A method for objectively evaluating the defect detection performance of in-situ monitoring systems

Henry C de Winton, Frederic Cegla, Paul A Hooper

Department of Mechanical Engineering, Imperial College London

Abstract

In-situ monitoring systems have the potential to assess material quality in additive manufacturing processes on-the-fly, paving the way for accelerated component qualification using defect digital twins. However, current systems vary widely in sensor technology and data analysis methods, leading to a lack of consensus in how the performance of these systems should be measured and compared. This work proposes a methodology and set of metrics, specifically Receiver Operating Characteristic (ROC) and Probability of Detection (POD) curves, to allow objective comparison of performance between any system, regardless of its underlying technology. We demonstrate this approach by comparing the ability to detect increases in part-wide porosity using two of the most common co-axial monitoring techniques in laser powder bed fusion; photodiodes and high-speed cameras. Using ROC curves, we show that melt pool metrics extracted from the camera offer a better trade off between detection and false alarms compared to the photodiode-based system when discriminating between samples at a 0.5% porosity threshold. POD curves were used to characterise detection capability across all porosity levels. It was found that the camera-based system can detect 43% of compromised parts (0.5% porosity), while the photodiode system detects 20%. However, for significantly compromised parts (5% porosity), the camera based method reaches 100%, while the photodiode only achieves 85%. The developed methodology shows that while the camera-based system is measurably superior, further improvement is needed before commercial implementation can be realised. Ultimately, the ROC-POD methodology allows objective assessments of detection performance, enabling quantifiable progress in the development of defect detection systems based on in-situ monitoring.

Keywords: Laser powder bed fusion, defect detection, porosity, in-situ monitoring.

1. Introduction

Any component manufactured for use in critical applications must be evaluated to ensure it can meet the demands placed upon it. This process of component qualification

*Corresponding author

Email address: paul.hooper@imperial.ac.uk (Henry C de Winton, Frederic Cegla, Paul A Hooper)

must be done with a high degree of certainty and often makes up a significant proportion of the total part cost and lead-time [1, 2]. In additive manufacturing, this is especially difficult due to the uncontrolled variation that is often inherent to the manufacturing process, leading to a lack of consistency in part properties and eroded confidence in use for critical applications.

In-situ monitoring offers a solution with unique benefits to additive manufacturing. Mapping defects during their creation would save significant time and cost in the qualification process . Furthermore, it offers the potential to reject or fix defective parts during the build process [3]. Ultimately, in-situ monitoring enabled additive manufacturing will lead to higher quality parts with a higher certainty in properties. However, the performance of current in-situ monitoring systems is not well understood, nor is there a good understanding of the required performance level to achieve the aim of accelerated qualification. Systems vary widely in sensor technology and data analysis methods, making direct comparisons of performance difficult. This paper aims to address these gaps and bring about a consensus in how the defect detection performance of in-situ monitoring systems should be measured.

AM components for the most structurally demanding applications tend to be metallic and manufactured using Laser Power Bed Fusion (LPBF) or Direct Energy Deposition (DED) [4]. Low porosity material is a fundamental requirement and porosity is a valuable indicator of part quality, often being the dominant factor behind material property variation including toughness and fatigue life [5, 6, 7]. Therefore, one basic requirement of the qualification process is to measure porosity and ensure it is within acceptable limits, either by in-situ methods or more traditonal ex-situ methods. While ex-situ methods work, they have several limitations. Archimedes' principle can only estimate a bulk porosity for a component and does not detect local regions of high porosity. Mechanical sectioning and imaging gives detailed information about porosity, but is destructive and labour intensive. Micro-CT gives a full 3D non-destructive measure of porosity throughout the part, however, resolution is highly dependent on part size, geometry and material x-ray attenuation [8]. Speed and cost of equipment are also concerns. In-situ monitoring used to estimate porosity has the potential to overcome many of these problems, as well as offering additional benefits of being able to take corrective action during a build, reducing wasted production time and raw materials.

Early research in the application of in-situ monitoring to LPBF consisted of co-axial camera and photodiode setups to monitor melt pool emissions [9, 10, 11, 12]. These works showed qualitatively that some level of process failure was detectable using in-situ process-monitoring and that process control was possible on that basis. However, the effectiveness was never quantified and the works discriminated between samples which would be unacceptable by current LPBF standards [13, 7]. Subsequent reviews of the field carried out by Sharratt in 2015 [14], by Everton et al in 2016[15] and by Grasso and Colosimo in 2017 [16] show that a wide variety of systems have been developed. In academia, designs have included those using optical, acoustic, depth and thermal measurements [14, 15, 16]. In industrial machines the focus has been on camera and photodiode based systems and every major manufacturer has now developed their own in-situ monitoring system [11, 12, 16]. These reviews also made the link between in-situ monitoring and the field of Non-Destructive Evaluation (NDE) [14]. However, the effectiveness of these systems to detect defective material or parts has rarely been quantitatively considered and no comparison between systems has been made on this basis.

While there is much work assessing process signatures from these systems [17, 18, 19, 20, 21, 22, 23], no single metric appears capable of predicting defective material or process failure across all modes [24, 25]. As a result, recent innovation in in-situ monitoring has focused on the application of machine learning to analyse in-situ monitoring data and combine metrics to improve detection capability [26, 25, 27, 28, 29, 30]. The methods applied vary widely in approach and some are inextricably linked to the sensors used to generate the raw data [31, 32]. In summary, almost every monitoring system is unique in both its hardware and software [25].

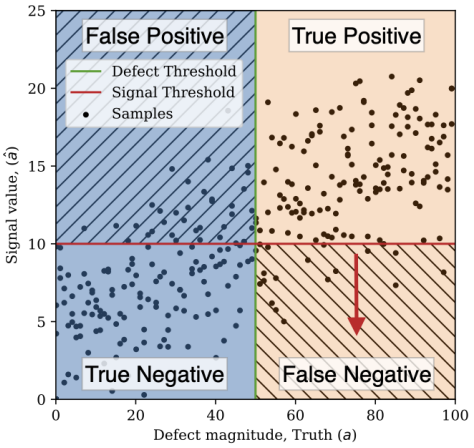
Research detecting process failures at a part, layer or melt pool level show that it is possible to correlate in-situ monitoring data with resulting material porosity [33, 34, 35, 36] and even identify individual pores [37, 38]. However, while these studies are notably more quantitative than the original in-situ monitoring work, detection effectiveness is still not thoroughly examined. Measures used to quantify detection effectiveness have included: accuracy and RMS error [36], F-score and Normalised Root Mean Squared Deviation [34] and Pearson's correlation coefficient [35]. While these measures are well-suited to fitting models, they are affected by the class skew of the data, making them poor absolute measures of detection capability. More descriptive measures used in the literature include Type I errors (false positives) and Type II errors (false negatives) [33] and True Positive Rate and False Positive Rate [37, 36]. However, all of these measures reduce the problem of matching signal strengths and defect sizes to two or three category classification. This greatly simplifies the problem and eliminates much understanding about the effect of defect size. Ultimately, while these studies significantly improve upon previous works, there is still a failing to characterise detection effectiveness consistently.

Without consistent quantification of detection capabilities, the effectiveness of different systems cannot be compared. What is needed is an objective method of evaluating defect detection systems, irrelevant of their specific approaches, sensors or algorithms. The field of Non-Destructive Evaluation (NDE) uses a variety of methods including Probability of Detection (POD) graphs and Receiver Operating Characteristic (ROC) curves [39]. These two curves are the industry standard when evaluating defect detection techniques including Ultrasonic C-scanning, Radiography, Magnetic Flux Leakage and Eddy Currents [40, 41, 42]. Furthermore these curves, in particular ROC curves, are seen widely in evaluating medical diagnostic tests [43, 44]. In these fields human operators are often part of the detection system and non-deterministic correlations between measurement and diagnostic output exist without issue. When combined with further information on the system resolution and types of defects detected these methods serve to fully characterise any system and allow quantitative comparison between them.

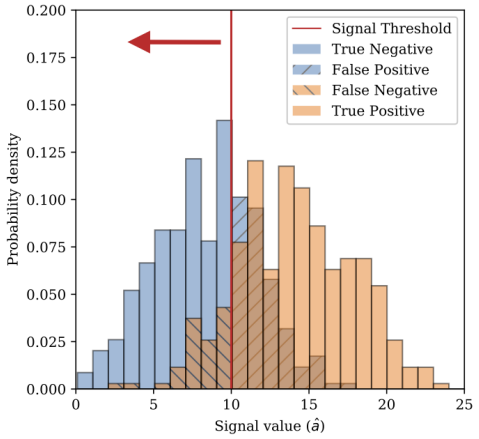
This article first describes the measures of detection effectiveness relevant to in-situ monitoring in section 2. This methodology is then applied in a case study in section 3 which compares four systems; two based on high-speed cameras and two based on photodiodes. The results of the comparison are presented and described in section 4. A discussion of the results of the case study and the evaluation methodology are contained in section 5.

2. Detection theory and its application to laser powder bed fusion

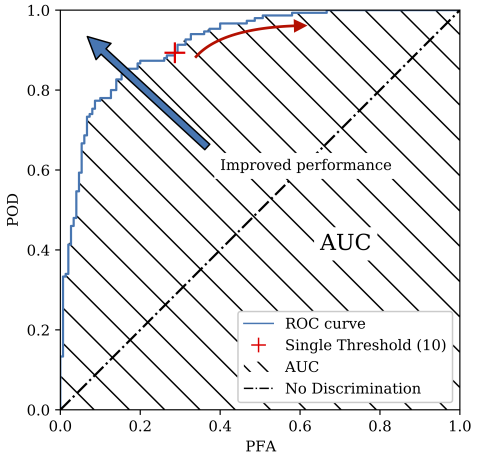
Detection performance comprises detection, characterisation and location. In this context detection purely refers to identifying a defect's existence. Finding its size or



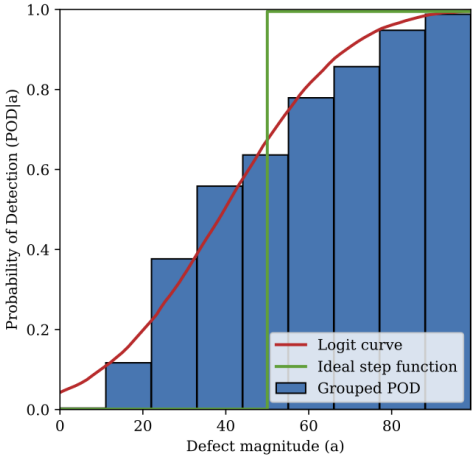
(a) Signal-Truth data with desired defect threshold, chosen signal threshold and four potential outcomes.



