Automatic quality assessments of laser powder bed fusion builds from photodiode sensor measurements

Sarini Jayasinghe[1] · Paolo Paoletti[1] · Chris Sutcliffe[1] · John Dardis[2] · Nick Jones[2] · Peter L. Green[1]

Abstract

While Laser powder bed fusion (L-PBF) machines have greatly improved in recent years, the L-PBF process is still susceptible to several types of defect formation. Among the monitoring methods that have been explored to detect these defects, camera-based systems are the most prevalent. However, using only photodiode measurements to monitor the build process has potential benefits, as photodiode sensors are cost-efficient and typically have a higher sample rate compared to cameras. This study evaluates whether a combination of photodiode sensor measurements, taken during L-PBF builds, can be used to predict measures of the resulting build quality via a purely data-based approach. Using several unsupervised clustering approaches build density is classified with up to 93.54% accuracy using features extracted from three different photodiodes, as well as observations relating to the energy transferred to the material. Subsequently, a supervised learning method (Gaussian Process regression) is used to directly predict build density with a RMS error of 3.65%. The study, therefore, shows the potential for machine-learning algorithms to predict indicators of L-PBF build quality from photodiode build measurements only. This study also shows that, relative to the L-PBF process parameters, photodiode measurements can contribute to additional information regarding L-PBF part quality. Moreover, the work herein describes approaches that are predominantly probabilistic, thus facilitating uncertainty quantification in machine-learnt predictions of L-PBF build quality.

Keywords Laser powder bed fusion · Automatic quality assessment · Machine learning · Uncertainty quantification

1 Introduction

The potential for Additive Manufacturing (AM) to reduce production steps and increase resource efficiency has encouraged the adoption of AM for serial production over conventional manufacturing processes such as milling, grinding, drilling, boring etc. Consequently, over a span of two decades, AM has become a multi-billion dollar industry [28]. The 2017 UK industrial strategy white paper [1] states that businesses have begun to exploit the potential of AM to make transformational improvements to productivity. Although the most widely used application of AM technologies so far has been rapid prototyping [28], the cost-effectiveness of the process and the ability of AM to fabricate geometrically complex and light-weight parts has increased the demand for AM-produced end-use products. These products include, for example, applications in aerospace and biomedical industries. Ensuring the quality of AM-produced parts is critical if we are to meet the certification constraints imposed by these sectors [17]. A lack of process robustness and repeatability has been noted as a major barrier that is preventing the full breakthrough of AM into risk-averse sectors [27, 29]. Thus, quality control in AM is an important issue which requires feasible solutions.

The current study focuses on laser powder bed fusion (L-PBF) processes (also known as Selective Laser Melting™ and Direct Metal Laser Sintering™), AM technologies which produce complex metallic parts from powder materials. The L-PBF process is a cycle of three steps. First, a powder deposition system deposits a thin layer of metal powder of 20–60 μm thickness. A laser then melts the powder following a predefined scanning path [22]. The bed which holds the part is then lowered and the cycle is repeated.

✉ Sarini Jayasinghe
  j.jayasinghe@liverpool.ac.uk

1 School of Engineering, University of Liverpool, Liverpool L69 3GH, UK

2 Renishaw Plc, New Mills, Kingswood, Wotton-under-Edge GL12 8JR, UK

Due to the layer-wise nature of the L-PBF process, defects are not always visible once the build has been completed. Consequently, CT scans are often employed post-build to identify defects. By introducing an on-line process monitoring system, part quality could be monitored during the build. This would allow the operator to assess the quality of the resulting parts without using relatively expensive post-build CT measurements. Moreover, such an on-line process monitoring system could, potentially, enable the implementation of automatic in-process corrective measures, thus facilitating process control for L-PBF. On-line process monitoring may also allow for faster part qualification in the R&D stage and decrease machine downtime [19].

Several studies have investigated the relationship between L-PBF process parameters, variations in photodiode signals and measures of final build quality. The work described in [26] classified parts according to their ultimate tensile strength via a semi-supervised machine learning algorithm, using features extracted from two co-axial photodiode sensors. Coeck et al. [7] analysed an off-axial photodiode sensor signal, extracting the co-ordinates of abrupt fluctuations ('DMP melt-pool events'). After exploring correlations between static tensile properties (ultimate tensile strength and plastic elongation) and DMP melt-pool events, an inverse relation between plastic elongation and the DMP melt-pool event density was observed in L-PBF parts. Coeck et al. [9] extended their study to automatically detect DMP melt-pool events in adjacent scan vectors. It was discovered that the location of DMP melt-pool event clusters were correlated to the pores found in the CT scans. The ratio between the readings from two co-axial photodiodes (each measuring different near-infrared wavelengths) was analysed in [2] under different process conditions. According to this study, as the energy per unit volume (transferred to the material) increases, both part density and photodiode signal ratio gradually increase until part density reaches 100%.

The aim of the current study is to investigate the feasibility of predicting L-PBF build density from a combination of different co-axial photodiode sensor measurements collected during the build process. Build density is used to quantify the quality of the parts, as it is a well-known indicator of global part quality [2]. It is noted that, while many developmental on-line process monitoring systems employ images collected via cameras to detect defects in L-PBF builds [3, 5, 8, 11, 16, 17, 21], using only photodiode measurements to monitor the build process is potentially beneficial as photodiode sensors are cost-efficient and typically have a higher sample rate compared to camera-based systems. The study involves using a Singular Value Decomposition (SVD) to extract features from datasets of photodiode measurements, before the extracted features are inputted into machine learning algorithms. The research exploits data collected during L-PBF builds fabricated in a Renishaw^{TM} AM 500M machine. Low density specimens are first identified via unsupervised clustering methods (K-means and a Gaussian mixture model (GMM)) before a supervised regression algorithm (Gaussian Process) is used to directly predict build density from the same extracted features.

The paper is organised as follows: "Literature review" further elaborates the motivation behind this research and discusses the state-of-the-art; "Experimental setup" provides the details of the experimental setup used for data collection; "Feature extraction" describes the feature extraction process and "Algorithm descriptions" introduces the machine learning methods used in the current paper; "Results and discussion" presents the results of the analysis. Finally, conclusions and suggestions for further extensions of the techniques are discussed in "Conclusions".

2 Literature review

According to Mani et al. [25] and Yadroitsev et al. [31], there are over 50 process parameters that can affect the final quality of L-PBF builds. The laser serves as the energy source in the heat transfer process, thus, parameters such as the peak power of the laser, pulse width, and pulse frequency impact the output of L-PBF builds. The focal point of the laser is moved across the build surface using a galvanometer scanner; the scan speed of the laser beam is also critical as it influences the energy applied to a particular spot of the build. Powder layer thickness, layer uniformity, powder temperature, and packing density can also significantly impact the heat transfer process [23].

The studies [2, 18] demonstrate an association between the energy density transferred to the material and L-PBF build density. According to Gu et al. [18] the energy density (i.e., energy per unit volume) transferred to the material is given by

$$E = \frac{P}{vht} \tag{1}$$

where P is the laser power, v is the scan speed, t is the layer thickness and h is the hatch distance. Consequently, the studies [2, 4, 18, 30] focused on the influence of these parameters on build quality. Arisoy et al. [4] discusses the effect of scan strategy, laser power, scan speed, and hatch distance on grain sizes, and showed that increasing the energy density results in larger grain sizes. The work described in [30] used scan speed and laser power to predict the porosity of metallic parts produced using L-PBF.

One common approach to melt-pool monitoring involves using camera-based sensing systems with CCD or CMOS detectors, which can achieve a relatively high spatial resolution. Such camera-sensing systems can allow

evaluation of the temperature profile and shape of the melt-pool [10]. The optical set-up suggested by Clijsters et al. [8] consisted of a high-speed NIR CMOS camera and a photodiode (sampling rate 10kHz) that was coaxial with the laser beam. Grasso et al. [17] utilised an off-axial camera sensing system with a CMOS detector which collected images at a rate of 300Hz. A pyrometer with an in-built CMOS detector (sampling rate 12.58Hz) was used by Khanzadeh et al. [21] to capture thermal images of the melt-pool.

Clijsters et al. [8] suggested an approach for in-situ monitoring where sensor measurements, taken from a photodiode and a high-speed NIR CMOS camera, were used to measure emitted light intensity and melt-pool area at a high sample rate (10kHz). Sensor data, mapped into a position-domain representation, were examined for anomalies using predefined expected values of emitted light intensity and areas of melt-pools for different scanning patterns (i.e., fill and contour). This system requires a position-dependent reference database unique to each part, generated by traditional validation techniques, to generate data representing expected behaviour. While the study stated that there is an excellent correlation between the detected errors and actual defects, the obtained accuracy was not reported.

A recent study carried out by Grasso et al. [17] used principal component analysis to compress image data of the melt-pool (collected via an off-axial camera sensing system at 300Hz) and output a statistical descriptor that can be used to identify defective areas. The proposed method was able to automatically detect and localise local defects related with overheating phenomena during the layer-by-layer L-PBF process.

A real-time porosity prediction method using morphological characteristics of the melt-pool boundary (captured using an IR camera with sampling rate 12.58Hz) was developed by Khanzadeh et al. [21]. The suggested method used a polar transformation to convert Cartesian co-ordinates of the melt-pool boundaries into polar co-ordinates (ρ, θ). The polar transformed melt-pool contours were represented as a function of θ. A functional principal component analysis (FPCA) was used to fit a cubic spline and extract key features that represent the morphological model of a melt pool. Defective and non-defective melt-pools were then classified using those key features. Melt-pools were classified with an accuracy level of 98.44% during fabrication of a thin-wall structure.

The studies [3, 11, 16] used off-axial high resolution imaging, obtained through visual cameras, to detect anomalies in L-PBF build layers. Aminzadeh *et al.* [3], captured build images using an 8.8 megapixel USB digital camera with high focus lenses while Gobert et al. [16] used a 36.3-megapixel DSLR camera (Nikon D800E) to take multiple images of the build layers.

The method suggested by Aminzadeh et al. [3], used a database of camera images with pre-identified zones (Zone I being where parts with no noticeable pores are created, Zone II being the 'high-energy zone' where small spherical, gas pores are created, and Zone III being the 'low-energy zone' where large irregular pores and lack of fusion occur) in layers of AM parts made with varying part quality, to train a Bayesian classifier. Appropriate features were selected by taking texture characteristics into consideration and by converting the images into the frequency domain. The resulting Bayesian classifier was able to achieve a 89.5% true positive rate and 82% true negative rate. However, due to the high resolution of the images used to capture the entire build layer, significant image post-processing was required due to the lack of image contrast between the part and powder.

Gobert et al. [16] proposed a method to identify discontinuities (e.g. porosity and cracks) in L-PBF builds. In the proposed approach the co-ordinates of anomalies and nominal voxels in the CT scan domain were first matched with layer-wise images. A binary classifier was then trained on features extracted from this matching layer-wise image stack to distinguish anomalies and nominal voxels. Discontinuities were detected with an 80% success rate, however, the approach heavily relied on embedded reference points for the calibration of coordinate transformations.

