A beginner's guide to image preprocessing techniques*. 2018: CRC Press.
28. Singh, D. and B. Singh, *Investigating the impact of data normalization on classification performance*. Applied Soft Computing, 2020. **97**: p. 105524.
29. Tennison, J., G. Kellogg, and I. Herman, *Model for tabular data and metadata on the web*. 2015.
30. Zhang, M., et al., *High cycle fatigue life prediction of laser additive manufactured stainless steel: A machine learning approach*. International Journal of Fatigue, 2019. **128**: p. 105194.
31. Lee, S., et al., *Data analytics approach for melt-pool geometries in metal additive manufacturing*. Science and technology of advanced materials, 2019. **20**(1): p. 972-978.
32. Jiang, J., et al., *Achieving better connections between deposited lines in additive manufacturing via machine learning*. Math. Biosci. Eng, 2020. **17**(4).
33. Rodríguez-Martín, M., et al., *Predictive models for the characterization of internal defects in additive materials from active thermography sequences supported by machine learning methods*. Sensors, 2020. **20**(14): p. 3982.
34. Montazeri, M., et al., *In-process monitoring of porosity in additive manufacturing using optical emission spectroscopy*. IISE Transactions, 2020. **52**(5): p. 500-515.
35. García, S., J. Luengo, and F. Herrera, *Data preprocessing in data mining*. Vol. 72. 2015: Springer.
36. Murray, J.D. and W. VanRyper, *Encyclopedia of graphics file formats*. Sebastopol: O'Reilly, 1996.
37. Saluja, A., J. Xie, and K. Fayazbakhsh, *A closed-loop in-process warping detection system for fused filament fabrication using convolutional neural networks*. Journal of Manufacturing Processes, 2020. **58**: p. 407-415.
38. Yeung, H., Z. Yang, and L. Yan, *A melt-pool prediction based scan strategy for powder bed fusion additive manufacturing*. Additive Manufacturing, 2020. **35**: p. 101383.
39. Caggiano, A., et al., *Automated Laser Polishing for surface finish enhancement of additive manufactured components for the automotive industry*. Production Engineering, 2021. **15**(1): p. 109-117.
40. DeCost, B.L., et al., *Computer vision and machine learning for autonomous characterization of am powder feedstocks*. Jom, 2017. **69**(3): p. 456-465.
41. Özel, T., et al., *Surface topography investigations on nickel alloy 625 fabricated via laser powder bed fusion*. The International Journal of Advanced Manufacturing Technology, 2018. **94**(9): p. 4451-4458.
42. Paulson, N.H., et al., *Correlations between thermal history and keyhole porosity in laser powder bed fusion*. Additive Manufacturing, 2020. **34**: p. 101213.
43. Zhang, J., P. Wang, and R.X. Gao, *Deep learning-based tensile strength prediction in fused deposition modeling*. Computers in industry, 2019. **107**: p. 11-21.

44. Zhang, X., J. Saniie, and A. Heifetz, *Detection of Defects in Additively Manufactured Stainless Steel 316L with Compact Infrared Camera and Machine Learning Algorithms*. JOM, 2020. **72**(12): p. 4244-4253.
45. Zhang, Z., et al., *Predicting flexural strength of additively manufactured continuous carbon fiber-reinforced polymer composites using machine learning*. Journal of Computing and Information Science in Engineering, 2020. **20**(6): p. 061015.
46. Bai, M., et al., *Anomaly detection of gas turbines based on normal pattern extraction*. Applied Thermal Engineering, 2020. **166**: p. 114664.
47. Gu, G.X., et al., *Bioinspired hierarchical composite design using machine learning: simulation, additive manufacturing, and experiment*. Materials Horizons, 2018. **5**(5): p. 939-945.
48. Ren, Y.M., et al., *Computational fluid dynamics-based in-situ sensor analytics of direct metal laser solidification process using machine learning*. Computers & Chemical Engineering, 2020. **143**: p. 107069.
49. Sanchez, S., et al., *Machine learning to determine the main factors affecting creep rates in laser powder bed fusion*. Journal of Intelligent Manufacturing, 2021: p. 1-21.
50. Snell, R., et al., *Methods for rapid pore classification in metal additive manufacturing*. JOM, 2020. **72**(1): p. 101-109.