(b) Histogram of signals split by the defect threshold in fig. 1a and showing the signal threshold in red.



(c) The ROC curve plotted for the data in fig. 1a and fig. 1b. Marked with a red cross is the response for the red threshold. The dotted-dashed black line shows the expected performance of random guessing. The Area Under the Curve (AUC) is shown shaded.



(d) The POD curve for the signal threshold in fig. 1c is shown in red, the ideal response based on the defect threshold in fig. 1b is shown in green.

Figure 1: Detection theory in practice. fig. 1a and fig. 1b show the data and thresholds along with the four potential outcomes for any data point. For the clear samples the outcome is either True Negative or False Positive and for the defective samples the outcome is either True Positive or False Negative. fig. 1c plots the ROC, the red arrow showing the effect of moving the threshold. fig. 1d shows the POD curve for the red threshold. The effect of moving the signal threshold in fig. 1a will change the shape and location of the curve.

severity is the domain of characterisation. Finally, the determination of a defect's location is a function of a systems' resolution, meaning the area analysed per reading. For each area analysed a system returns a continuous signal. The strength of the returned signal

depends on the defect (size, detectability) and the system (noise level, sensitivity). An example data set of defects (a) and signals ($\hat{a}$)is plotted in fig. 1a, being an example there are no units specified. The defect type will determine the units which could be porosity (%) or individual pore length (µm). The units of the signal are generally not important as long as they are consistent, since the result could be unit-less values generated by a regression method. Due to sensor noise a signal is still returned even when no defect is present. In the context of this article a detection system is a system which generates estimates of defectiveness (i.e. porosity, pore diameter in micrometers, etc) for a given region, these can be binary, discrete or continuous. Specifically the term detection system covers any and all stages of sensing, processing and analysis necessary to generate an estimate. Any estimate can be compared to truth data and evaluated, whether this estimate is generated with a sophisticated multi-sensory approach or a simple coin toss.

2.1. Probability of detection and false alarm

The detection of defects can be simplified into a binary classification problem. Assuming we have some results for which we have the signal value and the true defect magnitude, a truth/defect threshold can be decided. All results above the threshold can be declared 'defective' and all those below the threshold can be declared 'clear'. In fig. 1a this threshold is shown as a vertical green line. Splitting the data into two sets eliminates all information of the defect's size beyond its relation to the threshold, this severely limits any possible characterisation. Based on this threshold there are two types of signal based on truth, $signal|clear$ and $signal|defect$. Histograms of these two types of signal are shown in fig. 1b. These two ranges may overlap, making it difficult to ascertain the true state of the sample based on the signal.

The signal is then referenced to a signal threshold yielding four possible outcomes as shown in fig. 1a and fig. 1b. Two for 'clear' samples: true negative (TN) and false positive (FP), and two for the 'defective' samples: true positive (TP) and false negative (FN). Of these outcomes a false positive and false negative are also known as type 1 and type 2 errors respectively. In fig. 1a and fig. 1b the signal threshold is shown with a red line and the areas representing the outcomes are shaded, clear samples are blue and defective samples are orange with the errors hatched. A system returning a continuous signal when combined with a signal threshold creates a defect classifier. For the same system, different thresholds may be applied.

In the field of Non Destructive Evaluation, characterisation of sensor performance is defined by combing these four outcomes (true negative, false positive, true positive, false negative) into Probability of Detection (POD) and Probability of False Alarm (PFA). Together POD (eq. (1)) and PFA (eq. (2)) fully describe the detection outcomes; false negative is the complement of true positive and false positive is the complement of true negative. Crucially, POD and PFA are not dependent on each other and are determined solely by the threshold and the probability density distribution of either the $signal|defect$ or $signal|clear$ respectively. Included in the eq. (1) and eq. (2) are statistical definitions of POD and PFA, the only difference is whether the defect is $>$ or $<$ the chosen defect threshold.

$$POD = \frac{TP}{TP + FN} = P(signal > signal\ threshold \mid defect > defect\ threshold) \quad (1)$$

$$PFA = \frac{FP}{TN + FP} = P(signal > signal\ threshold \mid defect < defect\ threshold) \quad (2)$$

Other fields combine these four outcomes in different ways, for example; Recall and Precision in machine learning, Sensitivity and Specificity in medicine. It is important to consider why POD and PFA should be used in preference to these other metrics.

Class skew occurs when the number of 'clear' and 'defective' samples is different. POD and PFA each depend on a single distribution and so are insensitive to class skew, unlike measures such as Precision and by extension F-scores. This is important as defect frequency varies between data sets, even when the defects are equally detectable and the sensor equally effective.

Alternative measures such as Precision and F-score combine results from both 'clear' and 'defective' classes. This makes them useful in the context of optimising on a single data set [34, 45]. However, such measures should not be used when trying to generalise sensor performance [33, 44, 46], since the rate of defect occurrence can vary between data sets. For similar reasons, classification accuracy is an even poorer measure of detector performance. If the defect-to-clear ratio is 1:99 (and it may be much higher when evaluating high density material) then 99% accuracy may be achieved by simply classifying all instances as defect free.

POD and PFA are also descriptive statistics with regards to the resulting part or process. Increasing POD reduces the number of misses and leads to improved confidence in part life. Increasing PFA increases the false positives and by extension the manufacturing cost as more parts are discarded needlessly. This is useful as it constrains the problem of selecting the optimum threshold, rather than simply maximising POD alone which leads to the rejection of all parts to guarantee none are defective[47]. Ultimately, POD and PFA allow sensible comparison of detector performance [44, 46, 33].

However, a high POD or low PFA alone does not guarantee defect-free material or a more efficient manufacturing process. While POD and PFA are insensitive to class skew, it should still be taken into account. Assured high quality parts require a high POD sensor and a low probability of occurrence. That is to say, should defects be very common then there is a high chance a few will be missed, even by a highly sensitive detector. Likewise low PFA does not guarantee low waste. This will depend on the PFA and the frequency of the defect free instances. If most of the material is high quality then even if PFA is apparently low, many false alarms can still be raised. For example during 200 assessments searching for pores in 99.5% dense AM material one would only expect to find 1 true pore. However, taking an example POD = 0.9 and PFA = 0.1 would mean we are likely to detect the single pore, we would also expect to find 20 false alarms, rendering the system useless.

2.1.1. Receiver Operating Characteristic

Single POD and PFA values characterise a detector for an individual signal threshold and an individual defect level. However, the Receiver Operating Characteristic (ROC) curve (shown in fig. 1c) plots POD and PFA for the complete range of possible signal thresholds. Shown in fig. 1a is the raw signal-truth data where the defect level is green and signal threshold is red. The resulting ROC curve for the data in fig. 1a can be seen in fig. 1c and the outcome of the red threshold in fig. 1a and fig. 1b is marked by a

red cross in fig. 1c. The black dotted-dashed line shows no discrimination, the expected performance of random guessing. The ROC curve should not be below this line; should that occur then there is negative discrimination and the threshold may be reversed to invert the line. The red arrows in fig. 1a,fig. 1b and fig. 1c show the effect of changing threshold. Plotting for all threshold values allows a proper assessment of the range of potential trade offs between POD and PFA.

The Area Under the Curve (AUC), shown shaded in fig. 1c, can be used as a single metric to summarise detector effectiveness. Detector effectiveness can be understood as the range of POD and PFA across all potential thresholds [46]. This can be used as an alternative to F-score for optimisation purposes [48]. However, it is possible for higher AUC curves to perform worse than lower AUC curves in specific regions.

A simple method to create an ROC curve from pairs of continuous signal and truth data is to first binarise the truth data based on the desired defect threshold. The data points can then be ordered by ascending signal value before moving through the data points calculating the POD and PFA as if the signal value of the current data point was the signal threshold. The resulting POD/PFA pairs can then be plotted to form the ROC curve. For further information on creating and interpreting ROC curves the authors recommend Fawcett [46].

2.1.2. Probability of detection curves

POD and PFA metrics and ROC curves have so far been discussed for binary classification. This ignores the reality that there is a wide range of defect sizes. Some are worse than others. It is therefore necessary to understand how the classifier reacts to the whole spectrum of defect sizes. This allows us to answer important questions such as"what is the largest defect size which my system is likely to miss?". This is in contrast to the question "what is the smallest defect size my system can possibly detect". Reliably detecting the largest defects is of considerably more value than occasionally detecting the smallest defects.

Assessment of a classifier across the full range of defect sizes can be achieved using a POD curve as shown in fig. 1d. The specific classifier in this example is show in fig. 1c and is designated with a red cross. Figure 1d plots POD for a given defect size ($POD|a$) against defect size (a).$POD|a$ is $P(signal > signal\ threshold \mid defect = a)$, unlike POD on an ROC curve where the detection is for defects above a threshold, rather than a specific value. Ideally $POD|a$ is a step function with the step taking place at the chosen defect threshold. Detection of defects less than the desired defect threshold are considered false alarms, therefore the ideal classifier has no response to defects below the desired defect threshold and always responds to defects above it. A classifier with no discrimination would have a horizontal line.

ROC curves may be used for relative comparison and highlight the possible trade-offs of POD and PFA. However, they are sensitive to the distribution of defect sizes in a data set. If all the defects are tightly clustered around the defect threshold the detector may struggle to classify them well, whereas if the defects are grouped well above and well below the threshold, classification may be easier. POD curves capture a system's response to the full range of individual defect sizes and serve as an absolute assessment of a systems sensitivity. They directly link detection to defect size and are seen widely in NDE literature to summarise a detector's capability [40, 41, 42].

The simplest approach to generating a POD plot is to group defects into bins and calculate POD over smaller ranges, this is shown in the fig. 1d as a blue bar chart. The width of each bar shows the range of defects grouped to create the estimate. This technique has the benefit of making no assumptions about the underlying distribution. However, it has the drawback of creating a trade off between the accuracy of each estimate and the number of groups. Large volumes of data are therefore needed to assemble accurate POD curves using this method.

Two alternatives are commonly used to create a continuous line, logit models can be fitted to binarised signal data (this is shown as a red line in fig. 1d) or an a vs $\hat{a}$ model maybe fitted to continuous response data [39]. The fitting methods make assumptions about the underlying distribution, however, they generate very similar curves with the main benefit of the a vs $\hat{a}$ model being tighter confidence bounds due the increased volume of information in continuous rather that binary response data [49]. The logit model in fig. 1d is based on the red signal threshold in the previous three figures and the grouped and logit models of $POD|a$ are in good agreement.