Table 1 summarises some of the state-of-the-art in-situ monitoring systems designed for the L-PBF build process. Table 1 also reports the data processing and analysis techniques utilised in those systems.

Of the aforementioned work, [2, 7, 9, 26] utilised photodiode sensors. Photodiodes are spatially integrated single-channel detectors that provide a voltage corresponding to the amount of light collected by the detector at each focal point. Cost effectiveness, high sensitivity, robustness, and fast data collection rates (typically $\sim$ 50kHz) make these devices attractive for L-PBF in-process monitoring. These sensors are, however, typically sensitive only over a limited range of wavelengths. Thermal radiation from the melt-pool is generally in the visible to IR range (900 to 2300 nm) while plasma emission is near UV or visible wavelengths (400 to 650 nm) [24]. Camera-sensing systems provide a more detailed overview of the melt-pool characteristics, but data management is a challenge as amount of data that can be collected via these sensors increases with the number of pixels in the image [6]. Consequently, choosing a sensor system for melt-pool monitoring is a trade-off between the response rate, cost and usefulness of the data that can be collected. Thus, many studies ([6, 8, 10], for example) exploit both photodiodes and CCD or CMOS cameras to capture the process dynamics of the melt-pool. However, the overall response rate of these combined monitoring system is low as the response rate of cameras is quite low compared to that of photodiodes [28].

Table 1 Selected literature on in-situ monitoring for automated defect detection in L-PBF

Reference	Process signature	Data processing technique	Analysing technique
Kruth et al. [8]	Melt-pool geometry and plume characteristics	Sensor data processed by a FPGA chip is mapped into a 2D regular grid, according to their logged positions, to assign the melt-pool data to a particular pixel	Statistical Process Control is carried out with reference data, generated by previous built parts using traditional validation techniques (CMM, CT, Archimedes, microscope)
Grasso et al. [17]	Melt-pool geometry and plume characteristics	A statistical descriptor based on principal component analysis (PCA) is applied to image data, which is suitable for identifying defective areas within a layer	K-means Clustering for automated defect detection
Fu et al. [15]	Temperature profile	A 3D finite element thermal model was developed to investigate the thermal process of L-PBF using ABAQUS software	Melt-pool shape and dimensions was predicted using process parameters and then verified with previously collected experimental data
Khanzadeh et al. [21]	Melt-pool geometry and temperature profile	Polar transformation was applied to position co-ordinates of the melt-pool boundaries and then functional principal component analysis was performed to extract key characteristics of those boundaries	Four different methods were used for melt-pool classification: K-nearest neighbour, support vector machine, decision tree, discriminant analysis
Foster et al. [14]	Plume characteristics	Images of layers were taken both immediately after re-coating and after laser exposure	Post-processed and analysed images were stacked to create 3D models to find anomalies
Craeghs et al. [11]	Plume characteristics	Melt-pool length and width of the captured melt-pool images were determined using image processing algorithms with the help of a FPGA chip	Change in length to width ratio is used detect process abnormalities
Aminzadeh et al. [3]	Plume characteristics	Features, carefully selected based on physical intuition into the process, were extracted from the images of L-PBF builds	Extracted features were classified using a Bayesian classifier
Gobert et al. [16]	Plume characteristics	Multi-dimensional visual features were extracted from layer-wise image stacks	Extracted features were evaluated using binary classification techniques, i.e. a linear support vector machine (SVM)
Coeck et al. [9]	Plume characteristics	Identifying position co-ordinates of melt-pools where the signal is fluctuating within a scan vector	Identified defective melt-pools within close proximity were utilised to detect defective

The current study explores the feasibility of using a set of measurements collected through different co-axial photodiode sensors to predict the density of L-PBF builds without using camera-based sensors. The overall aim of this paper, then, is to develop a model that can predict L-PBF build quality via build-data acquired through photodiode sensors.

3 Experimental setup

Sixty-two Inconel718 cubes were build as test samples for this study using a Renishaw[TM] AM 500M machine.

The machine sensor system, shown in Fig. 1, consists of three high precision co-axial single-channel photodiodes. Two of the photodiodes are designed to capture melt-pool plume characteristics while the remaining sensor is designed to measure the intensity of the laser beam.

Photodiode-1 (PD1 - no. 4 in Fig. 1, wavelength − 700 to 1050 nm) is sensitive to plasma emissions whereas photodiode-2 (PD2 - no. 5 in Fig. 1, wavelength − 1080 to 1700 nm) is sensitive to thermal radiation. A third photodiode, photodiode-3 (PD3 - no. 10 in Fig. 1) measures the intensity of the laser beam. The optical window (no. 16 in Fig. 1), which is used to focus the laser beam, exhibits > 95% spectral transmission across the wavelength of interest. The machine has

a Galvo-scanner system (no. 19 in Fig. 1) which controls the movement of the laser focal point, such that it follows pre-defined (x, y) coordinates. In the following experiments, the (x, y) coordinates of the laser focal point were collected alongside the photodiode measurements at 100 kHz.

Build density of the specimens was measured using a 2D microscopy technique (OGP smartscope Zip Lite 300 at 75 magnification). As shown in Fig. 2, the edges of the images
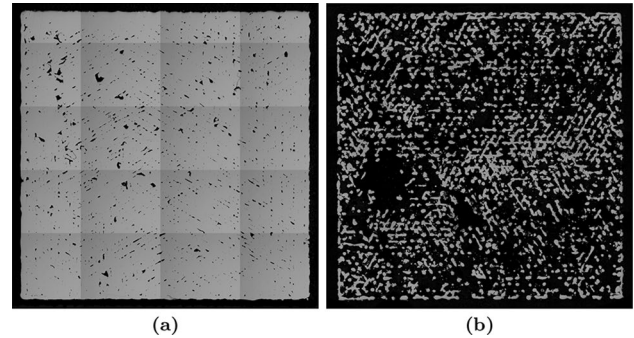

Fig. 2 2D image captured by OGP smartscope Zip Lite 300 microscope at 75 magnification **a** specimen 3 with edited edges, **b** specimen 13 with edited edges
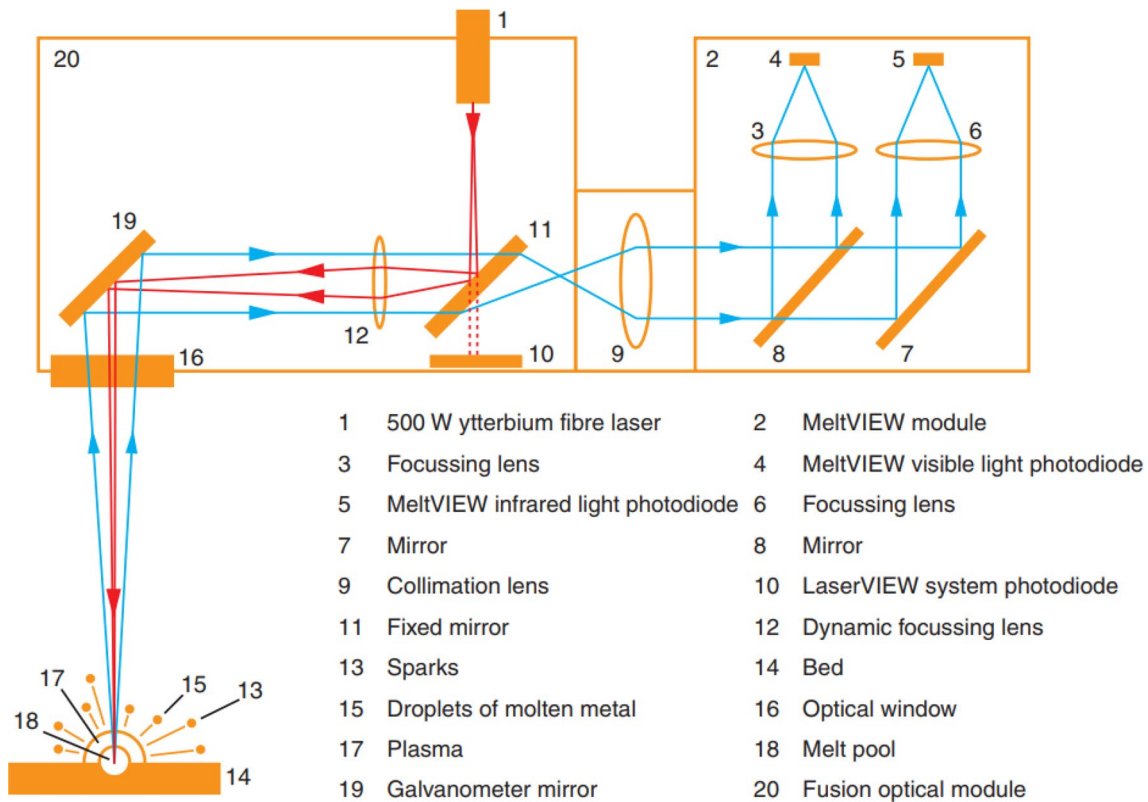

1	500 W ytterbium fibre laser	2	MeltVIEW module
3	Focussing lens	4	MeltVIEW visible light photodiode
5	MeltVIEW infrared light photodiode	6	Focussing lens
7	Mirror	8	Mirror
9	Collimation lens	10	LaserVIEW system photodiode
11	Fixed mirror	12	Dynamic focussing lens
13	Sparks	14	Bed
15	Droplets of molten metal	16	Optical window
17	Plasma	18	Melt pool
19	Galvanometer mirror	20	Fusion optical module

Fig. 1 Renishaw[TM] AM 500M machine schematic assembly of the optical sensing system

were filled using photoshop, so that the edges of the builds were not included in estimates of part bulk density.

4 Feature extraction

Sixty-two Inconel 718 cubes (Fig. 3a) were built using different combinations of process parameters (i.e., laser power, scan speed, exposure time and hatch distance) according to a Design of Experiments (DoE) procedure. This DoE procedure was carried out following standard practise implemented by Renishaw.