51. Han, Y., et al., *Quantitative microstructure analysis for solid-state metal additive manufacturing via deep learning*. Journal of Materials Research, 2020. **35**(15): p. 1936-1948.
52. Imani, F., et al., *Process mapping and in-process monitoring of porosity in laser powder bed fusion using layerwise optical imaging*. Journal of Manufacturing Science and Engineering, 2018. **140**(10).
53. Shorten, C. and T.M. Khoshgoftaar, *A survey on image data augmentation for deep learning*. Journal of Big Data, 2019. **6**(1): p. 1-48.
54. You, S., et al., *Mitigating Scattering Effects in Light-Based Three-Dimensional Printing Using Machine Learning*. Journal of Manufacturing Science and Engineering, 2020. **142**(8): p. 081002.
55. Yazdi, R.M., F. Imani, and H.J.J.o.M.S. Yang, *A hybrid deep learning model of process-build interactions in additive manufacturing*. Journal of Manufacturing Systems, 2020. **57**: p. 460-468.
56. Roach, D.J., et al., *Utilizing computer vision and artificial intelligence algorithms to predict and design the mechanical compression response of direct ink write 3D printed foam replacement structures*. Additive Manufacturing, 2021. **41**: p. 101950.
57. Khanzadeh, M., et al., *Porosity prediction: Supervised-learning of thermal history for direct laser deposition*. Journal of manufacturing systems, 2018. **47**: p. 69-82.
58. McHenry, K. and P. Bajcsy, *An overview of 3d data content, file formats and viewers*. National Center for Supercomputing Applications, 2008. **1205**: p. 22.
59. Szilvsi-Nagy, M. and G. Matyasi, *Analysis of STL files*. Mathematical and computer modelling, 2003. **38**(7-9): p. 945-960.
60. Zhang, Y., et al., *Predictive manufacturability assessment system for laser powder bed fusion based on a hybrid machine learning model*. Additive Manufacturing, 2021. **41**: p. 101946.

61. Herriott, C. and A.D. Spear, *Predicting microstructure-dependent mechanical properties in additively manufactured metals with machine-and deep-learning methods*. Computational Materials Science, 2020. **175**: p. 109599.
62. Gobert, C., et al., *Porosity segmentation in X-ray computed tomography scans of metal additively manufactured specimens with machine learning*. Additive Manufacturing, 2020. **36**: p. 101460.
63. Zhang, Y. and Y.F. Zhao, *Hybrid sparse convolutional neural networks for predicting manufacturability of visual defects of laser powder bed fusion processes*. Journal of Manufacturing Systems, 2021.
64. Kuschmitz, S., et al., *Design and Additive Manufacturing of Porous Sound Absorbers—A Machine-Learning Approach*. Materials, 2021. **14**(7): p. 1747.
65. Yang, S., et al., *Towards an automated decision support system for the identification of additive manufacturing part candidates*. Journal of Intelligent Manufacturing, 2020: p. 1-17.
66. Wu, H., Z. Yu, and Y. Wang, *Experimental study of the process failure diagnosis in additive manufacturing based on acoustic emission*. Measurement, 2019. **136**: p. 445-453.
67. Shevchik, S.A., et al., *Acoustic emission for in situ quality monitoring in additive manufacturing using spectral convolutional neural networks*. Additive Manufacturing, 2018. **21**: p. 598-604.
68. Wasmer, K., et al., *In situ quality monitoring in AM using acoustic emission: A reinforcement learning approach*. Journal of Materials Engineering and Performance, 2019. **28**(2): p. 666-672.
69. Okaro, I.A., et al., *Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning*. Additive Manufacturing, 2019. **27**: p. 42-53.
70. Yang, Z., et al., *Filament breakage monitoring in fused deposition modeling using acoustic emission technique*. Sensors, 2018. **18**(3): p. 749.
71. Hong, K., et al., *A vibration measurement system for health monitoring of power transformers*. Measurement, 2016. **93**: p. 135-147.
72. Zouhri, W., et al., *Optical process monitoring for Laser-Powder Bed Fusion (L-PBF)*. CIRP Journal of Manufacturing Science and Technology, 2020. **31**: p. 607-617.
73. Wu, D., Y. Wei, and J. Terpenney, *Predictive modelling of surface roughness in fused deposition modelling using data fusion*. International Journal of Production Research, 2019. **57**(12): p. 3992-4006.