The benefit of fitted models is that far fewer data points are required to fit the curve. In NDE this has previously been seen as an advantage as each data point may be very expensive to generate [40, 41, 42]. However, with additive manufacturing one might expect many thousands of data points from a single micro computed tomography scan, therefore applying the simple grouped method may be widely applicable. This would reduce the number of statistical assumptions which may not always hold for in-situ processing monitoring of laser powder bed fusion where melt pool mechanics can vary considerably. For complete instructions on generating POD curves the authors recommend appendix G of the Non-Destructive Evaluation System Reliability Assessment Handbook [40].

2.2. System Resolution

While ROC and POD curves serve to assess detection, knowing the location of a defect is also an important aspect of a system's performance. Knowing the porosity for each mm^3 of a part is much more valuable than having a single estimate for the entire part. However, system resolution is separate from the accuracy of each estimate.

Every system has a minimum resolution or area analysed, this dictates the accuracy and precision of estimates of defect location. For most off-axis systems this is determined solely by the system resolution. However, for co-axial systems the temporal resolution also plays a role. The speed of the scan head will determine the area of the sample covered by each reading. The resolution can also be affected by post-processing; if layer-wise images of a whole printer bed were used and pixel-wise segmentation was applied, the resolution would be the size of the pixels of the image. However, if one were to evaluate these images using simple classification then at best the resolution would be the whole layer visible in the image. Temporally, if several data points were to be averaged to represent a larger area, the resolution would also decrease to this larger area.

Resolution of the system is also constrained by the resolution of the reference technique used to evaluate the system. Even if the system collects data points in high resolution, if they are evaluated against estimates of part-wide porosity the system's measured performance is limited to this task. That is to say, each of the many readings are being assessed for their ability to predict part porosity, even if in reality they focus on a small region of the part.

2.3. Truth data

In order to evaluate the accuracy and precision of a system the true nature of the defects must be known. In NDE, either a second referee technique or some form of artificial defect may be used. In the case of additive manufacturing, porosity analysis is typically via micrographs or micro-CT may be used to map out the true defects [33, 34, 37]. However, the results of these techniques are still not the absolute truth and are subject to their own uncertainties and resolution limits. Consequently the system being assessed can never be shown to be more accurate than the truth data and the limitations of the techniques used to create it.

In order to ensure a wide range of defects to detect, samples may be further supplemented with some artificially generated defects. In the case of additive manufacturing these can be created by deliberately destabilising the laser powder bed fusion process such as by deliberately creating lack of fusion areas [37]. However, the creation of artificial defects will affect the applicability of the result. For instance, being able to detect defects generated by poor laser focus may not generalise to process failures generated by denudation.

3. Methods: Porosity detection using Photodiodes and Cameras

This case study applies the methodology in section 2 to assess the detection of sample porosity by four competing systems. Two systems are based on a synthetic photodiode (integrated image) and two systems are based on high-speed imaging and computer vision (returning melt pool metrics). Camera and photodiode based systems were used in early defect detection works [9, 12] and they have been offered on industrial machines for the last 5 years [15]. Photodiodes [33, 50, 34, 35] and melt pool metrics extracted from images [37, 38] continue to be seen in contemporary detection studies. However, no quantitative comparison of the two was found in the literature.

The aims of this case study is to compare the effectiveness of a photodiode based approach to that based on a camera and to determine at what porosity level either is effective. However, to illustrate the importance of post-processing and analysis in in-situ monitoring, systems based on raw signals are also included for comparison. Furthermore, with post-processing and analysis playing a large role in in-situ systems, our results should not be overly generalised to all photodiode or camera based systems. Improved post-processing and analysis methods should yield further increases in performance.

Consequently four sets of signal values are returned:

1. A raw photodiode signal (Integrated image).
2. A processed photodiode signal.
3. A single melt pool metric (Spatter number).
4. A machine learning model (K-Nearest Neighbours, KNN) combining several melt pool metrics.

The general stages used to evaluate system detection performance start with building samples and acquiring in-situ monitoring data, in this case with a co-axial high-speed camera. The in-situ data is processed to generate the predictive signal values for each system. The truth data must then be obtained, in this work the built samples are sectioned and imaged to obtain part porosity. Once the signal values and ground truth

data are collected the requisite ROC and POD curves can be generated for each system. These stages are discussed in more detail in the sections below.

3.1. Sensing hardware

A Renishaw AM250 machine was modified to include a co-axial, high-speed imaging capability as shown in fig. 2. The light returned by the melt pool was collected using a long-pass beam splitter at 1000nm. This was subsequently focused, filtered at 700nm and imaged using a Photron SA5 camera. The recordings were made at a resolution of 128×128 and a frame rate of 100kHz. 100kHz was chosen to capture most high-speed phenomena of the melt pool while ensuring reasonable exposure times since no illumination was used. Additionally, the laser X,Y positions and power were recorded. Further details of the monitoring system can be found in [51]. The AM250 machine uses a 200W, 1070nm laser with a 70 µm spot size. The feedstock was SS316L with a 10 µm to 50 µm particle distribution. The shielding gas was argon.

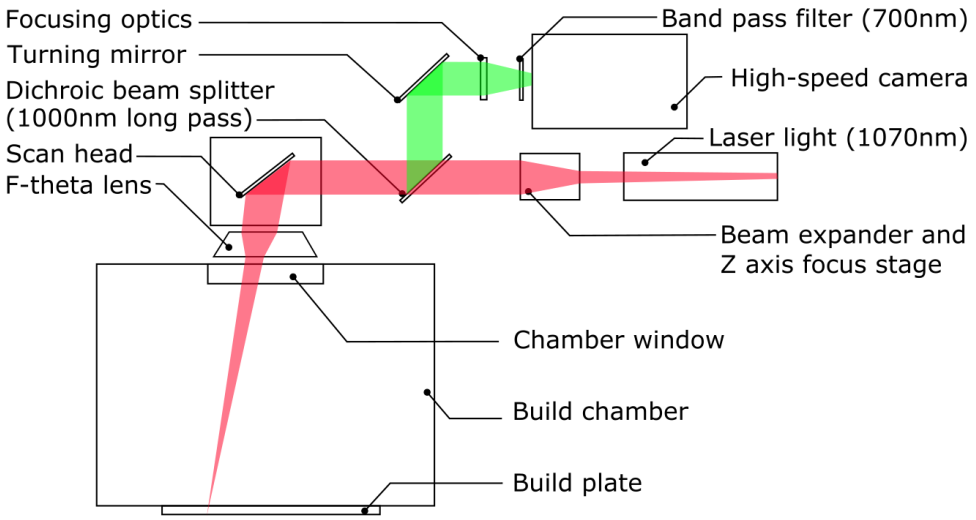


Figure 2: Diagram of the optical train for the co-axial imaging system. The laser-focusing stage is before the dichroic beam splitter and the image-focusing optics are directly before the camera allowing the laser to be focused independently of the camera. The imaging wavelength of the system is 700nm. More details of the system can be found in [51].

3.2. Porosity detection experiments

Samples with widely varying porosity were created by de-focusing the laser during the hatching section of the samples. Laser de-focusing happens naturally in the LPBF process due to thermal lensing where residue generated by melting is deposited on the laser window [52]. No other parameters were altered. This includes laser focus during the contour scan of the sample where the scan strategy is significantly different from hatching. It was intended to only compare the hatching areas since this represents the bulk of the sample.

Five builds of nine samples were performed in total, with the final build being a reference build with all samples having zero focus height. The remaining four builds each had samples with -16, -12, -8, -4, 0, 4, 8, 12, 16 mm of focus height for one of the nine samples. The Renishaw AM250 maintains laser spot size for at least ±1mm of focus [53], in order to achieve detectable laser de-focusing steps of ±4mm where chosen. These were randomly allocated to each sample in the build, for example sample one on the bed might have focus height −16 mm in the first build, 6 mm in the second and so on. This was to eliminate the effects of bed position such as those relating to gas flow and wiper direction.

The samples were cylinders 5 mm in diameter, 20 mm in length and stood vertically on the build plate with length in the build direction. Each build consisted of 9 cylinders in a 3×3 array with 50 mm spacing in the X and Y direction between samples to prevent any interference.

During the builds, the monitoring system was run twice yielding two sets of footage. This was done once early in the build on approximately the 50^{th} layer and later on approximately the 150^{th} layer. Each recording was 3.6 seconds long with the hatching section of each cylinder being 25,000 frames. The remaining 35% of the footage captured was contour scans and scan-head movement between cylinders. Post-build, the footage was split into blocks, one for each cylinder. Subsequently this was cut down to only footage of the hatching process, removing the contour melting and movement between each sample. The data processing took place after the experiments and was not completed in on-the-fly.

3.3. Detection systems and image processing methods

Four detection systems were developed for comparison, as shown in fig. 3, two using a photodiode based approach and two using an image based approach. The first system aimed to approximate a photodiode sensor. This was achieved by summing intensities over the central 50×50 pixels of the sensor to recreate a single 1 mm^2 sensor. This was done to recreate the higher-bit depth and large sensor size of photodiodes when compared to camera pixels. Furthermore, the 100 kHz sampling rate of the camera matched the highest rates seen in literature for photodiodes, such as those used in [34]. A high sampling rate compared to most cameras is traditionally one of the greatest benefits of using photodiodes [54, 9, 55]. The raw photodiode signal was used as the first detection system in its own right. However, performance was significantly improved using post-processing, creating a second system listed as the processed photodiode.

In the third and fourth systems camera images were analysed to identify the melt pool and spatter regions and record particular metrics such as the number of spatter particles and melt pool length. The third system used a single metric, spatter number, to perform detection. This was chosen since spatter is highly linked to porosity and defect formation in laser powder bed fusion [20, 22, 23]. The fourth system combined all image based metrics using k Nearest Neighbours regression.