Figure 3b shows the (x, y) coordinates of the laser point over a single layer (the first layer created after the initial support layers) of the build. Data points recorded when the laser was off (when the laser was moving in between specimens

and in between hatches) and when the laser was used on contours (edge of each specimen) was removed, as the technique proposed in this paper is focused on bulk density only. When the laser focal point was moved across the layer, each specified location was exposed to the laser beam for a pre-defined time period (exposure time), and the maximum intensity of the laser beam per each exposure, captured by PD3, was recorded alongside the corresponding PD1 and PD2 sensor readings. The sensor data was separated into 62 data sections, each corresponding to a single cube. The data separation was carried out by creating polygons marking the boundaries of each cube (see black outlines in Fig. 3b) and then indexing all the data points within each area as belonging to the same set. Examples of the resulting data segments are shown in Fig. 4a. The signal-to-noise ratio in decibels (SNR) of signals captured from PD1, PD2, and PD3 were 10.05, 14.10, and 16.66 respectively. The
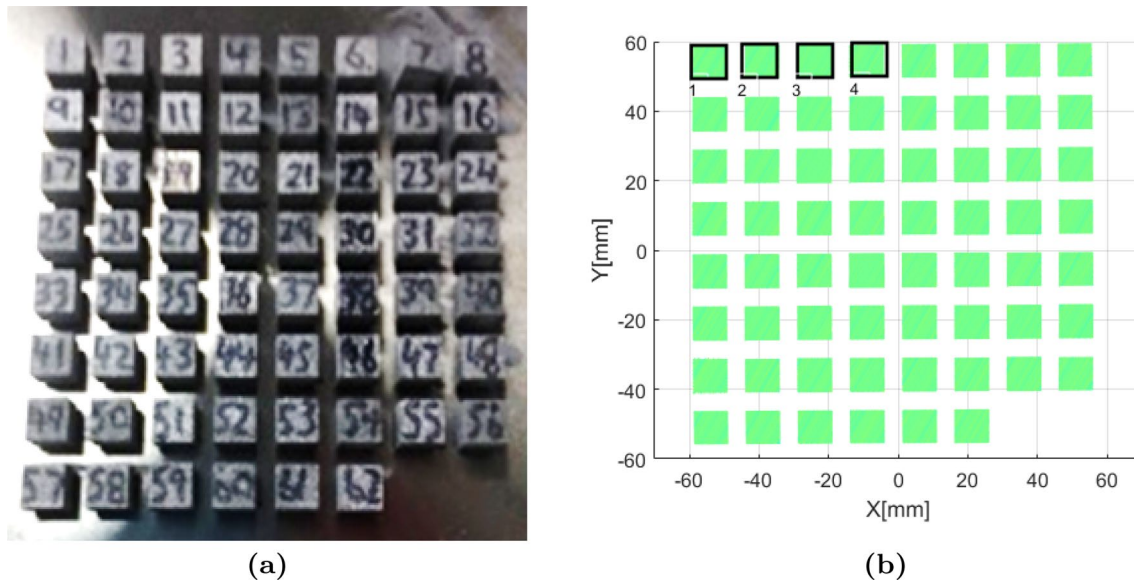

(a)

(b)

Fig. 3 **a** 62 cubes built with different combinations of process parameters such as laser power, scan speed, point distance and exposure time, **b** position coordinates of the laser point representing a single layer of the build

Fig. 4 **a** The separation of the signal data was carried out by indexing all the data points with respect to their location on the layer, **b** The number of signal samples corresponding to each specimen, for a single layer (the first layer created after the initial support layers) of each specimen

(a)

(b)
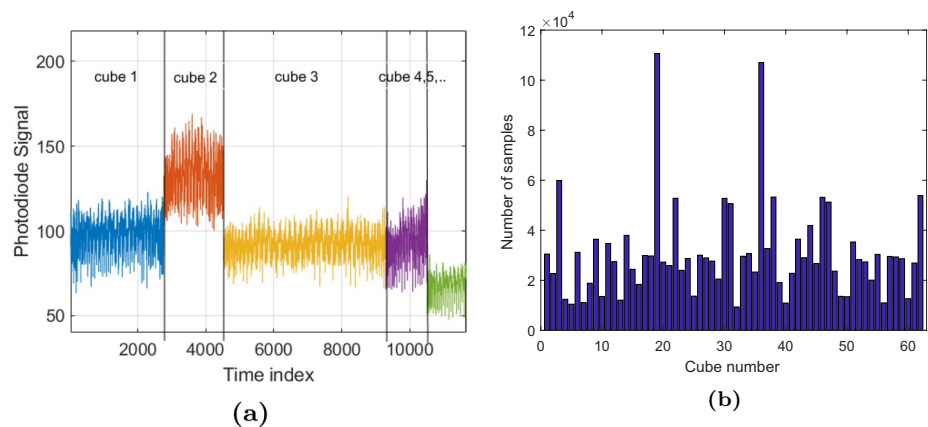
SNR of the signals, captured on each specimen, vary from one specimen to another as these specimens were fabricated using different process parameters. The aforementioned SNR values were calculated using signal mean and standard deviation captured on the specimens associated with high density values, SNR = 10 log ($\mu_{\text{signal}}/\sigma_{\text{signal}}$). While some data was removed when a single reading per exposure was extracted, the number of data points corresponding to each cube was recorded separately, as this value is proportional to the time spent by the laser on each bounded area; the number of data points collected per cube is shown in Fig. 4b. Subsequently, for each photodiode, the sensor data recorded for each cube was arranged into separate columns, creating three separate data matrices (one per photodiode). To ensure that these matrices had the same number of elements in each column, the minimum column length across each specimen was determined before the data was truncated accordingly. To ensure that these matrices had the same number of elements in each column, the minimum column length across each specimen was determined before the data was truncated accordingly. The minimum column length and number of specimens were equal to 1216 and 62 respectively, hence each of the resulting data matrices was of dimension 1216×62. A Singular Value Decomposition (SVD) was then performed on each data matrix to extract the features that, later in this study, are used as inputs to various machine learning approaches.

In the following, to illustrate the feature extraction procedure, we use A to denote a data matrix obtained from one of the 3 photodiode sensors. An SVD takes a data matrix, A, and factorises it into a product of three separate matrices:

$$A = U \Sigma V^T \tag{2}$$

If the original data matrix A is ($m \times n$), then U will be an ($m \times m$) orthogonal matrix and V will be a ($n \times n$) orthogonal matrix. Σ, an ($m \times n$) matrix, has the form

$$\Sigma = \begin{bmatrix} \Lambda \\ 0 \end{bmatrix}, \qquad \Lambda = \text{diag}(\Sigma_1, \Sigma_2, ..., \Sigma_n) \tag{3}$$

where the diagonal elements of Λ are sorted from largest to smallest. These diagonal elements are the square root of the eigenvalues of $A A^T$ and are commonly referred to as 'singular values'.

As a result of the SVD, each column of A can be expressed as a linear combination of basis vectors ($u_1, u_2, ..., u_m$). Specifically, by defining $C = \Sigma V^T$, one can write

$$A = UC \Rightarrow a_i = \sum_{j=1}^{m} u_j c_{j,i} \tag{4}$$

where a_i is the ith column of A and u_j is the jth column of U. Thus, each vector a_i is now associated with a constant $c_{j,i}$ for every basis vector u_j. Often, only a small subset of

singular values are significantly different from zero. In such cases, the sum in Eq. (4) can be truncated and each column of A can be approximated using only a limited number of $k < m$ basis vectors, $u_1, u_2, ..., u_k$. Specifically, by writing $\bar{C} = \Sigma_k V_k^T$ where V_k is an ($n \times k$) matrix, composed of the first k columns of V, and Σ_k represents a ($k \times k$) matrix which includes the first k singular values ($\Sigma_1, \Sigma_2 .... \Sigma_k$), we can write

$$A \approx \bar{A} = U_k \bar{C} \Rightarrow \bar{a}_i = \sum_{j=1}^{k} u_j \bar{c}_{j,i} \tag{5}$$

The difference between A and $\bar{A}$ can be quantified using the normalised Frobenius norm [20]

$$\|A - \bar{A}\|_{\text{Fro}} = \sqrt{\frac{\Sigma_{i=1}^n \Sigma_{j=1}^m (|a_{j,i} - \bar{a}_{j,i}| / |a_{j,i}|)^2}{mn}} \tag{6}$$

For each of the three data matrices, the number of basis vectors (k) employed for generating the approximated matrix $\bar{A}$ was increased up until $k=15$ and the Frobenius norm was used to quantify the contribution of each basis vector to the matrix $\bar{A}$.

Figure 5 shows how the Frobenius norm decreases as the number of basis vectors increases for all three data matrices. It can be seen that the approximation realised using only the first basis vector captures most of the data - in fact, the Frobenius norm for photodiode-1 is only 0.0934 (for context the Frobenius norm when $\bar{A} = 0$ is 1). Similarly, the Frobenius norm for photodiode-2 and photodiode-3, using 1 basis vector only, is 0.0388 and 0.0218 respectively. In the following, the constants relating to the first basis vectors of the three matrices are therefore used as extracted features to be inputted into the machine learning algorithms.

5 Algorithm descriptions

This section describes the supervised and unsupervised machine learning approaches that were used in the current study to classify and predict part density. The constants $c_{j,i}$ (defined in Eq. (5)) that are related to the first basis vectors of photodiode-1 (c^{pd1}), photodiode-2 (c^{pd2}) and photodiode-3 (c^{pd3}), as well as the total number of scan samples (s) used within each bounded area (as this is proportional to the time spent by the laser on each specimen) were used to create different feature spaces. Herein, a single component of the feature space is represented by the vector $x_i = \left(x_i^{(1)} . . x_i^{(D)} \right)^T$, where D is the dimension of the feature space and i is used to index each specimen. As an example, x_i can be a combination of c_i^{pd1}, the constants that are related to the first basis vectors of photodiode-1,
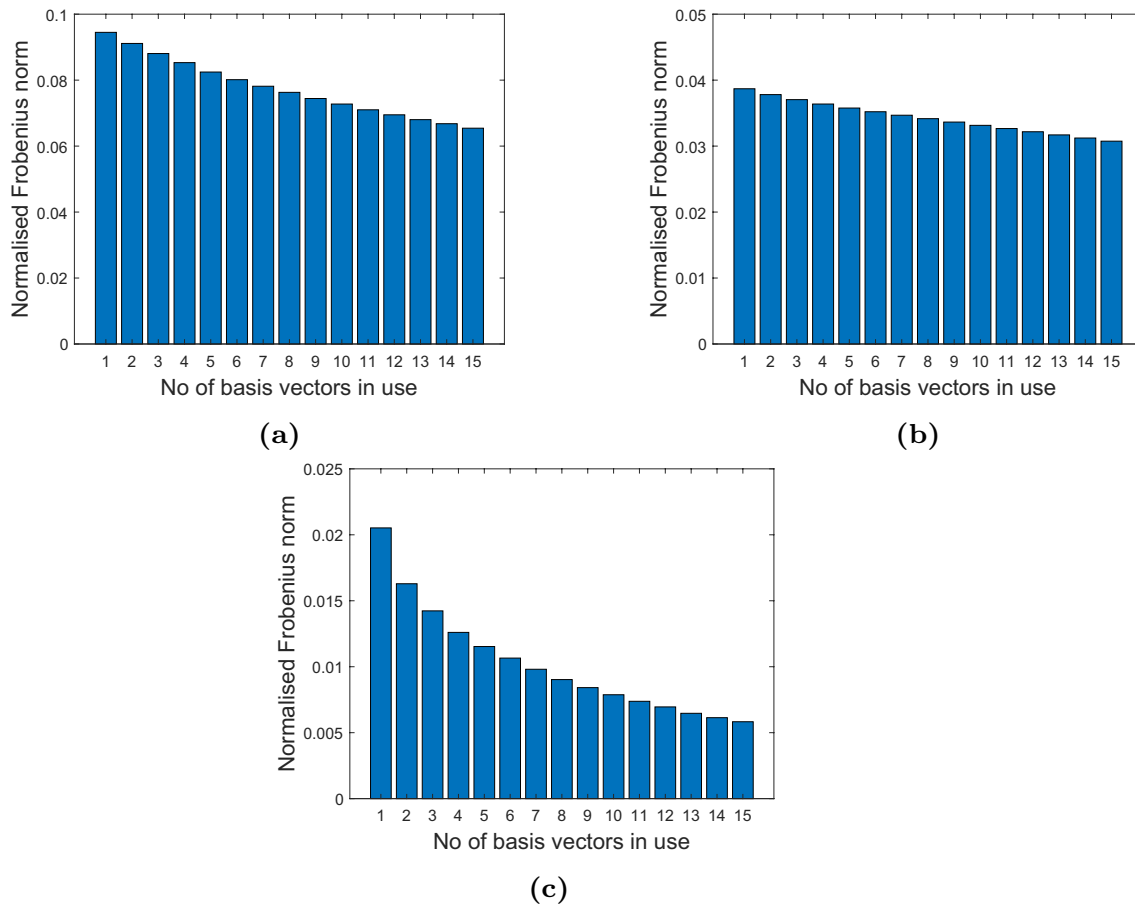
(a)

(b)

(c)

Fig. 5 Normalised Frobenius norm calculated using the first 15 basis vectors (out of a total of 1216 basis vectors) for **a** Photodiode-1, **b** Photodiode-2, **c** Photodiode-3

and c_i^{pd2}, the constants that are related to the first basis vectors of photodiode-2, which provides a 2D feature space (such that $D = 2$). In the following, all the feature vectors ($\boldsymbol{c}^{\text{pd1}}$, $\boldsymbol{c}^{\text{pd2}}$, $\boldsymbol{c}^{\text{pd3}}$, $\boldsymbol{s}$) were standardised before being used to create feature spaces.