74. Obaton, A.-F., et al., *A non-destructive resonant acoustic testing and defect classification of additively manufactured lattice structures*. Welding in the World, 2021. **65**(3): p. 361-371.
75. Zhang, Z., et al., *Prediction of Dimensional Changes of Low-Cost Metal Material Extrusion Fabricated Parts Using Machine Learning Techniques*. Metals, 2021. **11**(5): p. 690.
76. Garland, A.P., et al., *Pragmatic generative optimization of novel structural lattice metamaterials with machine learning*. Materials & Design, 2021. **203**: p. 109632.
77. Baturynska, I. and K. Martinsen, *Prediction of geometry deviations in additive manufactured parts: comparison of linear regression with machine learning algorithms*. Journal of Intelligent Manufacturing, 2021. **32**(1): p. 179-200.

78. Bessa, M.A., P. Glowacki, and M. Houlder, *Bayesian machine learning in metamaterial design: Fragile becomes supercompressible*. *Advanced Materials*, 2019. **31**(48): p. 1904845.
79. Jiang, J., et al., *Machine learning integrated design for additive manufacturing*. *Journal of Intelligent Manufacturing*, 2020: p. 1-14.
80. Yanamandra, K., et al., *Reverse engineering of additive manufactured composite part by toolpath reconstruction using imaging and machine learning*. *Composites Science and Technology*, 2020. **198**: p. 108318.
81. Després, N., et al., *Deep learning and design for additive manufacturing: a framework for microlattice architecture*. *Jom*, 2020. **72**(6): p. 2408-2418.
82. Minnema, J., et al., *CT image segmentation of bone for medical additive manufacturing using a convolutional neural network*. *Computers in biology and medicine*, 2018. **103**: p. 130-139.
83. Vafadar, A., et al., *Advances in metal additive manufacturing: a review of common processes, industrial applications, and current challenges*. *Applied Sciences*, 2021. **11**(3): p. 1213.
84. Bandyopadhyay, A., Y. Zhang, and S. Bose, *Recent developments in metal additive manufacturing*. *Current opinion in chemical engineering*, 2020. **28**: p. 96-104.
85. Masinelli, G., et al. *Artificial Intelligence for Monitoring and Control of Metal Additive Manufacturing*. in *International Conference on Additive Manufacturing in Products and Applications*. 2020. Springer.
86. Michopoulos, J.G., et al., *On the multiphysics modeling challenges for metal additive manufacturing processes*. *Additive Manufacturing*, 2018. **22**: p. 784-799.
87. Scime, L. and J. Beuth, *A multi-scale convolutional neural network for autonomous anomaly detection and classification in a laser powder bed fusion additive manufacturing process*. *Additive Manufacturing*, 2018. **24**: p. 273-286.
88. Stanisavljevic, D., et al., *Detection of interferences in an additive manufacturing process: an experimental study integrating methods of feature selection and machine learning*. *International Journal of Production Research*, 2020. **58**(9): p. 2862-2884.
89. He, H., Y. Yang, and Y. Pan, *Machine learning for continuous liquid interface production: Printing speed modelling*. *Journal of Manufacturing Systems*, 2019. **50**: p. 236-246.
90. Gardner, J.M., et al., *Machines as craftsmen: localized parameter setting optimization for fused filament fabrication 3D printing*. *Advanced Materials Technologies*, 2019. **4**(3): p. 1800653.
91. Bastani, K., P.K. Rao, and Z. Kong, *An online sparse estimation-based classification approach for real-time monitoring in advanced manufacturing processes from heterogeneous sensor data*. *IIE Transactions*, 2016. **48**(7): p. 579-598.
92. Osswald, P.V., et al., *Optimization of the production processes of powder-based additive manufacturing technologies by means of a machine learning model for the temporal prognosis of the build and cooling phase*. *Production Engineering*, 2020. **14**(5): p. 677-691.
93. Xiangyang, Y., et al., *Atomic simulations of melting behaviours for TiAl alloy nanoparticles during heating*. *Bulletin of Materials Science*, 2020. **43**(1): p. 1-9.
94. Caiazzo, F. and A. Caggiano, *Laser direct metal deposition of 2024 Al alloy: trace geometry prediction via machine learning*. *Materials*, 2018. **11**(3): p. 444.