3.3.1. Photodiode Post-Processing

For each sample there was a photodiode trace, with a single number per frame. However, when trying to differentiate between porous and non-porous samples based on individual frames, the raw photodiode performed poorly. This can be clearly seen in

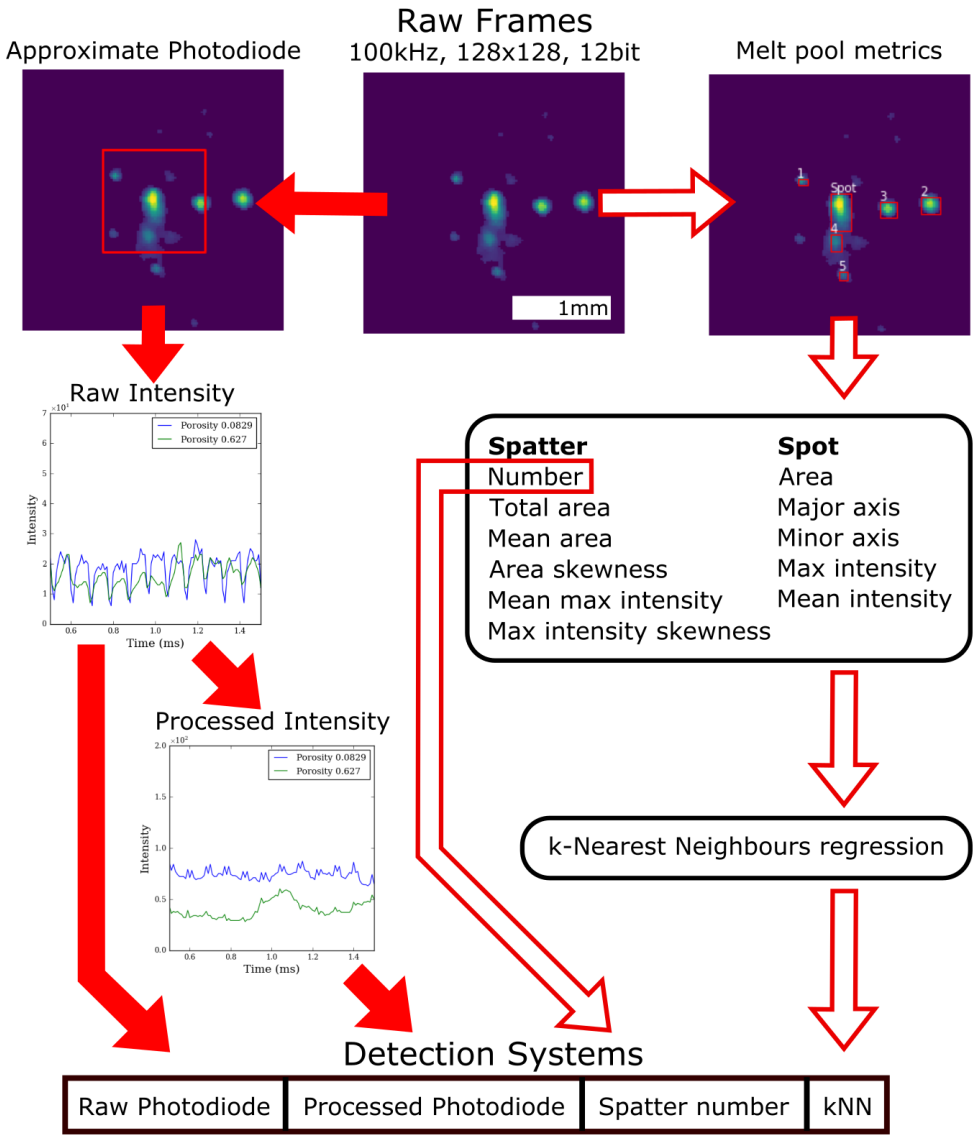
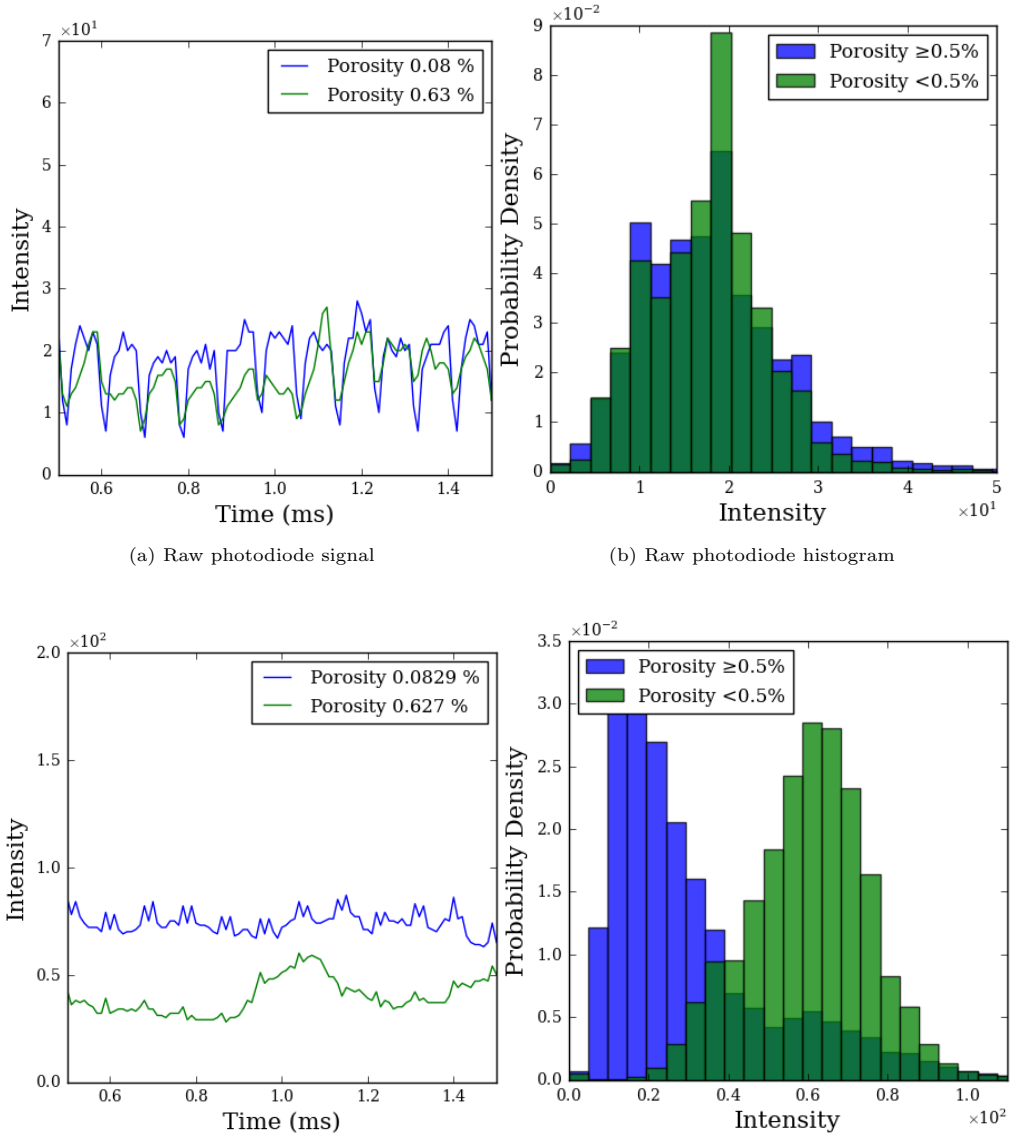


Figure 3: Detection system processing routes. Starting from a raw 12-bit monochrome frame there were two processing options. One recreated a photodiode; this summed the intensity values for pixels within a 50×50 pixel area yielding a single intensity value. The alternative option aimed to capture specific melt pool metrics based on an understanding of what was being imaged. To identify the relevant areas, the image was binarised and the spot and spatter identified. From this, metrics relating to the two regions of interest could be extracted. The metrics describe the melt pool in individual detail and the spatter particles as a group.

fig. 4a and fig. 4b, where a zoomed view shows that the signals clearly overlap and that the histograms for signals with more than 0.5% porosity are indistinguishable from those

with less than 0.5% porosity.

In order to make proper use of the photodiode values, they were post-processed to extract a more useful signal. The raw signal for a high density sample (shown in blue in fig. 4a) shows a repeating pattern. This is caused by the modulated laser power input (pulsed) on the AM250 machine, a phenomena which would likely not be present on a continuously firing laser system. A fully-focused laser creates a very bright central point in an otherwise small teardrop-shaped melt pool. When the laser pulse stops, the central spot rapidly decreases in brightness. When the laser is defocused the intensity of the central spot point is much reduced, creating less pronounced peaks and troughs in the signal. This can be seen in the signal for a lower density sample as shown in green in fig. 4b. To capture these pronounced dips in brightness the absolute difference between consecutive samples was summed over a small window of 20 samples. The resulting signal is referred to as intensity variability. The 20-sample-wide filter window was wide enough to always capture at least one laser pulse entirely. Pulses were on average 9 samples long. The results of this filtering process can be seen in fig. 4c where the denser sample clearly has a larger variability in intensity when compared to the less dense sample. This is reflected in the histogram in fig. 4d where the denser samples predominantly have a higher variability in intensity reflecting the stable cycling of laser pulses.



(a) Raw photodiode signal

(b) Raw photodiode histogram

(c) Photodiode variability signal

(d) Photodiode variability histogram

Figure 4: Raw and processed photodiode signals.

3.3.2. Melt pool metric extraction

The melt pool dynamics of laser powder bed fusion are complex and as yet not fully understood. However, several works have shown links between melt pool and spatter dimensions and melting regime, pore formation and build quality [56, 17, 20, 57, 19]. An automated process was created to extract some key descriptors of the melt pool situation. These descriptors included spot size (specifically the area and length of the major and minor axes) and the pool's maximum and mean intensity. With the spot identified, the

remaining visible particles were classified as spatter. Their number and size varied significantly, so descriptors were used to capture the nature of the spatter distribution as a whole, specifically their mean and median size and intensity. Unlike the raw photodiode results shown in fig. 4b, the melt pool metrics already showed reasonable levels of discrimination between porous and non-porous samples when assessing individual frames. Therefore, no further post-processing was deemed necessary.

3.3.3. Combining melt pool metrics using machine learning

The melt pool metrics were combined using an off-the-shelf machine learning algorithm to evaluate if this improved detection capability over a single metric in isolation. Several algorithms from the SKLearn python package were evaluated based on the maximum AUC score when discriminating between samples at the 0.5% porosity level. This was performed at the track length scale, 2,500 frames averaged to create a single value predicting density. Algorithms assessed included linear and logistic regression, decision trees, random forests, support vector machines and Gaussian naive Bayes. However, K-Nearest Neighbours (KNN) proved the most effective with limited tuning. All evaluations were performed using a 50-50 train-test split in the data and length scale averaging was performed prior to applying the model.