In the current work, the ability of two different unsupervised learning approaches (K-means clustering and a Gaussian Mixture Model (GMM), described in "Unsupervised learning") to cluster the feature space was assessed. Subsequently, it was investigated whether supervised learning (Gaussian Process (GP) regression with Automatic Relevance Determination, described in "Supervised learning") could be used to directly predict sample density using the same feature vectors.

5.1 Unsupervised learning

The two unsupervised learning approaches used in this study are briefly explained below.

5.1.1 K-means algorithm

Given a set of n observations $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ (in this case representing n points in the feature space, $n = 62$), K-means clustering aims to partition the observations into sets $\boldsymbol{h}_r = \boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_q$ ($q \leq n$). The K-means algorithm begins with an initial estimate of cluster centres ($\boldsymbol{o}_r$ where $r = 1, 2 \ldots, q$) before calculating the distance between each point and each centre. Each observation is then assigned to the cluster whose centre is closest. In the next iteration of the algorithm, the centres are recomputed by setting them equal to the mean of the newly classified points. The process of estimating cluster centres and assigning observations to clusters is then repeated until the cluster centers are judged to have converged. While more computationally efficient versions of K-means have been proposed in [12], they were not found to be necessary in the current study where only 62 data points were available.

5.1.2 Gaussian mixture model

K-means is a deterministic approach, based on the assumption of spherical clusters. Gaussian Mixture Models (GMMs) are probabilistic and can be used to identify non-spherical clusters. GMMs therefore provide more flexibility compared to the K-means algorithm. Within the GMM framework, it is assumed that each feature vector has been generated from one of H Gaussian distributions, such that

$$p(\mathbf{x}_i) = \sum_{b=1}^{H} \pi_b \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \; ; \; \sum_{b=1}^{H} \pi_b = 1 \tag{7}$$

where the π_b's are called *mixture proportions* and b indexes each cluster. The parameters $\boldsymbol{\mu}_b$ and $\boldsymbol{\Sigma}_b$ represent the mean and the covariance matrix of the b th Gaussian respectively.

Training a GMM involves identifying the parameters $\theta_{GMM} = \{\pi_b, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b\}$, where $b = 1, 2, ..., H$, which maximise the likelihood of witnessing the data $\mathbf{x}_1, .., \mathbf{x}_n$. This estimation procedure can be undertaken using the Expectation Maximisation (EM) algorithm, which consists of two main steps: the expectation step (E-step) and the maximisation step (M-step). The EM algorithm begins by randomly initialising the GMM parameters for each cluster. Given these parameters, the posterior probability that cluster c is responsible for generating $\mathbf{x}_i$ is calculated as part of the E-step. To do so, let us assume that each data point $\mathbf{x}_i$ is given a 'label' $\mathbf{z}_i \in \mathbb{R}^H$ describing which of the Gaussians from the mixture was used to generate $\mathbf{x}_i$, where the labels are such that $z_{i,b} \in \{0, 1\}$ subject to $\sum_{b=1}^{H} z_{i,b} = 1, \; \forall i$. These labels are hidden to the user (and are named 'latent variables') and they indicate the cluster each observation originates from ($z_{i,b} = 1$ means that the point x_i came from the b th Gaussian with probability 1). The expected value of $z_{i,b}$, at the t th iteration of the algorithm, is

$$\gamma(z_{i,b})^{(t)} = \frac{\pi_b \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_b^{(t)}, \boldsymbol{\Sigma}_b^{(t)})}{\sum_{b'=1}^{H} \pi_{b'}^{(t)} \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_{b'}^{(t)}, \boldsymbol{\Sigma}_{b'}^{(t)})} \tag{8}$$

Fixing each $z_{i,b}$ equal to its expected value, in the M-step, a new set of parameters for the Gaussian distributions are computed such that they maximise the lower bound of the log likelihood of all the observations in each cluster. This is called the M-step. It can be shown that, at the tth iteration of the algorithm, the M-step for a GMM involves setting

$$\boldsymbol{\mu}_b^{(t+1)} = \frac{1}{n_b} \sum_{i=1}^{n} \gamma(z_{i,b})^{(t)} \mathbf{x}_i \tag{9}$$

$$\boldsymbol{\Sigma}_b^{(t+1)} = \frac{1}{n_b} \sum_{i=1}^{n} \gamma(z_{i,b})^{(t)} (\mathbf{x}_i - \boldsymbol{\mu}_b^{(t+1)})(\mathbf{x}_i - \boldsymbol{\mu}_b^{(t+1)})^T \tag{10}$$

$$\pi_b^{(t+1)} = \frac{n_b}{n} \tag{11}$$

where

$$n_b = \sum_{i=1}^{n} \gamma(z_{i,b})^{(t)} \tag{12}$$

Once these parameters are calculated, the log likelihood is evaluated as

$$\ln p(\mathbf{x}_1, ..., \mathbf{x}_n \mid \boldsymbol{\mu}, \hat{\boldsymbol{\Sigma}}, \boldsymbol{\pi}) = \sum_{i=1}^{n} \ln \left\{ \sum_{b=1}^{H} \pi_b \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) \right\} \tag{13}$$

The log likelihood given in Eq. (13) can be used to analyse convergence of the EM algorithm as it runs over a number of iterations.

5.2 Supervised learning

In supervised learning, algorithms learn from labelled data. The approach used in this study, namely Gaussian Process regression, is briefly explained below.

5.2.1 Gaussian process regression

Let us assume there exists a function, f, that can predict the density, d, of each specimen from the extracted features, $\mathbf{x}$. However, f is unknown and must be estimated from the data. Here, it is assumed that

$$d_i = f(\mathbf{x}_i) + \epsilon, \qquad \epsilon \sim \mathcal{N}(\epsilon; 0, \sigma^2) \tag{14}$$

such that each observation, d_i, is equal to the function f evaluated at input $\mathbf{x}_i$ but corrupted with zero mean Gaussian noise of variance σ^2. By defining

$$\mathbf{f} = (f_1, f_2 ......, f_n)^T, \qquad f_i \equiv f(\mathbf{x}_i) \tag{15}$$

then, from the definition of a Gaussian process, $p(\mathbf{f})$ is a Gaussian whose mean is zero and whose co-variance is defined by a Gram matrix, $\mathbf{K}$, so that

$$p(\mathbf{f}) = \mathcal{N}(\mathbf{f}; \mathbf{0}, \mathbf{K}) \tag{16}$$

The matrix $\mathbf{K}$ is defined using a kernel function, k, chosen to ensure that $\mathbf{K}$ is a valid covariance matrix. In this study, the considered kernel is

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2} \sum_{j=1}^{D} \frac{1}{L_j^2} (x_j - x_j')^2 \right) \tag{17}$$

where x_j represents the jth component of $\mathbf{x}$. The 'hyperparameters' L_j ($j = 1, ..., D$) and σ are collected together into the vector $\theta_{GP} = (L_1, L_2, ..., L_D, \sigma)$.

The hyperparameter L_j controls the lengthscale of the kernel function for the jth input dimension. By allowing different lengthscales for each dimension of the input space, it is possible to control the sensitivity of the regression model to each input. The identification of the hyperparameteres $L_1, ..., L_D$ is referred to as Automatic Relevance Determination (ARD), as it helps to establish the relevance of different inputs to the model predictions. If, for example, the jth feature has no relevance for the regression problem, the optimum lengthscale, L_j, will be relatively large (thus reducing the sensitivity of the regression model to that feature). ARD can therefore be used to discard features that are less relevant.

From the noise model described by Eq. (14), the likelihood of observing $d = (d_1, ..., d_n)^T$, conditional on f, is

$$p(d \,|f) = \mathcal{N}(d \,|f, I\sigma^2) \tag{18}$$

where I is the identity matrix. By marginalising over f, the marginal likelihood is

$$p(d) = \int p(d\,|f)p(f)df = \mathcal{N}(d\,|0, K + \sigma^2 I) \tag{19}$$

Given the training dataset, the optimum hyperparameters θ_{GP} which maximise the marginal log likelihood can be estimated using gradient-based optimisation algorithms (e.g. conjugate gradients [13]). Once the optimum hyperparameters are identified, in response to a new input x_* (generated from specimens that was not included in the training data, for example), the probability that this sample has measured density d_* is given by

$$p(d_*|d) = \mathcal{N}(d_*|m_*, \sigma_*^2) \tag{20}$$

where

$$m_* = k^T G^{-1}d, \qquad \sigma_*^2 = \sigma^2 + k(x_*, x_*) - k^T G^{-1}k. \tag{21}$$

and

$$G = K + \sigma^2 I, \qquad k_i = k(x_i, x_*) \tag{22}$$

.

6　Results and discussion

As mentioned in "Algorithm descriptions", the constants relating to the first basis vectors of the data matrices of Photodiode-1 (c^{pd1}), Photodiode-2 (c^{pd2}), Photodiode-3 (c^{pd3}) and the total number of scan samples (s) of each specimen were used to create different feature spaces.

