In-Line Drift Detection Using Monitoring Systems and Machine Learning In Selective Laser Melting

Pinku Yadav[a,b], Vibhutesh Kumar Singh[d], Thomas Joffre[c], Olivier Rigo[a], Corinne Arvieu[b], Emilie Le Guen[b], Eric Lacoste[b]

[a]*SIRRIS, Rue du Bois Saint-Jean 12, 4102 Seraing, Belgium*

[b]*University of Bordeaux, CNRS, Arts et Metiers Institute of Technology, Bordeaux INP, INRAE, I2M Bordeaux, F-33400 Talence, France*

[c]*IPC - Centre Technique Industriel de la Plasturgie et des Composites, Rue Pierre et Marie Curie 2, 01100 Bellignat, France*

[d]*University College Dublin, Belfield, Dublin 4, Ireland*

Abstract

Direct metal laser sintering, an additive manufacturing technique, has a huge growing demand in industries like aerospace, biomedical, and tooling sector due to its capability to manufacture complex parts with ease. Despite many technological advancements, the reliability and repeatability of the process is still an issue. Therefore, there is a demand for in-line automatic fault detection and post-processing tools to analyze the acquired in-situ monitoring data aiming to provide better quality assurance to the user.

This study focuses on the treatment of the data obtained using the EOSTATE Optical Tomography monitoring system. A balanced dataset is obtained with the help of computer tomography of the certified part (Stainless Steel CX cylindrical samples), through which a feature matrix is prepared, and the layers of the part are classified either having "Drift" or "No-drift". The model is trained with the feature matrix and tested on benchmark parts (Maraging Steel) and on an industrial part (knuckle, automotive part) manufactured in AlSi10Mg. The proposed semi-supervised approach shows promising results for presented case studies. Thus, the semi-supervised machine learning approach, if adopted, could prove to be a cost-effective and fast approach to post-process the in-situ monitoring data with much ease.

Keywords: Direct Metal Laser Sintering, Quality Assurance, In-situ Monitoring, Machine Learning, Semi-supervised Learning, Optical Tomography;

1. Introduction

According to ISO/ASTM 52900, the term "Additive manufacturing" defined as the "process of joining materials to make parts from the 3D model data, usually layer upon layer, as opposed to subtractive manufacturing and formative manufacturing methodologies" [1]. Additive manufacturing (AM) has enormous potential in aerospace, medical, and automotive industries due to design and material flexibility. Nowadays, AM technology used for not only prototyping but also small batch production with improved repeatability and reliability. According to the Wohlers Report, the market of AM manufactured parts has grown to $2.9 bn as compared to $2.1 bn in 2016 [2]. In recent years, growing demand for AM parts pushed for industrial development for better quality assurance and monitoring of the process. Therefore, it is viable to implement an in-situ monitoring and feedback control loop to control the process for better reliability and part quality [3]. Apart from the process monitoring, there are several other advantages of these monitoring systems, such as no significant effect on the overall printing time, no need for post-analysis of the part using expensive techniques like CT/μCT. Even though the advancements in the monitoring systems using high precision sensors such as photodiodes, high-speed infrared cameras, and pyrometers, anomaly detection during the process is still a challenge. As the final quality of the part depends on hundreds of controllable factors (machine parameters, process parameters) starting from powder morphology but can also be influenced by uncontrollable factors such as non-uniformity of powder bed due to failure of recoater, failure of part due to high residual stresses [4]. Also, another motivation for better monitoring systems is in the biomedical and tooling sector, where every part printed is different, so it is complicated to have a reference part for comparison, unlike in case of automotive and aerospace sectors. Furthermore, the different geometrical features such as overhang, sharp corners, down-skin, and up-skin are also critical as they create variability of the measured signals [5-6]. One of the challenges that existing commercial in-situ monitoring systems pose is the generation of the enormous size of the data, which is very difficult to store and analyze. For example, for a batch of 10 mm height cylindrical samples, a data of 100 GB is being generated. Therefore, the need for a fast and easy approach to post-process the data and easy detectability of drift is inevitable. Thus, the use of machine learning techniques in the AM process has increased in the last few years.