3.3.4. Length scale averaging

Discrimination between high and low density samples was initially performed using values from each frame, allowing many assessments of each sample. However, large variance among even small number of frames meant even the best metrics were only weakly predictive. Therefore, the image statistics and photodiode values were subsequently averaged over 3 different time scales to capture trends at 3 different length scales. Since the imaging was co-axial the number of frames correlates with area of the sample monitored. The whole layer average used all available footage of the hatching section of a single layer which was approximately 25,000 frames in total. The track level average split the data into 2,500 frame blocks, averaging each independently and dropping any remaining frames. 2,500 frames corresponds to an estimated laser travel distance during

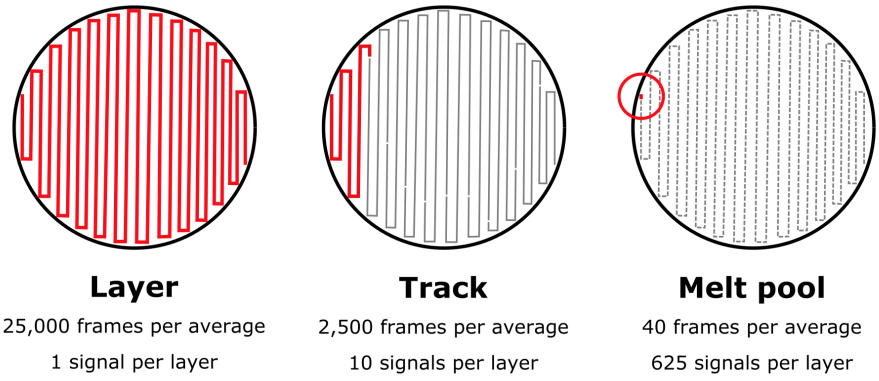


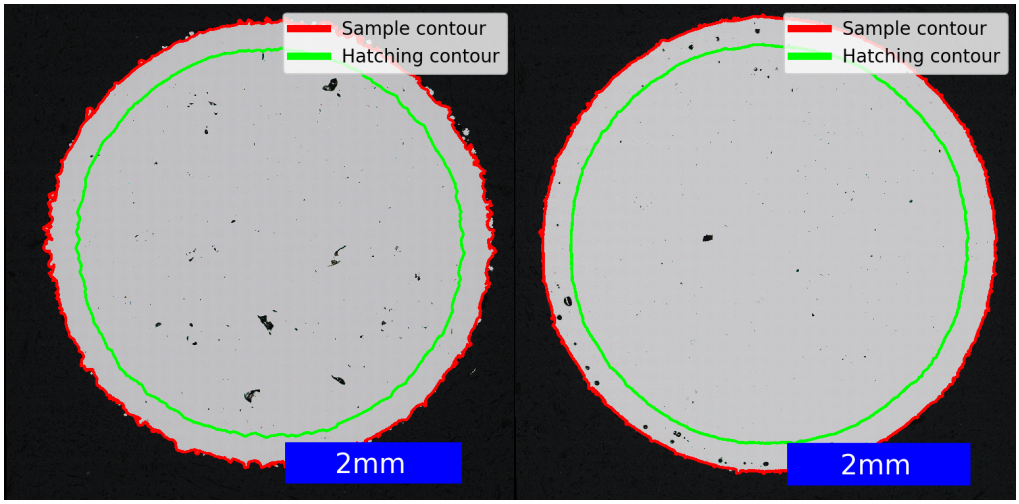
Figure 5: The three different length scales which the statistics where averaged over and how that relates scan path.

each recording of 18.75 mm or 278 laser exposures. The melt pool length scale was taken as 40 frames, with the laser averaging 750mm/s travel speed; this corresponds to approximately 300 µm in laser travel per recording, this coincides with the length of a typical melt pool 300-350 µm or approximately 4.5 laser exposures. Our estimate of melt pool length is based on the images acquired for this study and therefore may not be applicable to other machines, materials or scan strategies.

3.4. Porosity Measurement

The sample porosity truth data was generated using micrographs of each sample. The samples were cut perpendicular to the build direction approximately halfway up the sample, mounted and polished. The micrographs were acquired with a Hirox RH-2000 microscope at 600x optical zoom with a resolution of 0.64µm per pixel. Each micrograph was constructed from a stitched grid of approximately 200 images to create a high resolution micrograph of each entire sample as seen in fig. 6.

The micrographs were assumed to be representative of part-wide, global porosity. However, different scan strategies are employed for the hatched and contour sections of a part. When the parameters of the hatching section are varied, it has a large effect on porosity in both sections of the part. This can be seen in fig. 6a where the hatching section displays many large lack-of-fusion pores due to the defocused laser while the contour strategy is the default and is very solid. High density samples, seen in fig. 6b, use a well focused laser for the hatching but the same contour strategy. The hatching region is very dense while the contour region displays many large keyhole pores, likely due to a hotter sample temperatures reached due to the more effective energy input during the hatching section which used a more focused laser. Since only footage of hatching was being evaluated, an approach was taken to measure specifically the porosity in the hatched region of the sample. For each sample the outer contour was found and then reduced to allow porosity analysis of the hatching section alone. The parts were still assumed to be homogeneous in the build direction, a single porosity measurement was used as truth data for both sets of footage taken for each sample on separate layers.



(a) Sample porosity 0.54%, hatching porosity 0.67 %. A 24% increase in porosity.

(b) Sample porosity 0.37%, hatching porosity 0.29. %. A 21% decrease in porosity.

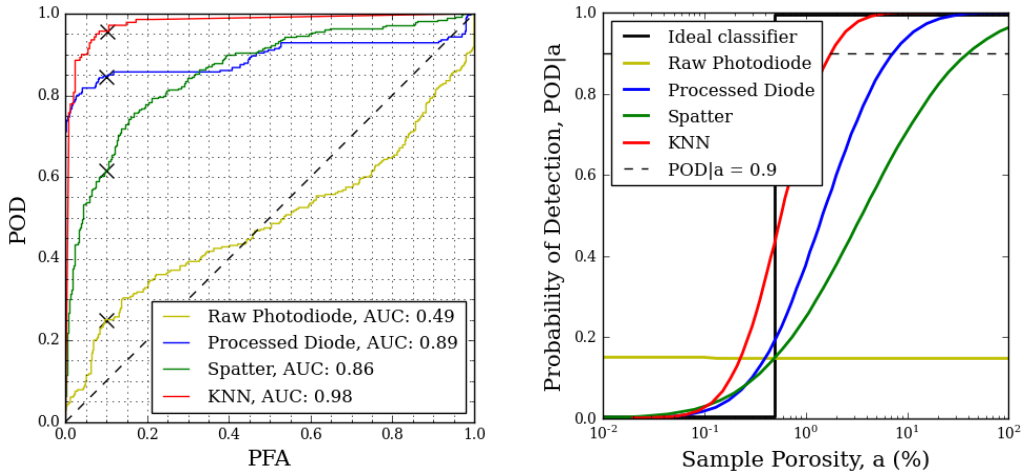
Figure 6: Micrographs of high and low porosity samples, fig. 6a and fig. 6b respectively. The samples are highlighted in red, the hatching areas are surrounded in green. The contour area is the region between these two lines.

4. Results

The results section has been split into 3 subsections assessing: system performance, discrimination at different porosity levels and the effect of different length scale averaging.

4.1. System comparison

Figure 7 uses the track level data to compare the four systems' ability to discriminate at the 0.5% level of porosity. This aimed to put the four systems in a level competition. The track level data was used as it had a sufficient number of data points with which to fit the KNN model while also offering significantly more effective classifiers than the melt pool level data. It also balances the performance of the four systems, since there was no discernible improvement in photodiode results when using layer data and the KNN and spatter results benefited from length scale averaging. The 0.5% level was chosen to represent the upper bound of porosity allowed in material which exhibits good mechanical properties. However, even lower porosity values were not chosen as the KNN results were clearly much better than those of any other system.



(a) ROC curves, defective $\geq$ 0.5% porosity.

(b) POD curves.

Figure 7: ROC and POD curves for raw and processed photodiode, spatter and KNN systems attempting to discriminate at 0.5% level of porosity. Marked with a black cross is the exact threshold which created the corresponding POD curves. The thresholds are chosen to all lie at a PFA of 0.1 in fig. 7a.

4.1.1. ROC curves

In fig. 7a an ROC curve is plotted for each system. The raw photodiode does not compete with any of the other systems. While the initial section is mildly above the non-discrimination line, the overall trend is very close in performance to non-discrimination and is reflected in the AUC score of 0.49. However, the poor performance of the raw photodiode is expected due to the modulated laser power input generating a noise dominated signal.