Initially, all the combinations of feature pairs were taken into consideration. Figs. 6 and 7 shows several combinations of pairs of feature vectors, plotted on a 2-dimensional
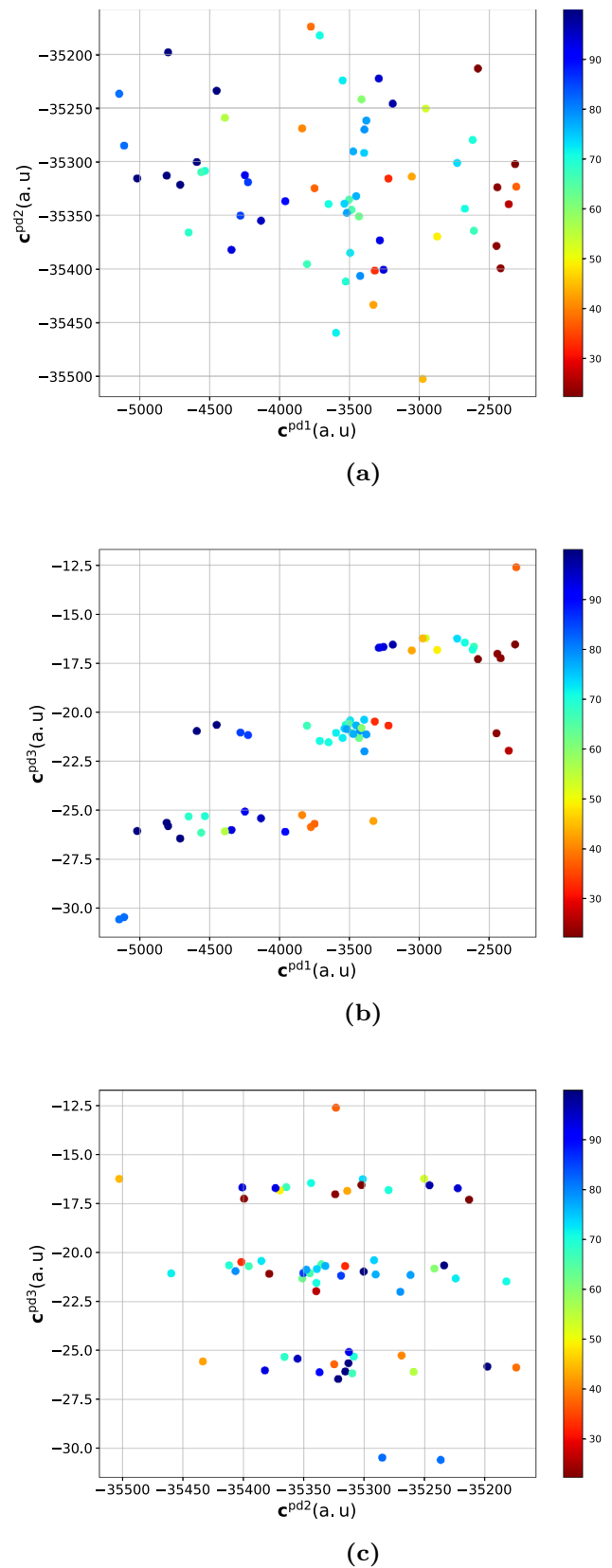

(a)

(b)

(c)

Fig. 6 **a** Features extracted from photodiode-1 (c^{pd1}) plotted against features extracted from photodiode-2 (c^{pd2}), **b** c^{pd1} plotted against c^{pd3}, **c** c^{pd2} plotted against c^{pd3} and coloured depending on the density (0–100%) of each specimen

Fig. 7 a Features extracted from photodiode-1 (c^{pd1}) plotted against ▶ number of signal samples collected on each specimen (**s**), **b** c^{pd2} plotted against **s**, **c** c^{pd3} plotted against **s** and coloured depending on the density (0–100%) of each specimen

space, where the data points have been coloured depending on the density of the corresponding specimen.

When observing the distribution of data points in Figs. 6 and 7, it is evident that when the coefficients relating to the first basis vector of photodiode-2 (c^{Pd2}) and photodiode-3 (c^{pd3}) are plotted with the coefficients relating to the first basis vector of photodiode-1 (c^{pd1}) (Fig. 6a and b), or when c^{pd2} and c^{pd3} are plotted against each other (Fig. 6c), the categories of density do not appear in separate clusters. Again, when c^{pd2} are plotted against the total number of scan samples (**s**), there are no evident clusters (Fig. 7b). However, when c^{pd1} are plotted against **s** (Fig. 7a), the categories of density appear to be better separated. Furthermore, when c^{pd3} are plotted against **s**, there are also roughly recognisable clusters (Fig. 7c). The ability of these feature combinations to visualise data points in separate clusters is summarised in Table 2.

6.1 Unsupervised learning

In the following, specimens with a density higher than or equal to 99% were categorised as 'high density' parts and the specimens with a density less than 99% were categorised as 'low density' parts. Unsupervised learning was therefore applied to a binary classification problem, such that each specimen could be classified as either 'low density' or 'high density'.

The feature combinations with recognisable clusters (c^{pd1} plotted against **s** and c^{pd3} plotted against **s**) were used as input features for the unsupervised machine learning algorithms. It is noted that the number of samples was always included in the feature space. For the sake of completeness, several features were added to the feature pairs recognised in the previous stage to be successful in separating categories of densities, creating both 3D and 4D feature spaces.

In the following, 50% of the available data was used for algorithm training such that the remaining 50% was used to test each algorithm's predictive capabilities. The training and testing datasets therefore consisted of 31 data points each. 2-fold cross validation was performed whereby the role of the two datasets was reversed (such that the training data became the testing data and *vice-versa*), before the predictive ability for each fold was averaged to estimate the final accuracy level. Specifically, the average accuracy level was calculated as
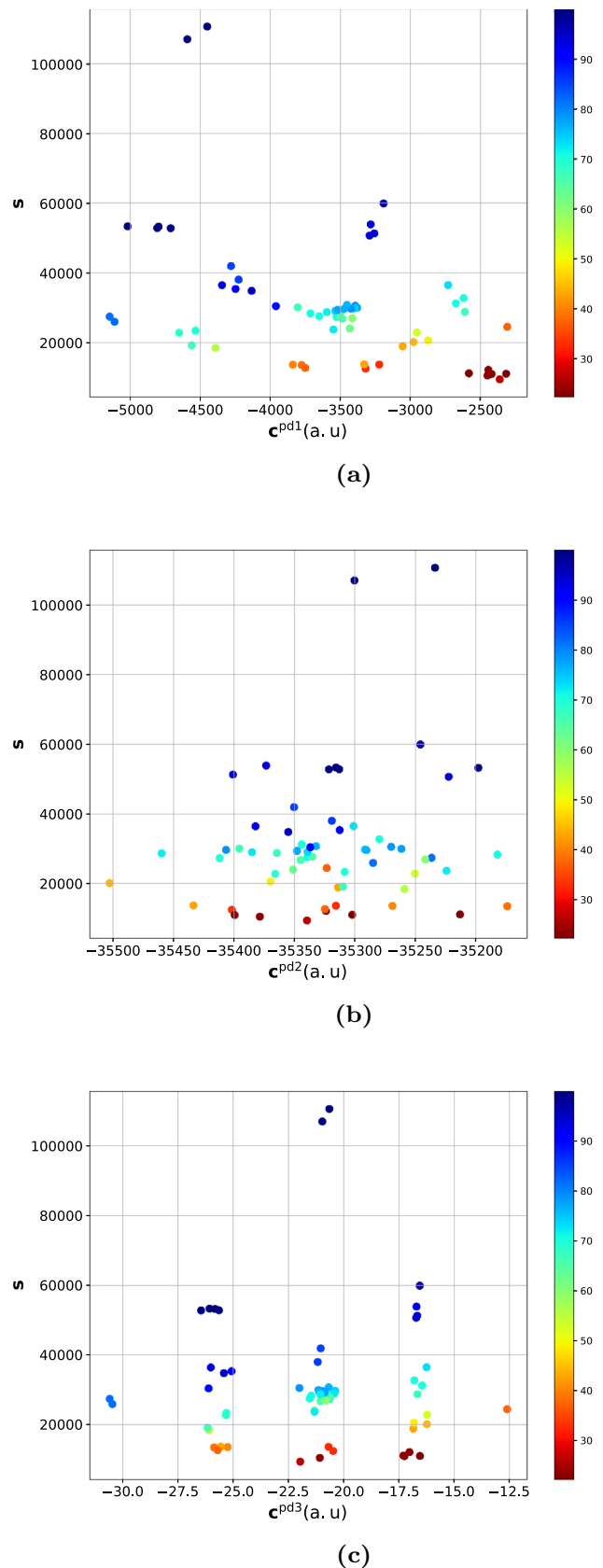
Table 2 Recognising clusters in the 2D feature space

Extracted feature vectors				Evident clusters
c^{pd1}	c^{pd2}	c^{pd3}	s	
•	•			✗
•		•		✗
	•	•		✗
•			•	✓
	•		•	✗
		•	•	✓

Table 3 Classification accuracy calculated using Eq. (23)

Normalised extracted features				K-means (%)	GMM (%)
c^{pd1}	c^{pd2}	c^{pd3}	s		
✓			✓	80.65	93.54
	✓		✓	77.42	93.54
✓	✓		✓	74.19	79.03
	✓	✓	✓	83.87	83.87
✓		✓	✓	75.81	90.03
✓	✓	✓	✓	66.13	79.03

$$\text{Avg. success rate\%} = \frac{\sum_{i=1}^{2}\left(\frac{N_{correct,i}}{N_{total,i}}\right) \times 100}{2} \quad (23)$$

where $N_{correct,i}$ is the number of correctly predicted specimens and $N_{total,i}$ is the total number of specimens in the ith fold.

We note that, with the GMM, the 'threshold probability' was set equal to 0.5 such that if the probability of a particular data point being in one particular cluster is more than 0.5, the point is considered part of that cluster. We first used K-means to cluster the feature space, as it provides a starting point (mean initialisation) for the GMM and because the produced results provide a reference to validate the subsequent GMM results.

All the feature space combinations and resulting accuracy levels are shown in Table 3. It can be observed that, in the case of the GMM, the highest success rate (93.54%) was achieved when c^{pd1} and s were used as input features. The same success rate was obtained when c^{pd3} and s were used as input features. Comparing the results obtained with K-means clustering for those feature pairs, the highest accuracy was achieved when c^{pd1} and s were used as input features. Adding c^{pd2} to the feature pair $\left(c^{pd1}, s\right)$ was detrimental, as the success rate of both clustering techniques was reduced. Adding coefficients relating to c^{pd3} to the same feature pair $\left(c^{pd1}, s\right)$ also slightly reduced the success rate of both clustering techniques. Therefore, in the following, we use the features relating to PD1 and number of samples (c^{pd1} plotted against s) to further quantify the performance of the GMM.

Figures 8 and 9 show the results obtained for validation data in the first and second fold, respectively, when s and c^{pd1} were clustered using the GMM. Figs. 8a and 9a illustrate the measured part density (parts with high density in green and low density in red) while Figs. 8b and 9b illustrate the GMM results.