95. Oehlmann, P., et al., *Modeling Fused Filament Fabrication using Artificial Neural Networks*. *Production Engineering*, 2021. **15**(3): p. 467-478.

96. Desai, P.S. and C.F. Higgs, *Spreading process maps for powder-bed additive manufacturing derived from physics model-based machine learning*. *Metals*, 2019. **9**(11): p. 1176.
97. Wang, T., et al., *In-situ droplet inspection and closed-loop control system using machine learning for liquid metal jet printing*. *Journal of manufacturing systems*, 2018. **47**: p. 83-92.
98. Wang, Z., et al., *Analysis of Critical Velocity of Cold Spray Based on Machine Learning Method with Feature Selection*. *Journal of Thermal Spray Technology*, 2021: p. 1-13.
99. Caggiano, A., et al., *Machine learning-based image processing for on-line defect recognition in additive manufacturing*. *CIRP Annals*, 2019. **68**(1): p. 451-454.
100. Scime, L., 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*, 2020. **36**: p. 101453.
101. Mativo, T., C. Fritz, and F. Ismail, *Cyber acoustic analysis of additively manufactured objects*. *The International Journal of Advanced Manufacturing Technology*, 2018. **96**(1-4): p. 581-586.
102. Nguyen, L., J. Buhl, and M. Bambach, *Continuous Eulerian tool path strategies for wire-arc additive manufacturing of rib-web structures with machine-learning-based adaptive void filling*. *Additive Manufacturing*, 2020. **35**: p. 101265.
103. Thomas, M., et al., *Distributed heterogeneous compute infrastructure for the study of additive manufacturing systems*. *MRS Advances*, 2020. **5**(29): p. 1547-1555.
104. Zhou, B. and T. Tian, *A path planning method of lattice structural components for additive manufacturing*. *The International Journal of Advanced Manufacturing Technology*, 2021: p. 1-24.
105. Zohdi, T., *Electrodynamic machine-learning-enhanced fault-tolerance of robotic free-form printing of complex mixtures*. *Computational Mechanics*, 2019. **63**(5): p. 913-929.
106. Zhu, Q., Z. Liu, and J. Yan, *Machine learning for metal additive manufacturing: predicting temperature and melt pool fluid dynamics using physics-informed neural networks*. *Computational Mechanics*, 2021. **67**(2): p. 619-635.
107. Kumar, H.A., et al., *Faster temperature prediction in the powder bed fusion process through the development of a surrogate model*. *Optics & Laser Technology*, 2021. **141**: p. 107122.
108. Roy, M. and O. Wodo, *Data-driven modeling of thermal history in additive manufacturing*. *Additive Manufacturing*, 2020. **32**: p. 101017.
109. Ren, K., et al., *Thermal field prediction for laser scanning paths in laser aided additive manufacturing by physics-based machine learning*. *Computer Methods in Applied Mechanics and Engineering*, 2020. **362**: p. 112734.
110. Zhou, Z., et al., *Thermal field prediction for welding paths in multi-layer gas metal arc welding-based additive manufacturing: A machine learning approach*. *Journal of Manufacturing Processes*, 2021. **64**: p. 960-971.
111. Ren, K., et al., *Integrated numerical modelling and deep learning for multi-layer cube deposition planning in laser aided additive manufacturing*. *Virtual and Physical Prototyping*, 2021: p. 1-15.
112. Mondal, S., et al., *Investigation of melt pool geometry control in additive manufacturing using hybrid modeling*. *Metals*, 2020. **10**(5): p. 683.

113. Yuan, B., et al., *Machine -learning -based monitoring of laser powder bed fusion*. Advanced Materials Technologies, 2018. **3**(12): p. 1800136.
114. Donegan, S.P., E.J. Schwalbach, and M.A. Groeber, *Zoning additive manufacturing process histories using unsupervised machine learning*. Materials Characterization, 2020. **161**: p. 110123.
115. Haghghi, A. and L. Li, *A hybrid physics-based and data-driven approach for characterizing porosity variation and filament bonding in extrusion-based additive manufacturing*. Additive Manufacturing, 2020. **36**: p. 101399.
116. Alejandrino, J.D., et al., *A machine learning approach of lattice infill pattern for increasing material efficiency in additive manufacturing processes*. Int. J. Mech. Eng. Robot. Res., 2020. **9**(9): p. 1253-1263.