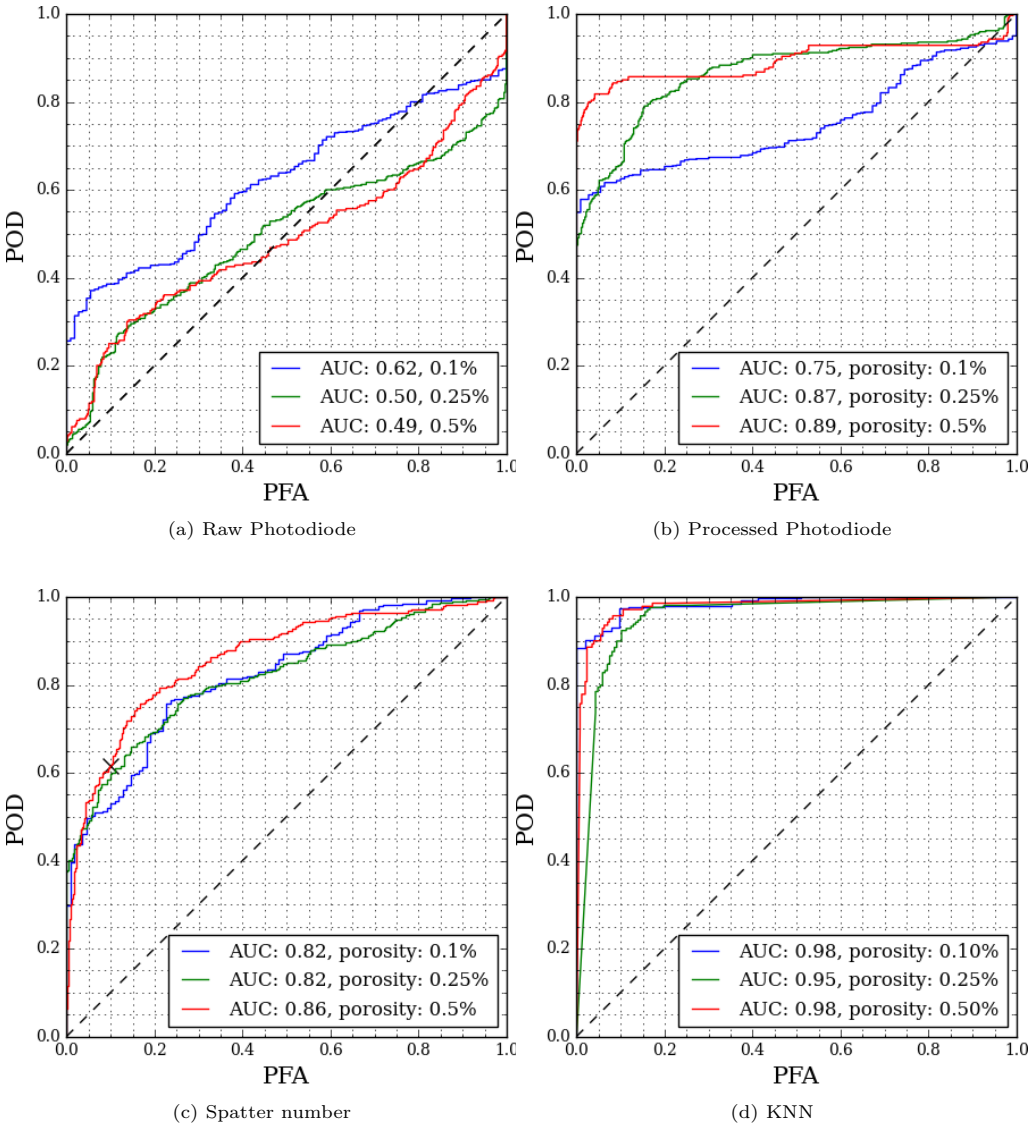
Spatter number (in green) displays a smooth, consistent curve with an AUC of 0.86. The processed photodiode has a similar AUC of 0.89. However, the shape of the processed photodiode curve (in blue) exhibits a very steep initial section, where it competes with the KNN, followed by several plateaus where it is eventually surpassed by the spatter system. The KNN rises steeply and continues this gradient for longer than the processed photodiode resulting in an AUC of 0.98, significantly higher than the other systems. At a PFA of 0.1 the KNN reaches a POD of 0.95, the processed photodiode reaches 0.84, spatter reaches 0.62 and the raw photodiode reaches 0.25. These points are marked with black crosses on each curve. All curves exhibit a broadly decreasing gradient across their lengths. However, the KNN's continued position above the other systems shows it is the optimal system [58].

4.1.2. POD curves

In fig. 7b a $POD|a$ curve is shown for each of the systems. The thresholds marked in fig. 7a all align at a PFA of 0.1. In this way, the system's probability of detection is compared for the same PFA. The KNN is the steepest line and is closest to the ideal step function. Processed photodiode and spatter number exhibit progressively less steep curves and the raw photodiode is almost flat. Despite being designed to detect porosity at the 0.5% level, the POD|0.5% porosity is 0.45 for the KNN, 0.20 for the processed

photodiode and 0.17 for spatter and the raw photodiode. Furthermore, reliable detection, when $POD|a = 0.9$ (where a is defect magnitude), is only reached by the KNN for 1.7% porosity, by the processed photodiode at 7%, spatter at 13%, while the raw photodiode never reaches this point.

4.2. Porosity Levels



(a) Raw Photodiode

(b) Processed Photodiode

(c) Spatter number

(d) KNN

Figure 8: ROC curves for each system when discriminating at 0.5, 0.25 and 0.1% porosity levels. All curves use the track level data (2,500 frames per data point).

Figure 8 shows the changes in detection outcome when discriminating at 0.5%, 0.25% and 0.1% porosity levels. The raw photodiode data in fig. 8a shows that at all levels the result is an AUC of approximately 0.5, although there is a slight improvement when detecting at the 0.1% level.

The spatter number in fig. 8c shows a small decrease in performance, initially having a high AUC score of 0.86 for 0.5%. This declines to 0.82 when moving to the 0.25% porosity level. However, the AUC plateaus and is the same for the 0.1% porosity level and the curve shapes are almost identical shapes. Ultimately, spatter seems relatively similarly effective across porosity levels.

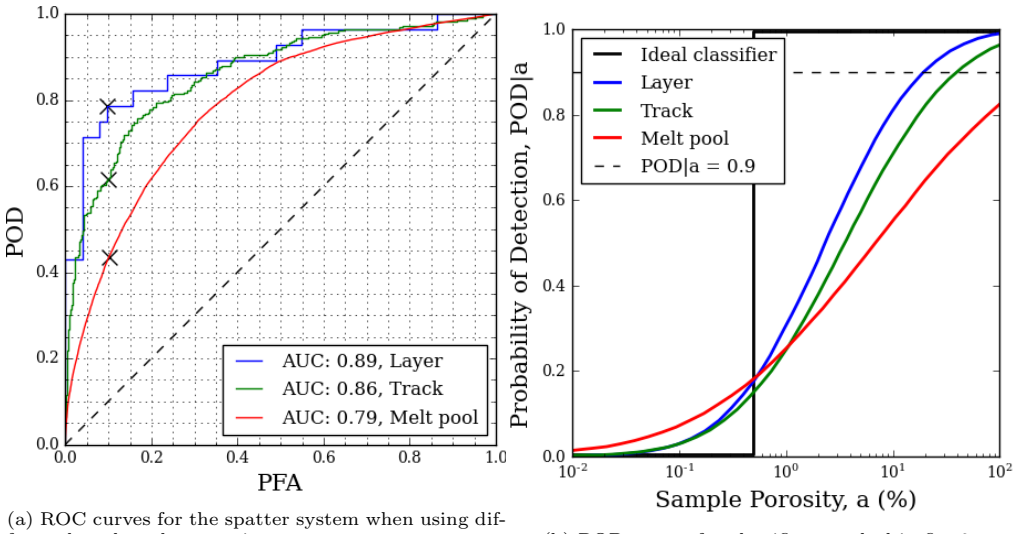
The processed photodiode data displayed in fig. 8b initially shows only a very moderate decrease in detection capability between 0.5% to 0.25% porosity levels from 0.89 to 0.87. However, the initial steep section is significantly shorter and performance in the low PFA range has significantly decreased even if the AUC drop is minor. Below 0.25% there is a significant drop to an AUC of 0.75, this results in a system which has a significantly lower AUC than the equivalent spatter based system.

Finally the KNN demonstrates consistent detection across porosity levels. While the shapes of the curves do subtly change, an AUC of 0.98 is achieved for 0.5% and 0.1% with a minor decrease to 0.95 at 0.25% porosity.

4.3. Length scale averaging

Figure 9 shows three ROC curves for the spatter number using layer, track and melt pool length scale averaging. The layer curve is the coarsest with the fewest points, with each point averaging over 25,000 frames. As a result it exhibits the highest AUC score. The track curve uses 2,500 frames per point and the curve is thereby significantly smoother when compared to the layer curve, while suffering a small penalty in AUC. The melt pool curve averages only 40 frames per reading and therefore has many more points from which to create the curve. Consequently, the curve is by far the smoothest. However, once again there is a detection penalty for using less data per point and the AUC drops to 0.79. For brevity only the spatter data is included. However, the effect is similar for the photodiode and KNN.

The POD curves shown in fig. 9b validates the trends shown fig. 9a, overall the track and layer curves almost similar with a small but discernable penalty. The melt pool curve has a higher chance of detection below the threshold and lower above the threshold when compared to the track and layer curves. This indicates that the melt pool scale averaging is a discernibly worse classifier than the layer and track alternatives despite having higher probability of detection at the specific defect threshold size.



(a) ROC curves for the spatter system when using different length scale averaging.

(b) POD curves for classifiers marked in fig. 9a

Figure 9: ROC and POD curves using different averaging lengths for the spatter number discriminating at the 0.5% porosity level. Layer: 25,000 frames, Track: 2,500 frames, Melt pool: 40 frames.

5. Discussion

5.1. ROC comparison

When interpreting ROC curves the question they appear to answer is: "What is the performance when detecting defects of magnitude X and above?". However, this is misleading. For purely binary problems this is the case. However, for problems which are continuous, much information is lost when binarising the data. It limits knowledge on the effects of defect magnitude beyond whether the defect is above or below the threshold.

Instead, the ROC curve tells us "Which classifier offers the optimum trade off between POD and PFA for this dataset". This includes classifiers which are derived from different thresholds of the same system as well as those from totally different systems. Furthermore, different datasets will create different results for the same classifier offering an insight into the generalisation of the system between different defect generating scenarios.

In the case study, ROC curves showed that the KNN outperformed the other sensors across all scenarios. ROC curves also allowed a fair choice of system thresholds based on PFA value, a dimension not captured on POD plots. ROC curves, and the POD, PFA and AUC values determined from them, may not serve as an absolute measure of a system's detection ability but they summarise the key dimensions of system performance to allow for comparison.

5.2. POD curve comparison

The primary question in industrial defect detection is not "what is the smallest defect detectable?" it is instead: "what is the largest defect which my system is likely to miss?" [40, 41, 42]. If one were to answer the first question it would appear from this case

study that the raw photodiode is significantly better than the other sensors at detecting porosity less than 0.1%. However, a flat $POD|a$ curve represents non-discrimination and is the equivalent of randomly picking 16% of samples to be defective.

When investigating the second question routine detection ($POD|a = 0.9$) is only reached by the KNN for parts which are 1.7% porous. This is despite the chosen KNN threshold having a POD of 0.95 when discriminating between samples with >0.5% porosity on the ROC plot. In the POD plot at 0.5% porosity the $POD|a$ is 0.45. Spatter and processed photodiode system struggled to a greater extent reaching reliable detection at 7% and 13% respectively. The porosity levels at which we are reliably discriminating are significantly worse than what could be termed good material. The POD curves also showed the total lack of discrimination of the raw photodiode data.

These questions highlight the benefits of POD curves; the detection probability is known for all defect magnitudes individually, allowing predictions of future performance to be estimated for different defect conditions in a manner not possible from ROC curves.

5.3. Photodiode vs camera based detection systems

Photodiode intensity can be affected by factors including the number, size and brightness of spatter particles and size and brightness of the melt pool. While photodiodes may be useful in detecting deviations from a known process baseline, they appear inappropriate to serve as an absolute measure of melt pool state. They could perhaps be used to measure relative deviation from a system baseline determined using a density cube included in each build.

High-speed imaging generates substantially more data which can be used to directly assess melt pool physics, such as melt pool size and spatter distribution. The KNN and photodiode both serve as a method of combining process signatures into a single number. However, it has proved preferable to do this using machine learning. This makes high-speed imaging a more direct measure of melt pool state and by extension more suitable as an absolute measure of part quality.