To test whether similar performance could be obtained by considering only normalised process parameters, and in particular laser speed and power (as they are directly linked
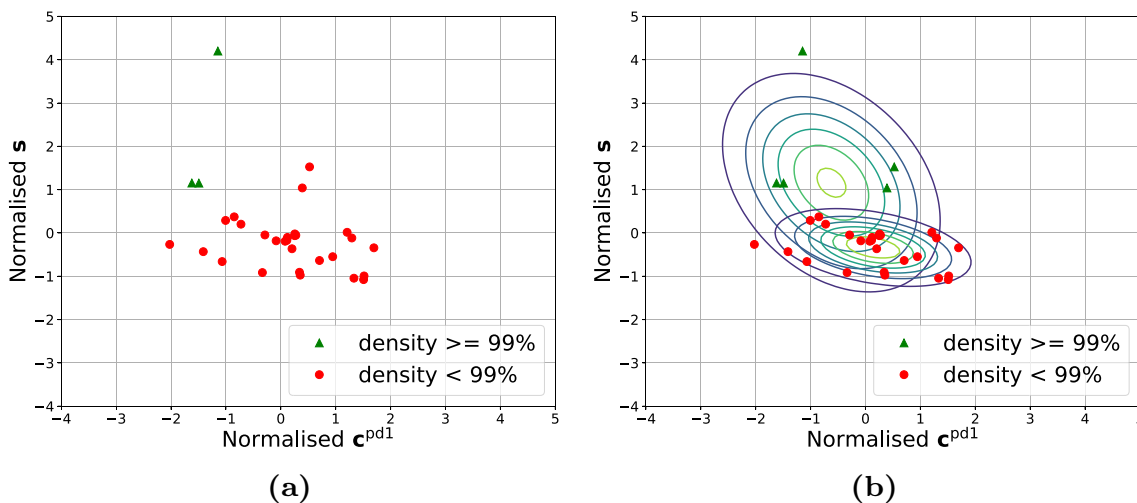

Fig. 8 Combinations of feature vectors, plotted in a 2-dimensional space, where the data points have been coloured depending on the density of the each specimen. Normalised c^{pd1} vs. normalised s plotted for the validation data in the first fold of the cross-validation **a** colour green represents samples with measured density ≥ 99% and red represents samples with measured density < 99%, **b** colour green represents cluster-1 (predicted density ≥ 99%) and red represents cluster-2 (predicted density < 99%)

Fig. 9 Normalised c^{pd1} vs. normalised **s** plotted for the validation data in the second fold of the cross-validation **a** colour green represents samples with measured density $\geq 99\%$ and red represents samples with measured density $< 99\%$, **b** colour green represents cluster-1 (predicted density $\geq 99\%$) and red represents cluster-2 (predicted density $< 99\%$)
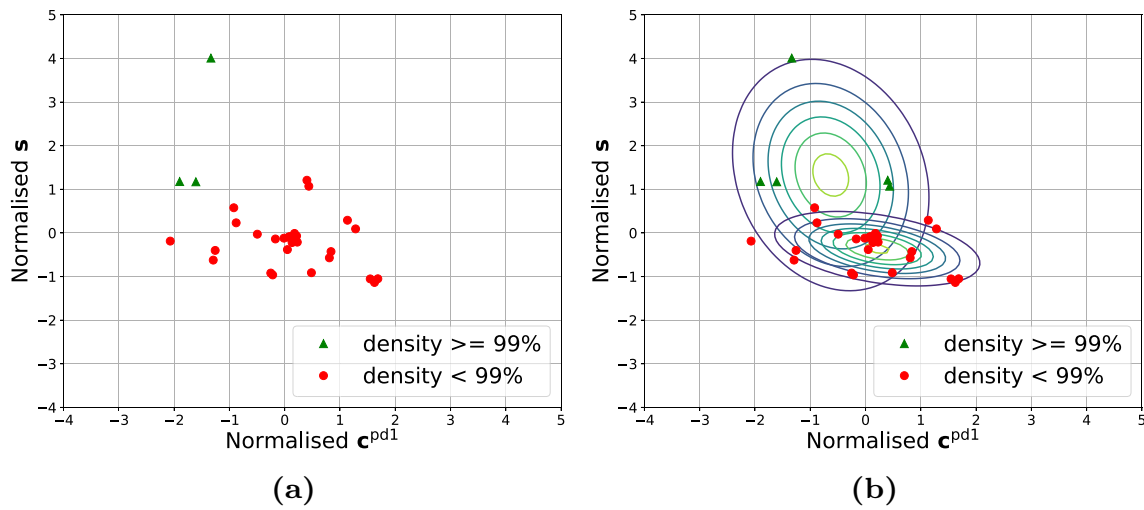
to energy density in Eq. (1)), clustering was performed using such parameters as input features. The results, shown in Figs. 10 and 11 (laser speed vs. laser power), highlight that the accuracy of classification went down to 29.03%. When all the process parameters (laser power, scan speed, point distance and exposure time) were used the accuracy increased up to 61.29%, still much lower than the 93.54% obtained using sensor data. Such results highlight that sensor readings contain additional useful information compared to the L-PBF process parameters.

To further quantify the performance of the GMM, a Receiver Operating Characteristic (ROC) curve was created. A Receiver Operating Characteristic curve is a

graphical plot that illustrates the diagnostic ability of a probabilistic classification algorithm as the algorithm's threshold probability is varied from 0 to 1. To create a ROC curve, a classifier's False Positive Rate (FPR) and True Positive Rate (TPR) are recorded for a range of different threshold probabilities before being plotted against one another. For the current example, the TPR is defined as the ratio of correctly identified low density parts, relative to the total number of low density parts. Likewise, the FPR is defined as the ratio of falsely identified low density parts, relative to the total number of high density parts. The Area Under the Curve (AUC) represents a measure of separability. The closer the AUC is to the
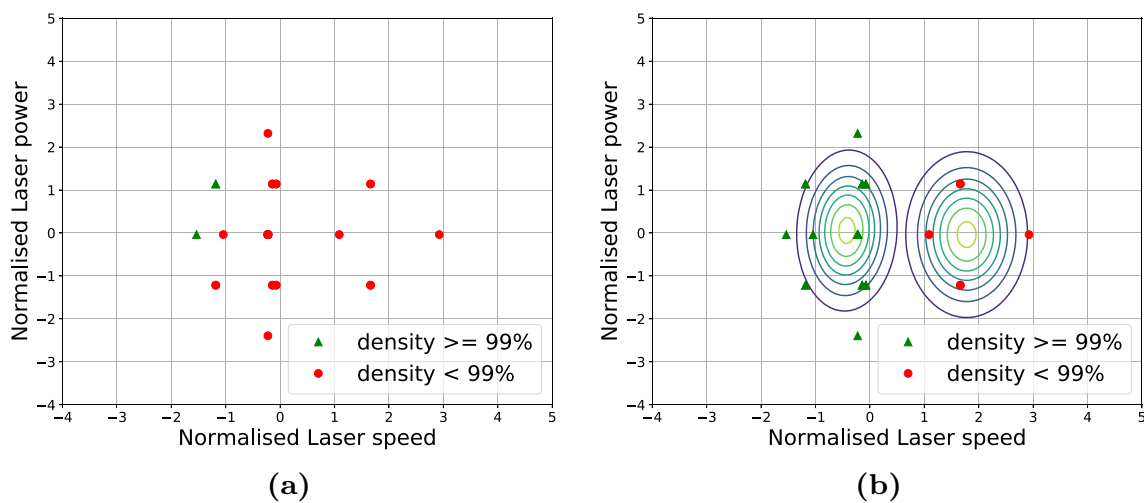

Fig. 10 Normalised laser speed vs. normalised laser power plotted for the validation data in the first fold of the cross-validation. **a** colour green represents samples with measured density $\geq 99\%$ and red represents samples with measured density $< 99\%$, **b** colour green represents cluster-1 (predicted density $\geq 99\%$) and red represents cluster-2 (predicted density $< 99\%$)
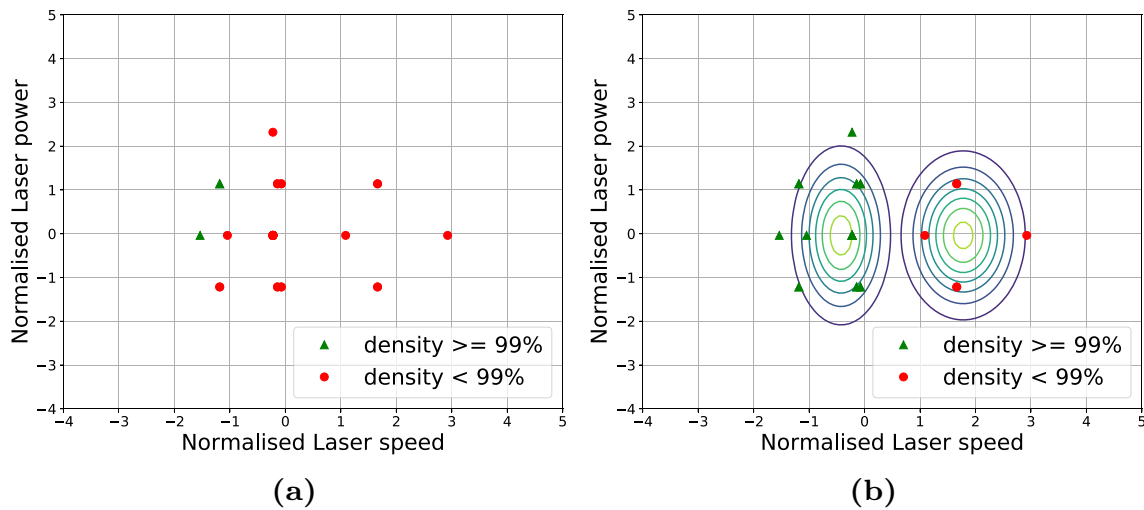
Fig. 11 Normalised laser speed vs. normalised laser power plotted for the validation data in the second fold of the cross-validation. **a** colour green represents samples with measured density ≥ 99% and red rep- resents samples with measured density < 99%, **b** colour green repre- sents cluster-1 (predicted density ≥ 99%) and red represents cluster-2 (predicted density < 99%)

top-left of the plot, the better the model is at distinguishing between parts with high density and low density. AUC = 0.5 indicates that the model has no class separation capac- ity, while AUC = 1.0 indicates that the model can separate the classes with 100% accuracy.

To analyse the performance of the GMM for a variety of threshold probabilities, an ROC curve was plotted for the case where the feature space consisted of normalised c^{pd1} and normalised s. The results obtained in both folds of the validation data were used to plot the ROC curve shown in Fig. 12. The AUC value for both first fold and second fold is 0.946, which indicates that GMM is capable of accurately classifying parts according to their densities for a range of threshold probabilities.