117. Schaechtl, P., B. Schleich, and S. Wartzack, *Statistical tolerance analysis of 3d-printed non-assembly mechanisms in motion using empirical predictive models*. Applied Sciences, 2021. **11**(4): p. 1860.
118. Khanzadeh, M., et al., *Quantifying geometric accuracy with unsupervised machine learning: Using self-organizing map on fused filament fabrication additive manufacturing parts*. Journal of Manufacturing Science and Engineering, 2018. **140**(3).
119. Khan, M.F., et al., *Real-time defect detection in 3D printing using machine learning*. Materials Today: Proceedings, 2021. **42**: p. 521-528.
120. Bisheh, M.N., S.I. Chang, and S. Lei, *A layer-by-layer quality monitoring framework for 3D printing*. Computers & Industrial Engineering, 2021. **157**: p. 107314.
121. Westphal, E. and H. Seitz, *A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks*. Additive Manufacturing, 2021. **41**: p. 101965.
122. Ko, H., et al., *Machine learning and knowledge graph based design rule construction for additive manufacturing*. Additive Manufacturing, 2021. **37**: p. 101620.
123. Korneev, S., et al., *Fabricated shape estimation for additive manufacturing processes with uncertainty*. Computer-Aided Design, 2020. **127**: p. 102852.
124. Ding, D., et al., *The first step towards intelligent wire arc additive manufacturing: An automatic bead modelling system using machine learning through industrial information integration*. Journal of Industrial Information Integration, 2021. **23**: p. 100218.
125. Zhang, Z., I. Fidan, and M. Allen, *Detection of material extrusion in-process failures via deep learning*. Inventions, 2020. **5**(3): p. 25.
126. Amini, M. and S.I. Chang, *MLCPM: A process monitoring framework for 3D metal printing in industrial scale*. Computers & Industrial Engineering, 2018. **124**: p. 322-330.
127. Jin, Z., et al., *Precise localization and semantic segmentation detection of printing conditions in fused filament fabrication technologies using machine learning*. Additive Manufacturing, 2021. **37**: p. 101696.
128. Baumgartl, H., et al., *A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring*. Progress in Additive Manufacturing, 2020: p. 1-9.
129. Wu, M., Z. Song, and Y.B. Moon, *Detecting cyber-physical attacks in CyberManufacturing systems with machine learning methods*. Journal of intelligent manufacturing, 2019. **30**(3): p. 1111-1123.
130. Li, R., M. Jin, and V.C. Paquit, *Geometrical defect detection for additive manufacturing with machine learning models*. Materials & Design, 2021. **206**: p. 109726.

131. Zhu, Z., et al., *Machine learning in tolerancing for additive manufacturing*. CIRP Annals, 2018. **67**(1): p. 157-160.
132. Guo, Y., W.F. Lu, and J.Y.H. Fuh, *Semi-supervised deep learning based framework for assessing manufacturability of cellular structures in direct metal laser sintering process*. Journal of Intelligent Manufacturing, 2021. **32**(2): p. 347-359.
133. Elbadawi, M., et al., *M3DISEEN: A novel machine learning approach for predicting the 3D printability of medicines*. International Journal of Pharmaceutics, 2020. **590**: p. 119837.
134. Schur, R., et al., *Mechanical anisotropy and its evolution with powder reuse in Electron Beam Melting AM of Ti6Al4V*. Materials & Design, 2021. **200**: p. 109450.
135. Adnan, M., et al., *A New Architectural Approach to Monitoring and Controlling AM Processes*. Applied Sciences, 2020. **10**(18): p. 6616.
136. Coatanéa, E., et al., *Systematic manufacturability evaluation using dimensionless metrics and singular value decomposition: a case study for additive manufacturing*. The International Journal of Advanced Manufacturing Technology, 2020: p. 1-17.
137. Mycroft, W., et al., *A data-driven approach for predicting printability in metal additive manufacturing processes*. Journal of Intelligent Manufacturing, 2020. **31**(7): p. 1769-1781.
138. Mahmoudi, M., et al., *On the printability and transformation behavior of nickel-titanium shape memory alloys fabricated using laser powder-bed fusion additive manufacturing*. Journal of Manufacturing Processes, 2018. **35**: p. 672-680.