5.4. System spatial resolution

All measures of system performance presented have been assessments of the ability to predict part-wide porosity post build. While averaging across a full layer gives the best performance, the track length scale is the optimum since it has a negligible reduction in performance while greatly increasing the number of data points. This is despite the truth data for the study being derived from micrographs of samples which only yield a value for part wide, global porosity. This assumes the part is homogeneous, which was expected since the parts are small, simple, uniform shapes and well-distanced from each other on the bed to avoid interference.

A significant drop in performance was experienced by all systems when using the melt pool scale averaging. The melt pool length scale is much closer to the length scale on which individual pores form. Consequently each reading no longer represents the sample as a whole. Instead they relate to a specific defect or pore and are therefore subject to much greater variation. To properly assess the system results at this increased spatial resolution, a new referee technique of equal performance would be required, such as micro-CT.

5.5. Applicability and limitations

The methodology presented in this paper is applicable to any detection system and there is no requirement for the connection between sensor signal and defect to be deterministic or even logical in any way. The methodology makes no assumptions about how the detection of a defect is performed and is equally capable of evaluating results from deterministic interpretations of sensor signals as it is fitted stochastic models or "black box" machine learning methods. All that is required is that the detection system generates predictions, these can be continuous signals or discrete classifications since all of these can be reduced to binary classification.

For quantitative comparisons between detection systems the definition of what a defect is should be as similar as possible. For example, some systems seen in the literature detect "part-quality" as determined by laser parameters [27], while others detect individual pores [37]. ROC and POD curves can be generated for detectors of both of these defect types, however, any comparison would need to take account of the grossly different nature of the defects detected. Fundamentally, the limiting factor in quantitative comparison is the truth data, not the detectors themselves. As seen in this work, despite having a very high fidelity system it cannot be evaluated at a finer resolution than part wide porosity due to the truth data.

5.6. Truth data

The truth data for porosity was generated via sample sectioning and micrographs. Efforts where made to reduce the effect of intra-layer regional porosity differences between the contour and hatching sections. However, it was still assumed the part was homogeneous in the build direction. The systems derived POD is dependant on this phenomena, as a consequence the ROC and POD results are for measures of part wide porosity as estimated by the micrographs.

The cause of the highly porous parts (mainly those >1%) in this study was due to deliberate laser defocusing. As a result, not all the defects in this study were completely natural. Many of the most dense samples (<1%) were those with optimal focusing. Therefore, these results can be expected to generalise to laser powder bed fusion as a whole. However, those samples above 1% porosity represent only one method of gross process failure. Laser powder bed fusion can destabilise in a variety of ways and defects generated during one layer may not always be present at the end of a build due to remelting. While there would likely be a correlation between focus height and the detection systems used in this work, the results were correlated to porosity measured via micrographs post-build. The primary mechanism for porosity generation is expected to be laser defocusing, particularly in the highly porous samples. The systems effectiveness may be reduced when detecting porosity generated by other mechanisms.

Ultimately, the derived ROC and POD estimates are only as good as the truth data used to create them. A POD of 1 is not in fact perfect detection but instead the POD of the referee technique. Furthermore, the uncertainty in the truth data propagates to any derived measurements. Even if the system being assessed is more accurate than the referee technique, the ROC and POD results can only show the assessed system is as good as the referee technique, never better.

5.7. Desired levels of performance in AM scenarios

Depending on the application of the system or the resulting part there is an important trade off between POD and PFA. In the case of re-melting, a detection system may be used to discover defects enabling the defective regions to be remelted. The aim is to fix as many defects as possible and fixing false alarms raised by setting a low detection threshold incurs minimal additional cost, mainly time. Therefore detection systems and thresholds may be chosen which optimise for POD at the expense of PFA.

In the case of certification of high density components one would expect many more clear than defective regions, at least 100:1. A low PFA is vital in these circumstances in order to prevent excessive part rejection. In high quality material the overwhelming number of inspections should be clear (eg: 99.9% clear, 0.1% defective). Therefore, even with a relatively low PFA (eg: 0.01), one would expect a significant number of false alarms (1% of all inspections). Furthermore, even with a POD of 1, true detections would only make up 0.1% of total inspections, resulting in 10 out of 11 rejections being false alarms.

6. Conclusion

In this study we presented a methodology to quantitatively compare the detection capability of in-situ monitoring systems in additive manufacturing. We demonstrated this approach by comparing part porosity detection from photodiode data and melt-pool metrics from high-speed camera footage. Specifically raw and post-processed photodiodes, a single melt pool metric (spatter number) and a fusion of several metrics (describing the melt pool and spatter) using a K-Nearest Neighbours (KNN) algorithm.

ROC analysis demonstrated that the KNN performed better than the processed photodiode at a variety of porosity levels (0.1%, 0.25% and 0.5%) and length scales (layer, track and melt pool). The post-processed photodiode was marginally better than spatter number. The raw photodiode was found to have negligible improvement over non-discrimination. The KNN was shown to excel at the lowest porosity level (0.1%), a level where the processed photodiode and spatter were at their lowest levels.

Despite ROC analysis presenting some encouraging results for the systems, POD curve analysis showed that the KNN could only reliably detect part porosity when it was at 1.7% or higher ($POD|1.7\% = 0.9$). Furthermore, POD curve assessment showed that reliable detection $POD|a = 0.9$ for the other systems never exceeded 7% porosity. The POD curve assessment methodology allows us to conclude that none of the systems evaluated here are ready for industrial usage where levels of porosity of less than 0.5% are required.

Ultimately, we have shown that the ROC-POD methodology allows objective and quantitative assessments of detection performance, irrelevant of system mechanics. This facilitates measurable progress in the development of defect detection systems using in-situ data, enabling accelerated component qualification in additive manufacturing processes.

7. Acknowledgements

The authors thank AWE plc (contract 30338995) and EPSRC (EP/K503733/1 and EP/R513052/1) for their financial support for this research.

References

[1] H. Kunzmann, T. Pfeifer, R. Schmitt, H. Schwenke, A. Weckenmann, Productive metrology-adding value to manufacture, CIRP annals 54 (2) (2005) 155–168.

[2] Product Verification – Growing UK Productivity and Competitiveness -, Tech. rep., Aerospace Technology Institute (2017).
URL https://www.ati.org.uk/media/1uqlpyen/insight05-product-verification.pdf

[3] B. M. Colosimo, S. Cavalli, M. Grasso, A cost model for the economic evaluation of in-situ monitoring tools in metal additive manufacturing, International Journal of Production Economics 223 (2020) 107532. doi:https://doi.org/10.1016/j.ijpe.2019.107532.
URL https://www.sciencedirect.com/science/article/pii/S0925527319303597

[4] I. Campbell, O. Diegel, J. Kowen, N. Mostow, T. Wohlers, Wohlers Report 2021: Additive Manufacturing and 3D Printing State of the Industry; Annual Worldwide Progress Report, Wohlers Associates, 2021.

[5] S. Leuders, M. Thöne, A. Riemer, T. Niendorf, T. Tröster, H. A. a. Richard, H. J. Maier, On the mechanical behaviour of titanium alloy TiAl6V4 manufactured by selective laser melting: Fatigue resistance and crack growth performance, International Journal of Fatigue 48 (2013) 300–307.

[6] S. Leuders, T. Lieneke, S. Lammers, T. Tröster, T. Niendorf, On the fatigue properties of metals manufactured by selective laser melting–The role of ductility, Journal of Materials Research 29 (17) (2014) 1911–1919.

[7] T. Ronneberg, C. M. Davies, P. A. Hooper, Revealing relationships between porosity, microstructure and mechanical properties of laser powder bed fusion 316L stainless steel through heat treatment, Materials & Design 189 (2020) 108481.

[8] W. W. Wits, S. Carmignato, F. Zanini, T. H. Vaneker, Porosity testing methods for the quality assessment of selective laser melted parts, CIRP Annals - Manufacturing Technology 65 (1) (2016) 201–204. doi:10.1016/j.cirp.2016.04.054.

[9] J. Kruth, P. Mercelis, J. V. Vaerenbergh, T. Craeghs, Feedback control of Selective Laser Melting, Proceedings of the 3rd International Conference on Advanced Research in Virtual and Rapid Prototyping (2007) 1–7.

[10] J.-P. Kruth, J. Duflou, P. Mercelis, J. Van Vaerenbergh, T. Craeghs, J. De Keuster, On-line monitoring and process control in selective laser melting and laser cutting, in: Proceedings of the 5th Lane Conference, Laser Assisted Net Shape Engineering, Vol. 1, 2007, pp. 25–28.

[11] J.-P. Kruth, P. Mercelis, Procedure and apparatus for in-situ monitoring and feedback control of selective laser powder processing (8 2009).

[12] S. Berumen, F. Bechmann, S. Lindner, J. P. Kruth, T. Craeghs, Quality control of laser- and powder bed-based Additive Manufacturing (AM) technologies, in: Physics Procedia, Vol. 5, Elsevier B.V., 2010, pp. 617–622. doi:10.1016/j.phpro.2010.08.089.

[13] LPW Technology, Testing powder for optimal processing window, Tech. Rep. 0, LPW Technology LTD (2017).
URL https://am-uk.org/wp-content/uploads/2018/12/Case-Study-08-Powder-Processing-Window-FINAL.pdf

[14] B. M. Sharratt, Non-destructive techniques and technologies for qualification of additive manufactured parts and processes, Sharratt Research and Consulting Inc., Victoria, BC, Technical Report No. DRDC-RDDC-2015-C035. http://cradpdf. drdc-rddc. gc. ca/PDFS/unc200/p801800_A1b. pdf.

[15] S. K. Everton, M. Hirsch, P. I. Stavroulakis, R. K. Leach, A. T. Clare, Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing (2016). doi:10.1016/j.matdes.2016.01.099.

[16] M. Grasso, B. M. Colosimo, Process defects and in situ monitoring methods in metal powder bed fusion: A review, Measurement Science and Technology 28 (4). doi:10.1088/1361-6501/aa5c4f.

[17] C. Zhao, K. Fezzaa, R. W. Cunningham, H. Wen, F. De Carlo, L. Chen, A. D. Rollett, T. Sun, Real-time monitoring of laser powder bed fusion process using high-speed X-ray imaging and diffraction, Scientific Reports 7 (1) (2017) 1–11. doi:10.1038/s41598-017-03761-2.