6.2 Supervised learning

Given the clustering results reported in the previous section, which encourage the use of extracted features from photodi- ode sensor data in assessing the quality of the L-PBF builds, it was investigated whether build density can be directly pre- dicted from the aforementioned features using a regression technique. Consequently, Gaussian Process (GP) regression with Automatic Relevance Determination (ARD) was used to predict the density of the specimens. To train the algo- rithm, 50% of the data (the first 31 datasets) was used and, once the optimum GP hyper-parameters were estimated, the remaining 50% of the data was used to validate the algo- rithm. 2-fold cross validation was again performed to meas- ure the predictive power of the algorithm outside the training set. The average RMS error was calculated using
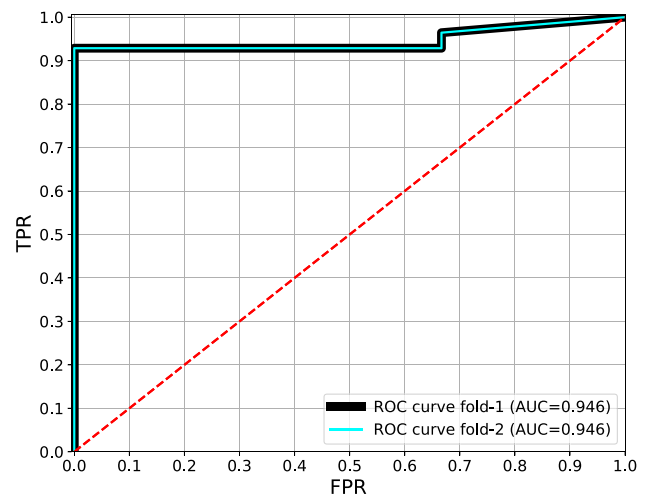

Fig. 12 The ROC curve plotted when GMM was used for the case where the feature space is consisted of normalised c^{pd1} and s. ROC curve was plotted for the results obtained in both folds of the valida- tion data

$$\text{Average RMS error} = \frac{\sum_{j=1}^{2} \left(\frac{1}{31} \sum_{i=1}^{31} (d_{*(i,j)} - d_{(i,j)})^2 \right)^{1/2}}{2} \tag{24}$$

where $d_{*(i,j)}$ represents the predicted density of the ith part in the jth fold while $d_{(i,j)}$ represents the measured density of the ith part in the jth fold.

The predicted density is plotted against the measured density in Fig. 13a. The black error bars represent a sin- gle standard deviation from the mean of each predicted
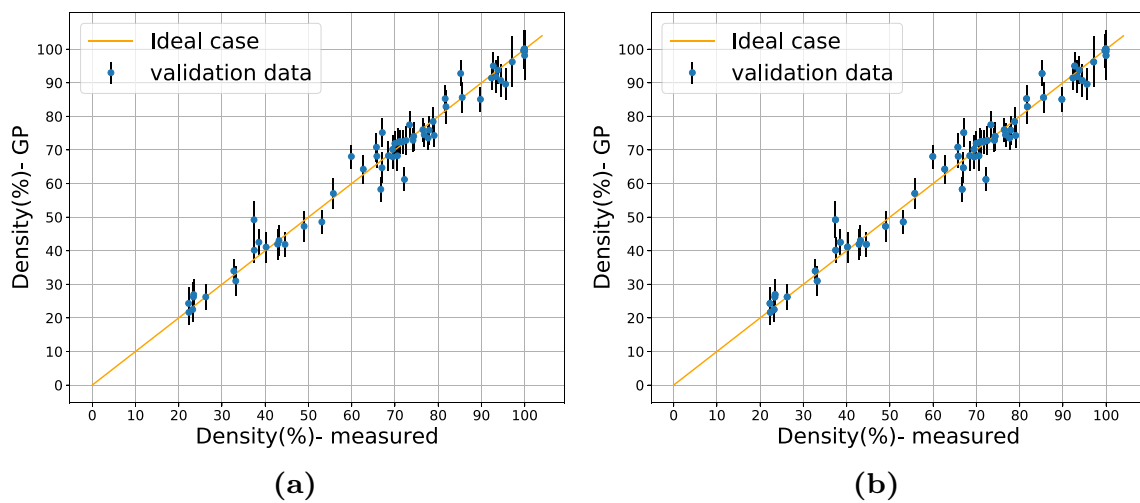
(a)

(b)

Fig. 13 Measured density compared against the predicted density obtained via GP regression. The blue points represent the validation data points. The black bars on each data point represent the standard deviation of each predicted density and the orange diagonal line represent the predictions of an ideal model. **a** The results obtained with all the extracted features, **b** results obtained without c^{pd1} and c^{pd2}

Table 4 Results of GP regression with ARD

	Density range	
	20–100%	90–100%
Length scale for **s**	1.08	2.60×10^{-1}
Length scale for c^{pd1}	3.21×10^{4}	6.77×10^{-1}
Length scale for c^{pd2}	2.04×10^{4}	3.61×10^{3}
Length scale for c^{pd3}	5.13	1.25×10^{3}

density. An ideal predictive model that can predict the exact measured density would place all predictions over the orange diagonal line. The results follow the line closely with sensible confidence bounds, indicating that the algorithm is capable of accurately predicting build density. The calculated average RMS error was 3.65%.

As explained in "Supervised learning", if the *j*th input has little predictive relevance, then the corresponding estimated lengthscale hyper-parameter, L_j, will be large (effectively removing the *j*th feature). Thus, referring to the lengthscale parameters reported in Table 4, it can be observed that the highest relevance was exhibited by **s** and c^{pd3} when the entire specimen set (density range from 20 to 100% - 62 specimens) was used to train the predictive model. According to the ARD results obtained for the entire specimen set, c^{pd1} and c^{pd2} can be discarded as model inputs as their associated lengthscales are relatively high compared to the other two lengthscales. As a result, the GP regression model was retrained without using c^{pd1} and c^{pd2} (results plotted in Fig. 13b). The average RMS error in this case was 3.6522%, which is relatively close to the previous average RMS error 3.6521%. It can therefore be concluded that c^{pd1} and c^{pd2}

have very little predictive relevance (which confirms the results obtained via ARD). GP regression is a non-parametric model hence part density cannot be explicitly expressed as a function of the input vectors. The surface plot shown in Fig. 14, however, shows the inferred relationship between part density and the 2 input vectors that were identified as having the highest importance (**s** and c^{pd3}).

To investigate the relevance of the input features in a relatively high density region, the specimens within the density range of 90–100% (13 specimens in total) were used to train the predictive model. As shown in Table 4 according to obtained ARD results the highest relevance was exhibited by **s** and c^{pd1}. Therefore, these results indicate that even if **s** and c^{pd3} help predict density for the density range of 20–100% as we go into a relatively high density range of 90–100%, **s** and c^{pd1} become more prominent in predicting build quality.

The work in the current section shows the potential for predicting the quality of L-PBF builds from sensor measurements, collected from different photodiodes. We note that build density is often required to be between 99–100%, but the average RMS error recorded for the overall specimen set was above 1%. However, the predicted density for the six specimens with density above 99%, (plotted in Fig. 15) had an RMS error of only 0.78%. As the experiment was originally designed to create specimens with a wide range of densities, the number of specimens found to have density above 99% was a very small portion of the entire specimen set. Thus, we believe that by expanding the sample size and focusing the experiment only on specimens with higher density the overall accuracy of the model in this region can be improved.

Since the most relevant features (**s** and c^{pd3}) for the overall specimen set directly correlates to the scan speed and the
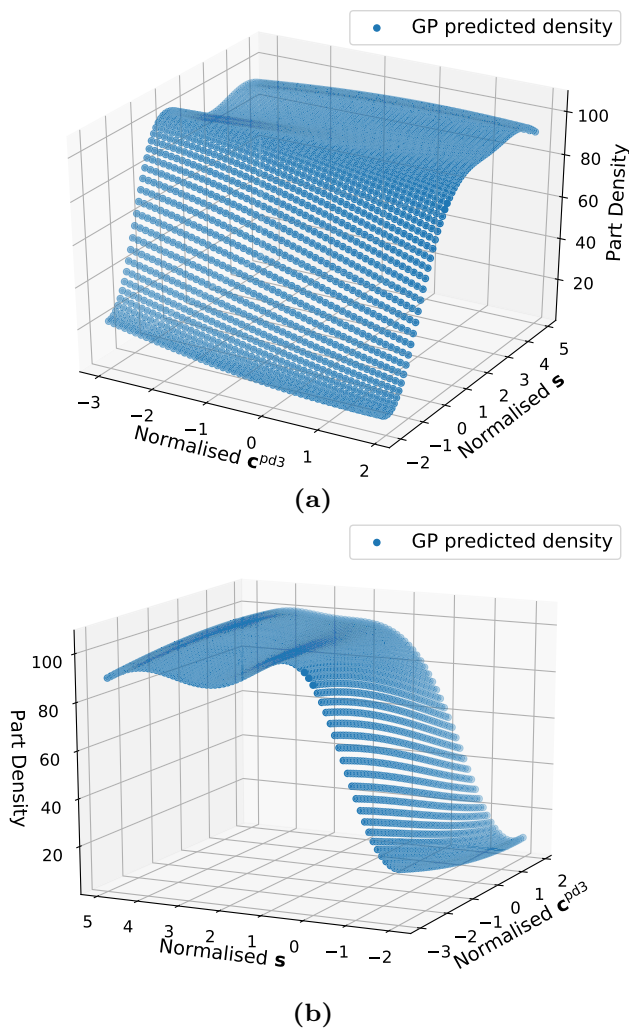
Fig. 14 Surface plot drawn for the part density predicted using input vectors s and c^{pd3}
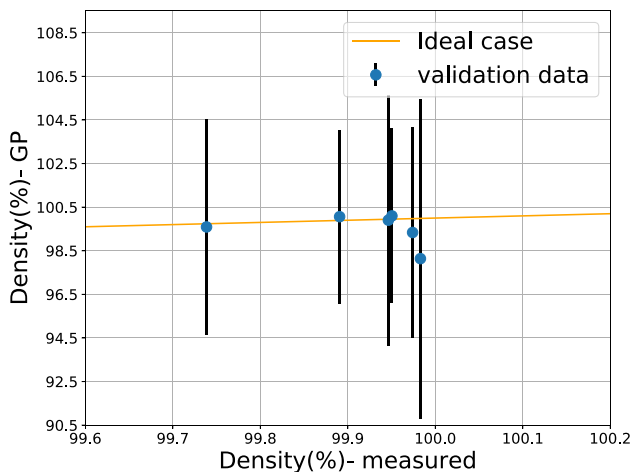

Fig. 15 Measured density compared against the predicted density zoomed in for the density range 99–100% with reference to Fig. 13a

laser power, it was investigated whether build density could be predicted with those process parameters. When only scan speed and laser power were used to predict build density, the average RMS error was 8.04%, as shown in Fig. 16a. With all the process parameters in use, density was predicted with an average RMS error of 5.06%, as shown in Fig. 16b. These results indicate that using the complete set of process parameters helps to increase prediction accuracy, but that features extracted from both photodiode sensor data and related to the energy transferred to the material provide additional useful information for predicting build density. A similar phenomenon was observed for the classification results, corroborating the hypothesis that, relative to the L-PBF process parameters, sensor data contributes additional information regarding part quality. As shown in Table 5, 48 out of 62 specimens were within one standard deviation when extracted features were used to predict build density but only 42 out of 62 specimens were within one standard deviation when process parameters were used to predict build density.

7 Conclusions

The absence of a robust quality control system in Additive Manufacturing introduces uncertainties regarding end-products' quality, hindering the adoption of AM technology in safety critical applications. Regarding Laser Powder Bed Fusion (L-PBF), most of the developmental monitoring systems in the literature have focused on image-based approaches. Using only photodiode sensor measurements for online process monitoring is an open challenge that is yet to be explored in depth. Advancing this approach would be beneficial as photodiodes are cost efficient, robust, and have a relatively high sample rate.