139. Chen, L., et al., *Rapid surface defect identification for additive manufacturing with in-situ point cloud processing and machine learning*. Virtual and Physical Prototyping, 2021. **16**(1): p. 50-67.
140. Samie Tootooni, M., et al., *Classifying the dimensional variation in additive manufactured parts from laser-scanned three-dimensional point cloud data using machine learning approaches*. Journal of Manufacturing Science and Engineering, 2017. **139**(9).
141. Mitchell, J.A., et al., *Linking pyrometry to porosity in additively manufactured metals*. Additive Manufacturing, 2020. **31**: p. 100946.
142. Tian, Q., S. Guo, and Y. Guo, *A physics-driven deep learning model for process-porosity causal relationship and porosity prediction with interpretability in laser metal deposition*. CIRP Annals, 2020. **69**(1): p. 205-208.
143. Liu, R., S. Liu, and X. Zhang, *A physics-informed machine learning model for porosity analysis in laser powder bed fusion additive manufacturing*. The International Journal of Advanced Manufacturing Technology, 2021. **113**(7): p. 1943-1958.
144. Zhang, X., et al., *Quality control of additively manufactured metallic structures with machine learning of thermography images*. JOM, 2020. **72**(12): p. 4682-4694.
145. García-Moreno, A.-I., et al., *Image-based porosity classification in Al-alloys by laser metal deposition using random forests*. The International Journal of Advanced Manufacturing Technology, 2020. **110**(9): p. 2827-2845.
146. Shevchik, S.A., et al., *Deep learning for in situ and real-time quality monitoring in additive manufacturing using acoustic emission*. IEEE Transactions on Industrial Informatics, 2019. **15**(9): p. 5194-5203.
147. Bartlett, J.L., et al., *Prediction of microstructural defects in additive manufacturing from powder bed quality using digital image correlation*. Materials Science and Engineering: A, 2020. **794**: p. 140002.
148. Montazeri, M., et al., *Heterogeneous sensor-based condition monitoring in directed energy deposition*. Additive Manufacturing, 2019. **30**: p. 100916.

149. García-Moreno, A.-I., et al., *Ex-situ porosity classification in metallic components by laser metal deposition: A machine learning-based approach*. Journal of Manufacturing Processes, 2021. **62**: p. 523-534.
150. Scime, L. and J. Beuth, *Using machine learning to identify in-situ melt pool signatures indicative of flaw formation in a laser powder bed fusion additive manufacturing process*. Additive Manufacturing, 2019. **25**: p. 151-165.
151. Snow, Z., et al., *Toward in-situ flaw detection in laser powder bed fusion additive manufacturing through layerwise imagery and machine learning*. Journal of Manufacturing Systems, 2021. **59**: p. 12-26.
152. Kapusuzoglu, B. and S. Mahadevan, *Physics-informed and hybrid machine learning in additive manufacturing: Application to fused filament fabrication*. JOM, 2020. **72**(12): p. 4695-4705.
153. Li, J., et al., *Machine Learning-Enabled Competitive Grain Growth Behavior Study in Directed Energy Deposition Fabricated Ti6Al4V*. JOM, 2020. **72**(1): p. 458-464.
154. Liu, C., et al., *Digital twin-enabled collaborative data management for metal additive manufacturing systems*. Journal of Manufacturing Systems, 2020.
155. Yadav, P., et al., *Inline Drift Detection Using Monitoring Systems and Machine Learning in Selective Laser Melting*. Advanced Engineering Materials, 2020. **22**(12): p. 2000660.
156. Scime, L. and J. Beuth, *Anomaly detection and classification in a laser powder bed additive manufacturing process using a trained computer vision algorithm*. Additive Manufacturing, 2018. **19**: p. 114-126.
157. Zhu, Y., et al., *Unraveling pore evolution in post-processing of binder jetting materials: X-ray computed tomography, computer vision, and machine learning*. Additive Manufacturing, 2020. **34**: p. 101183.
158. Baturynska, I., *Application of machine learning techniques to predict the mechanical properties of polyamide 2200 (PA12) in additive manufacturing*. Applied Sciences, 2019. **9**(6): p. 1060.
159. Chang, T.-W., et al., *Predicting magnetic characteristics of additive manufactured soft magnetic composites by machine learning*. The International Journal of Advanced Manufacturing Technology, 2021. **114**(9): p. 3177-3184.