[18] C. L. A. Leung, S. Marussi, M. Towrie, R. C. Atwood, P. J. Withers, P. D. Lee, The effect of powder oxidation on defect formation in laser additive manufacturing, Acta Materialia 166 (2019) 294–305. doi:10.1016/j.actamat.2018.12.027.

[19] R. Cunningham, C. Zhao, N. Parab, C. Kantzos, J. Pauza, K. Fezzaa, T. Sun, A. D. Rollett, Keyhole threshold and morphology in laser melting revealed by ultrahigh-speed x-ray imaging, Science 363 (2019) 849–852.
URL http://science.sciencemag.org/

[20] S. Ly, A. M. Rubenchik, S. A. Khairallah, G. Guss, M. J. Matthews, Metal vapor micro-jet controls material redistribution in laser powder bed fusion additive manufacturing, Scientific Reports 7 (1). doi:10.1038/s41598-017-04237-z.

[21] T. Furumoto, K. Egashira, K. Munekage, S. Abe, Experimental investigation of melt pool behaviour during selective laser melting by high speed imaging, CIRP Annals 67 (1) (2018) 253–256. doi:10.1016/j.cirp.2018.04.097.

[22] A. R. Nassar, M. A. Gundermann, E. W. Reutzel, P. Guerrier, M. H. Krane, M. J. Weldon, Formation processes for large ejecta and interactions with melt pool formation in powder bed fusion additive manufacturing, Scientific Reports 9 (1) (2019) 5038. doi:10.1038/s41598-019-41415-7.
URL http://www.nature.com/articles/s41598-019-41415-7

[23] J. Volpp, Spattering effects during selective laser melting, Journal of Laser Applications 32 (2) (2020) 22023.

[24] E. Malekipour, H. El-Mounayri, Common defects and contributing parameters in powder bed fusion AM process and their classification for online monitoring and control: a review, International Journal of Advanced Manufacturing Technology 95 (1-4) (2018) 527–550. doi:10.1007/s00170-017-1172-6.

[25] P. Yadav, O. Rigo, C. Arvieu, E. Le Guen, E. Lacoste, In situ monitoring systems of the SLM

process: On the need to develop machine learning models for data processing, Crystals 10 (6) (2020) 524.

[26] C. Wang, X. P. Tan, S. B. Tor, C. S. Lim, Machine learning in additive manufacturing: State-of-the-art and perspectives, Additive Manufacturing (2020) 101538.

[27] L. Scime, 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 25 (2019) 151–165. doi:https://doi.org/10.1016/j.addma.2018.11.010.
URL https://www.sciencedirect.com/science/article/pii/S2214860418306869

[28] I. A. Okaro, S. Jayasinghe, C. Sutcliffe, K. Black, P. Paoletti, P. L. Green, Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning, Additive Manufacturing 27 (2019) 42–53. doi:https://doi.org/10.1016/j.addma.2019.01.006.
URL https://www.sciencedirect.com/science/article/pii/S2214860418306092

[29] M. Grasso, B. M. Colosimo, A statistical learning method for image-based monitoring of the plume signature in laser powder bed fusion, Robotics and Computer-Integrated Manufacturing 57 (2019) 103–115. doi:https://doi.org/10.1016/j.rcim.2018.11.007.
URL https://www.sciencedirect.com/science/article/pii/S073658451830139X

[30] L. Scime, D. Siddel, S. Baird, V. Paquit, 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) 101453.

[31] S. A. Shevchik, G. Masinelli, C. Kenel, C. Leinenbach, K. Wasmer, Deep Learning for In Situ and Real-Time Quality Monitoring in Additive Manufacturing Using Acoustic Emission, IEEE Transactions on Industrial Informatics 15 (9) (2019) 5194–5203. doi:10.1109/TII.2019.2910524.

[32] G. Mohr, S. J. Altenburg, A. Ulbricht, P. Heinrich, D. Baum, C. Maierhofer, K. Hilgenberg, In-Situ Defect Detection in Laser Powder Bed Fusion by Using Thermography and Optical Tomography—Comparison to Computed Tomography (2020). doi:10.3390/met10010103.

[33] M. Montazeri, R. Yavari, P. Rao, P. Boulware, In-process monitoring of material cross-contamination defects in laser powder bed fusion, Journal of Manufacturing Science and Engineering, Transactions of the ASME 140 (11). doi:10.1115/1.4040543.

[34] M. Montazeri, A. R. Nassar, A. J. Dunbar, P. Rao, In-process monitoring of porosity in additive manufacturing using optical emission spectroscopy, IISE Transactions (2019) 1–16doi:10.1080/24725854.2019.1659525.
URL https://www.tandfonline.com/doi/full/10.1080/24725854.2019.1659525

[35] A. Gaikwad, F. Imani, H. Yang, E. Reutzel, P. Rao, In Situ Monitoring of Thin-Wall Build Quality in Laser Powder Bed Fusion Using Deep Learning, Smart and Sustainable Manufacturing Systems 3 (1) (2019) 20190027. doi:10.1520/SSMS20190027.
URL http://www.astm.org/doiLink.cgi?SSMS20190027

[36] S. Jayasinghe, P. Paoletti, C. Sutcliffe, J. Dardis, N. Jones, P. Green, PRE-PRINT: Automatic Quality Assessments of Laser Powder Bed Fusion Builds from Photodiode Sensor Measurements.

[37] J. A. Mitchell, T. A. Ivanoff, D. Dagel, J. D. Madison, B. Jared, Linking pyrometry to porosity in additively manufactured metals, Additive Manufacturing 31 (2020) 100946. doi:10.1016/j.addma.2019.100946.

[38] N. H. Paulson, B. Gould, S. J. Wolff, M. Stan, A. Greco, Correlations between thermal history and keyhole porosity in laser powder bed fusion, Additive Manufacturing (2020) 101213.

[39] B. D. Olin, W. Q. Meeker, Applications of statistical methods to nondestructive evaluation, Technometrics 38 (2) (1996) 95–112. doi:10.1080/00401706.1996.10484451.

[40] MIL-HNBK-1823, Non-Destructive Evaluation System Reliability Assessment, ASC/ENSI, 2530 Loop Road West, Bldg 560, Wright-Patterson AFB OH 45433-7101, 1999.

[41] G. A. Matzkanin, H. T. Yolken, Probability of detection (POD) for Nondestructive Evaluation (NDE), Tech. rep., NTIAC, Austin, Texas (2001).

[42] G. A. Georgiou, Probability of Detection (POD) curves: derivation, applications and limitations, Jacobi Consulting Limited Health and Safety Executive Research Report 454.

[43] N. A. Obuchowski, J. A. Bullen, Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine, Physics in Medicine & Biology 63 (7) (2018) 07TR01.

[44] S. M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, T. Back, M. Chesus, G. C. Corrado, A. Darzi, M. Etemadi, F. Garcia-Vicente, F. J. Gilbert, M. Halling-Brown, D. Hassabis, S. Jansen, A. Karthikesalingam, C. J. Kelly, D. King, J. R. Ledsam, D. Melnick, H. Mostofi, L. Peng, J. J. Reicher, B. Romera-Paredes, R. Sidebottom, M. Suleyman, D. Tse, K. C. Young, J. De Fauw, S. Shetty, International evaluation of an AI system for breast cancer screening, Nature 577 (7788) (2020) 89–94. doi:10.1038/s41586-019-1799-6.

[45] J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves, in: ACM International Conference Proceeding Series, Vol. 148, ACM Press, New York, New York, USA, 2006, pp. 233–240. doi:10.1145/1143844.1143874.
URL http://portal.acm.org/citation.cfm?doid=1143844.1143874

[46] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters 27 (8) (2006) 861–874. doi:10.1016/j.patrec.2005.10.010.

[47] D. G. A. E. Sturges, Elements of Detection Theory applied to NDE (1993).

[48] T. Calders, S. Jaroszewicz, Efficient AUC optimization for classification, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 4702 LNAI, Springer, Berlin, Heidelberg, 2007, pp. 42–53. doi:10.1007/978-3-540-74976-9{_}8.

[49] P. W. Hovey, A. P. Berens, Statistical evaluation of NDE reliability in the aerospace industry, Review of Progress in Quantitative Nondestructive Evaluation 7 B (1988) 1761–1768. doi:10.1007/978-1-4613-0979-6{_}108.

[50] M. Montazeri, P. Rao, Sensor-based build condition monitoring in laser powder bed fusion additive manufacturing process using a spectral graph theoretic approach, Journal of Manufacturing Science and Engineering 140 (9).

[51] P. A. Hooper, Melt pool temperature and cooling rates in laser powder bed fusion, Additive Manufacturing 22 (May) (2018) 548–559. doi:10.1016/j.addma.2018.05.032.
URL https://doi.org/10.1016/j.addma.2018.05.032

[52] L. R. Goossens, Y. Kinds, J.-P. Kruth, B. Van Hooreweder, On the influence of thermal lensing during selective laser melting, in: Solid Freeform Fabrication Symposium Proceedings (SFF Symp 2018), Austin; University of Texas, 2018, pp. 2267–2274.

[53] M. Saunders, Focusing on laser melting performance, Tech. Rep. 0, Renishaw plc. (2017).

[54] J.-P. Kruth, L. Froyen, J. Van Vaerenbergh, P. Mercelis, M. Rombouts, B. Lauwers, Selective laser melting of iron-based powder, Journal of materials processing technology 149 (1-3) (2004) 616–622.

[55] M. Mani, B. M. Lane, M. A. Donmez, S. C. Feng, S. P. Moylan, R. Fesperman, Measurement Science Needs for Real-time Control of Additive Manufacturing Powder Bed Fusion Processes, Tech. rep., NIST (2015). doi:dx.doi.org/10.6028/NIST.IR.8036.

[56] S. A. Khairallah, A. T. Anderson, A. M. Rubenchik, W. E. King, Laser powder-bed fusion additive manufacturing: Physics of complex melt flow and formation mechanisms of pores, spatter, and denudation zones, Acta Materialia 108 (2016) 36–45. doi:10.1201/9781315119106.
URL http://dx.doi.org/10.1016/j.actamat.2016.02.014

[57] C. L. A. Leung, S. Marussi, R. C. Atwood, M. Towrie, P. J. Withers, P. D. Lee, In situ X-ray imaging of defect and molten pool dynamics in laser additive manufacturing, Nature Communications 9 (1) (2018) 1–9. doi:10.1038/s41467-018-03734-7.

[58] F. Provost, T. Fawcett, Robust classification for imprecise environments, Machine Learning 42 (3) (2001) 203–231. doi:10.1023/A:1007601015854.