The aim of the present research, therefore, was to investigate the feasibility of predicting the density of L-PBF builds via photodiode sensor data collected during the build process. Firstly, an unsupervised clustering approach was investigated, whereby L-PBF parts were separated into 2 classes depending on their density. Features, extracted from photodiode measurements via Singular Value Decomposition, were used as inputs to two different unsupervised learning algorithms (K-means and a Gaussian Mixture Model). It was shown that the L-PBF builds could be clustered, depending on density, with accuracy levels of up to 93.54%.

Given the promising nature of the results realised using unsupervised approaches, Gaussian Process regression, a supervised approach, was then used to directly predict L-PBF build density. Again, these predictions were made using features extracted from photodiode measurements. The Gaussian Process was able to predict build density with an average RMS error of 3.65%. While the sample size of the training data has limited the accuracy of our approach,
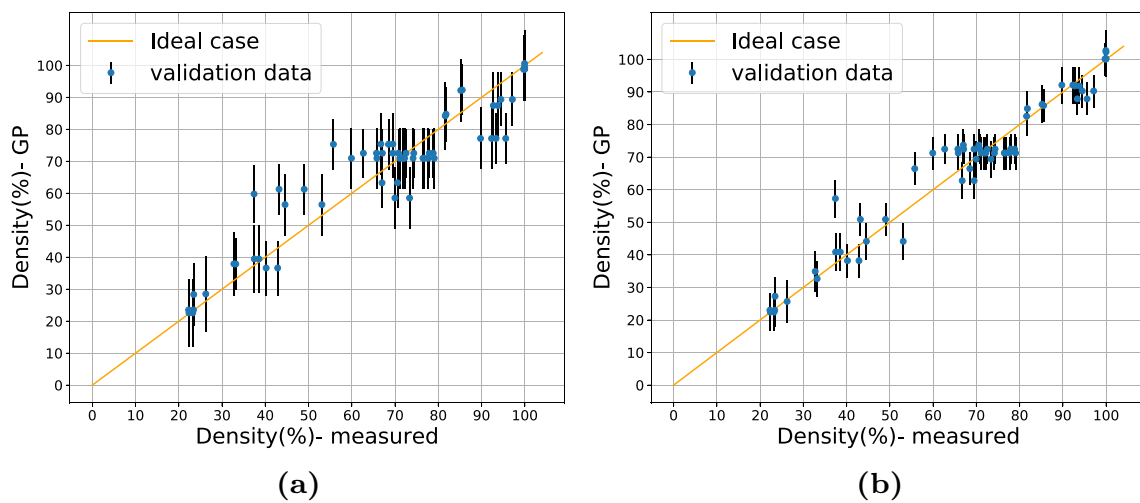
Fig. 16 Measured density compared against the predicted density. **a** The results obtained with normalised laser power and normalised laser speed, **b** results obtained with all the varying process parameters

Table 5 Results comparison

	Process parameters	Extracted features
Clustering		
GMM with EM accuracy	61.29%	93.54%
Regression		
Average RMS error	5.04%	3.65%
Number of points within σ	42 (out of 62)	48 (out of 62)
Number of points within 3σ	61 (out of 62)	61 (out of 62)

preliminary results show that, on high density samples (99–100%), the RMS error was only 0.78%. We believe that, by increasing the sample size in the high density region, the accuracy level in that region could be further improved.

The results presented in this paper illustrate that sensor data collected during the build process contain more information compared to process parameters set that are known *a priori*. The classification accuracy obtained using process parameters was only 61.29% and the average RMS error in GP regression was 5.04% while, when features extracted from the photodiode data were used, a classification accuracy of 93.54% and regression average RMS error of 3.65% were realised. For new materials, a considerable amount of the literature focuses only on the selection of process parameters in the parameter development stage. We propose focusing on using photodiode sensor data to assess and optimise part quality. We note that, for new materials, new specimen sets will be fabricated in the parameter optimisation stage, following standard practice. Hence, we believe that we may be able to use data generated during this initial phase to train the model when build parameters for new materials are being optimised. However, the applicability of

the approach across different materials is currently unknown and this will need to be investigated further in the future.

Although there may be some variability in the parts produced across different machines of the same model type, such variability is expected to be minimal for well-tuned machines and therefore we believe our model is robust enough to successfully predict part density fabricated across different machines of the same model type. Moreover, nowadays many AM machines provide co-axial or off-axial photodiodes that capture light emitted by the melt-pool (similar to the setup used for our study), therefore the proposed method can be easily adopted across different machines by re-training the model with suitable datasets. A rigorous study of robustness of the proposed methodology with respect to different machine models and sensor setups is beyond the scope of this paper, but it is an interesting avenue for future investigations.

A natural progression of this work is to increase the sample size and focus the experiment on specimens with higher density, so that accuracy can be further improved in the region of the part density that is the most relevant for end-users. The model can also be tested on new materials and machines to establish the model's generalisation capability.

Declarations

References

1. Industrial strategy: building a britain fit for the future. https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future. Accessed 18 Feb 2020

2. Alberts D, Schwarze D, Witt G (2017) In situ melt pool monitoring and the correlation to part density of Inconel® 718 for quality assurance in selective laser melting. In: International Solid Freeform Fabrication Symposium, Austin, TX, USA, pp 1481–1494

3. Aminzadeh M, Kurfess TR (2019) Online quality inspection using bayesian classification in powder-bed additive manufacturing from high-resolution visual camera images. J Intell Manuf 30(6):2505–2523

4. Arısoy YM, Criales LE, Özel T, Lane B, Moylan S, Donmez A (2017) Influence of scan strategy and process parameters on microstructure and its optimization in additively manufactured Nickel alloy 625 via laser powder bed fusion. Int J Adv Manuf Technol 90(5–8):1393–1417

5. Baumgartl H, Tomas J, Buettner R, Merkel M (2020) A deep learning-based model for defect detection in laser-powder bed fusion using in-situ thermographic monitoring. Progress in Additive Manufacturing pp 1–9

6. Berumen S, Bechmann F, Lindner S, Kruth JP, Craeghs T (2010) Quality control of laser-and powder bed-based Additive Manufacturing (AM) technologies. Physics Procedia 5:617–622

7. Bisht M, Ray N, Verbist F, Coeck S (2018) Correlation of selective laser melting-melt pool events with the tensile properties of Ti-6al-4v eli processed by laser powder bed fusion. Addit Manuf 22:302–306

8. Clijsters S, Craeghs T, Buls S, Kempen K, Kruth JP (2014) In situ quality control of the selective laser melting process using a high-speed, real-time melt pool monitoring system. Int J Adv Manuf Technol 75(5–8):1089–1101

9. Coeck S, Bisht M, Plas J, Verbist F (2019) Prediction of lack of fusion porosity in selective laser melting based on melt pool monitoring data. Addit Manuf 25:347–356

10. Craeghs T, Bechmann F, Berumen S, Kruth JP (2010) Feedback control of layerwise laser melting using optical sensors. Physics Procedia 5:505–514

11. Craeghs T, Clijsters S, Yasa E, Kruth JP (2011) Online quality control of selective laser melting. In: Proceedings of the Solid Freeform Fabrication Symposium, Austin, TX, pp 212–226

12. Elkan C (2003) Using the triangle inequality to accelerate k-means. In: Proceedings of the 20th international conference on Machine Learning (ICML-03), pp 147–153

13. Fletcher R (2013) Practical methods of optimization. John Wiley & Sons

14. Foster B, Reutzel E, Nassar A, Hall B, Brown S, Dickman C (2015) Optical, layerwise monitoring of powder bed fusion. In: Solid Freeform Fabrication Symposium, Austin, TX, Aug, pp 10–12

15. Fu C, Guo Y (2014) 3-dimensional finite element modeling of selective laser melting Ti-6Al-4V alloy. In: International

16. Gobert C, Reutzel EW, Petrich J, Nassar AR, Phoha S (2018) Application of supervised machine learning for defect detection during metallic powder bed fusion additive manufacturing using high resolution imaging. Addit Manuf 21:517–528

17. Grasso M, Laguzza V, Semeraro Q, Colosimo BM (2017) In-process monitoring of selective laser melting: spatial detection of defects via image data analysis. J Manuf Sci Eng 139(5):051001

18. Gu H, Gong H, Pal D, Rafi K, Starr T, Stucker B (2013) Influences of energy density on porosity and microstructure of selective laser melted 17-4ph stainless steel. In: International Solid Freeform Fabrication Symposium, Austin, TX, USA, p 474

19. Gupta N, Weber C, Newsome S (2012) Additive manufacturing: status and opportunities. Science and Technology Policy Institute, Washington

20. Horn RA, Mathias R (1990) An analog of the Cauchy–Schwarz inequality for Hadamard products and unitarily invariant norms SIAM. J Matrix Anal Appl 11(4):481–498

21. Khanzadeh M, Chowdhury S, Marufuzzaman M, Tschopp MA, Bian L (2018) Porosity prediction: supervised-learning of thermal history for direct laser deposition. J Manuf Syst 47:69–82

22. Kruth JP, Mercelis P, Van Vaerenbergh J, Froyen L, Rombouts M (2005) Binding mechanisms in selective laser sintering and selective laser melting. Rapid Prototyping J 11(1):26–36

23. Kurzynowski T, Chlebus E, Kuźnicka B, Reiner J (2012) Parameters in selective laser melting for processing metallic powders. In: High power laser materials processing: lasers, beam delivery, diagnostics, and applications, vol. 8239, p 823914. International Society for Optics and Photonics

24. Lott P, Schleifenbaum H, Meiners W, Wissenbach K, Hinke C, Bültmann J (2011) Design of an optical system for the in situ process monitoring of selective laser melting (SLM). Physics Procedia 12:683–690

25. Mani M, Feng S, Lane B, Donmez A, Moylan S, Fesperman R (2015) Measurement science needs for real-time control of additive manufacturing powder bed fusion processes. US Department of Commerce, National Institute of Standards and Technology

26. Okaro IA, Jayasinghe S, Sutcliffe C, Black K, Paoletti P, Green PL (2019) Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. Addit Manuf 27:42–53

27. Sames WJ, List F, Pannala S, Dehoff RR, Babu SS (2016) The metallurgy and processing science of metal additive manufacturing. Int Mater Rev 61(5):315–360

28. Spears TG, Gold SA (2016) In-process sensing in selective laser melting (SLM) additive manufacturing. Integrating Mater Manuf Innovation 5(1):2

29. Tapia G, Elwany A (2014) A review on process monitoring and control in metal-based additive manufacturing. J Manuf Sci Eng 136(6):060801

30. Tapia G, Elwany A, Sang H (2016) Prediction of porosity in metal-based additive manufacturing using spatial gaussian process models. Addit Manuf 12:282–290

31. Yadroitsev I, Bertrand P, Smurov I (2007) Parametric analysis of the selective laser melting process. Appl Surface Sci 253(19):8064–8069