160. Hassanin, H., et al., *Controlling the properties of additively manufactured cellular structures using machine learning approaches*. Advanced Engineering Materials, 2020. **22**(3): p. 1901338.
161. Hu, C., et al., *The fabrication of long carbon fiber reinforced polylactic acid composites via fused deposition modelling: Experimental analysis and machine learning*. Journal of Composite Materials, 2021. **55**(11): p. 1459-1472.
162. Zhang, Z., et al., *Data-driven predictive modeling of tensile behavior of parts fabricated by cooperative 3d printing*. Journal of Computing and Information Science in Engineering, 2020. **20**(2): p. 021002.
163. Aladesanmi, V., et al., *Regression analysis of hardness property of laser additive manufactured (LAM) Ti and TiB₂ metal matrix composite*. Materials Today: Proceedings, 2021. **44**: p. 1249-1253.
164. Zhan, Z. and H. Li, *A novel approach based on the elastoplastic fatigue damage and machine learning models for life prediction of aerospace alloy parts fabricated by additive manufacturing*. International Journal of Fatigue, 2021. **145**: p. 106089.

165. Zhan, Z. and H. Li, *Machine learning based fatigue life prediction with effects of additive manufacturing process parameters for printed SS 316L*. International Journal of Fatigue, 2021. **142**: p. 105941.
166. Li, Z., et al., *Prediction of surface roughness in extrusion-based additive manufacturing with machine learning*. Robotics and Computer-Integrated Manufacturing, 2019. **57**: p. 488-495.
167. Xia, C., et al., *Modelling and prediction of surface roughness in wire arc additive manufacturing using machine learning*. Journal of Intelligent Manufacturing, 2021: p. 1-16.
168. Rankouhi, B., et al., *Compositional grading of a 316L-Cu multi-material part using machine learning for the determination of selective laser melting process parameters*. Additive Manufacturing, 2021. **38**: p. 101836.
169. Hajjalizadeh, F. and A. Ince, *Integration of artificial neural network with finite element analysis for residual stress prediction of direct metal deposition process*. Materials Today Communications, 2021. **27**: p. 102197.
170. Wu, Q., et al., *Residual stresses in wire-arc additive manufacturing—Hierarchy of influential variables*. Additive Manufacturing, 2020. **35**: p. 101355.
171. Liu, S., et al., *Machine learning for knowledge transfer across multiple metals additive manufacturing printers*. Additive Manufacturing, 2021. **39**: p. 101877.
172. Marmarelis, M.G. and R.G. Ghanem, *Data-driven stochastic optimization on manifolds for additive manufacturing*. Computational Materials Science, 2020. **181**: p. 109750.
173. Srinivasan, S., B. Swick, and M.A. Groeber, *Laser Powder Bed Fusion Parameter Selection via Machine-Learning-Augmented Process Modeling*. JOM, 2020. **72**(12): p. 4393-4403.
174. Chan, S.L., Y. Lu, and Y. Wang, *Data-driven cost estimation for additive manufacturing in cybermanufacturing*. Journal of manufacturing systems, 2018. **46**: p. 115-126.
175. Meißner, P., et al., *Artificial neural networks-based material parameter identification for numerical simulations of additively manufactured parts by material extrusion*. Polymers, 2020. **12**(12): p. 2949.
176. Nagarajan, H.P., et al., *Knowledge-based design of artificial neural network topology for additive manufacturing process modeling: A new approach and case study for fused deposition modeling*. Journal of Mechanical Design, 2019. **141**(2): p. 021705.
177. Park, H.S., et al., *Machine learning-based optimization of process parameters in selective laser melting for biomedical applications*. Journal of Intelligent Manufacturing, 2021: p. 1-16.
178. Gaikwad, A., et al., *Heterogeneous sensing and scientific machine learning for quality assurance in laser powder bed fusion—A single-track study*. Additive Manufacturing, 2020. **36**: p. 101659.
179. Lee, X.Y., et al., *Automated detection of part quality during two-photon lithography via deep learning*. Additive Manufacturing, 2020. **36**: p. 101444.
180. Williams, G., et al., *Design repository effectiveness for 3D convolutional neural networks: Application to additive manufacturing*. Journal of Mechanical Design, 2019. **141**(11).
181. Huang, D.J. and H. Li, *A machine learning guided investigation of quality repeatability in metal laser powder bed fusion additive manufacturing*. Materials & Design, 2021. **203**: p. 109606.

182. Hertlein, N., et al., *Prediction of selective laser melting part quality using hybrid Bayesian network*. Additive Manufacturing, 2020. **32**: p. 101089.
183. Garland, A.P., et al., *Deep convolutional neural networks as a rapid screening tool for complex additively manufactured structures*. Additive Manufacturing, 2020. **35**: p. 101217.
184. Liu, Q., et al., *Machine-learning assisted laser powder bed fusion process optimization for AlSi10Mg: New microstructure description indices and fracture mechanisms*. Acta Materialia, 2020. **201**: p. 316-328.
185. Shrestha, R., et al. *Effect of Build Orientation on the Fatigue Behavior of Stainless Steel 316L Manufactured Via A Laser-Powder Bed Fusion Process*. in *27th Annual Solid Freeform Fabrication Symposium Proceedings*. 2016.
186. Bhoi, N.K., *Mendeley data repository as a platform for research data management*. *Marching Beyond Libraries: Managerial Skills and Technological Competencies*, 2018: p. 481-487.
187. Ayers, M., *DOE Data Explorer: The Data*. *Issues in Science and Technology Librarianship*, 2008(55).
188. Ding, L., et al. *Data-gov wiki: Towards linking government data*. in *2010 AAAI Spring Symposium Series*. 2010.
189. Brase, J. *DataCite-A global registration agency for research data*. in *2009 fourth international conference on cooperation and promotion of information resources in science and technology*. 2009. IEEE.
190. Mahmoudi, M., et al., *Mechanical properties and microstructural characterization of selective laser melted 17-4 PH stainless steel*. *Rapid Prototyping Journal*, 2017. **23**(2): p. 280-294.
191. Lu, Y., P. Witherell, and A. Donmez. *A collaborative data management system for additive manufacturing*. in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. 2017. American Society of Mechanical Engineers.
192. Alwosheel, A., S. van Cranenburgh, and C.G. Chorus, *Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis*. *Journal of choice modelling*, 2018. **28**: p. 167-182.
193. Cho, J., et al., *How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?* arXiv preprint arXiv:1511.06348, 2015.
194. Kavzoglu, T. and P. Mather, *The use of backpropagating artificial neural networks in land cover classification*. *International journal of remote sensing*, 2003. **24**(23): p. 4907-4938.
195. Jain, A.K. and B. Chandrasekaran, *39 Dimensionality and sample size considerations in pattern recognition practice*. *Handbook of statistics*, 1982. **2**: p. 835-855.
196. Abu-Mostafa, Y.S., *Hints*. *Neural Computation*, 1995. **7**(4): p. 639-671.
197. Haykin, S., *Neural networks and learning machines, 3/E*. 2010: Pearson Education India.
198. Erdmann, M., et al. *From manual to semi-automatic semantic annotation: About ontology-based text annotation tools*. in *Proceedings of the COLING-2000 Workshop on Semantic Annotation and Intelligent Content*. 2000.
199. Ernst, M., et al., *MolNetEnhancer: Enhanced molecular networks by integrating metabolome mining and annotation tools*. *Metabolites*, 2019. **9**(7): p. 144.
200. Costello, C., et al. *Dragonfly: Advances in non-speaker annotation for low resource languages*. in *Proceedings of the 12th Language Resources and Evaluation Conference*. 2020.

201. Pineau, J., et al., *Improving Reproducibility in Machine Learning Research*. Journal of Machine Learning Research, 2021. **22**: p. 1-20.
202. Deng, L., *The mnist database of handwritten digit images for machine learning research [best of the web]*. IEEE Signal Processing Magazine, 2012. **29**(6): p. 141-142.
203. Johnson, A.E., et al., *MIMIC-III, a freely accessible critical care database*. Scientific data, 2016. **3**(1): p. 1-9.
204. Chang, A.X., et al., *Shapenet: An information-rich 3d model repository*. arXiv preprint arXiv:1512.03012, 2015.
205. Zhang, H., et al., *A knowledge transfer framework to support rapid process modeling in aerosol jet printing*. 2021. **48**: p. 101264.