This paper specifically focuses on Laser Powder Bed Fusion (L-PBF), which is one of the metal AM techniques, also known as Direct metal laser sintering (DMLS) and Selective laser melting (SLM). The monitoring system acquires data through an off-axis Infra-red camera-equipped installed in the EOS M290 machine, where bandpass filters are used to monitor the melt pool radiations eliminating the reflected laser radiations. Establishing the link

between the acquired data and part quality is a challenging task. However, in recent years many researchers have used machine learning algorithms and computer vision techniques on large datasets for decision making [7-8]. Grasso *et al*. [9], detected the over melting via a statistical comparison between layer-wise emitted light intensity profiles. Some studies have also linked the statistical descriptors of spatters (spatter count and size) and vapor plume with the processing parameters and their influence on flaw formation during printing [10-11]. Khanzadeh *et al.* [12] proposed supervised learning based model to predict the porosity. The pattern from the individual melt pool images is extracted to predict the probability of class labels. Recently, Okaro *et al.* proposed another perspective of the ML approach for data treatment and predicted the quality of the parts based on their mechanical property as a descriptor. Key features were extracted from the photodiode signals, and a semi-supervised classification algorithm called "Gaussian Mixture Model-Expectation-Maximization (GMM-EM)" was applied to classify the samples as "faulty" and "acceptable" based on the ultimate tensile strength of the tensile bars. This approach showed the possibility for automatic certification of the L-PBF parts based on their mechanical properties. But Okaro *et al.* method does not illustrate any possibility to detect the defects in the parts, which is the cause for inferior mechanical properties [13]. Whereas, Scime *et al.* [14] studied the flaw detection using the powder bed layer images. They correlated the effect of the bad powder layer on the final part quality using a computer vision (CV) algorithm. Therefore, ML approaches can be promising steps to certify the quality of the final parts. It can be used as an additional verification step with currently used techniques such as computer tomography.

In this work, a commercially available EOSTATE Exposure Optical Tomography (OT) module installed on EOS M290 is used for in-situ monitoring. The acquired in-situ data is post-processed to check for drift layers in an automotive part and benchmark part using a machine learning (ML) based classification approach. The "drift layer" is classified based on the localized hotspots being generated due to localized variation in melt pool signatures in a particular layer of the part, which significantly affects the final part quality. Any ML problem, in theory, can be broadly classified into two categories: (I) supervised and (II) unsupervised. In a supervised ML approach, a set of labeled data is used to train the classifier, and the trained classifier is used to predict the outcome of unlabeled data. In contrast, an unsupervised approach is used to ascertain labels through correlations/patterns in unlabeled data and is explicitly dependent on the distinction in the dataset. Unsupervised learning is also helpful for labeling the unknown data or can be used to organize the data in clusters, anomaly detection, or association. A more detailed explanation about these approaches could be found in the literature [15-16]. However, there exist an ML approach, which utilizes

both the supervised and unsupervised ML approaches in a particular pipeline, called "Semi-supervised learning approach." The semi-supervised learning approach exists since, in many real-world applications, one cannot directly apply either a supervised or unsupervised as the availability of a fully labeled dataset is scarce [17]. The main principle behind a semi-supervised ML approach is that an unsupervised learning approach is first applied to the unlabeled dataset to obtain initial labels, and to obtain learning parameters like suitable distance metric, .etc. Then the newly labeled data is used to train a supervised learning approach based classifier, a similar approach could be found in the literature [18-20]. This hybrid pipelined approach is beneficial for applications such as additive manufacturing, where getting labeled data is a challenge.

In this study, we propose a semi-supervised learning-based approach to post-process and detect the drift in the acquired in-situ OT data. Firstly, an unsupervised learning algorithm (*K*-means clustering) is used to expand the available limited labeled dataset and finding the best-suited distance matric to generate a training dataset. Secondly, a supervised learning-based classification algorithm (*k*-nearest neighbor) is trained with the labeled dataset (obtained from *K*-means clustering) and tested with real case studies. In this study, the in-situ monitoring dataset is treated as a binary classification problem, and the labeled data have labels, namely: "no-drift" and "drift," which signifies the possibility of drift in a particular layer. The following points signify the important contributions of this paper:

- A decision on the training features from the dataset based on the histogram mapping of the layer-wise OT data and validated with CT data.
- Use of *K*-means clustering for labeling a limited labeled dataset based on the selected data features. The cross-validation with CT data is performed to decide the suitable distance metric and other classification parameters.
- Use of the labeled data to train/test a *k*-nearest neighbor (*k*-NN) classifier and judge its accuracy.
- The application of the trained *k*-NN classifier, on case studies of complex geometry parts and cross-validation of *k*-NN results using the EOSTATE Exposure OT analysis tool.

2. Definitions

- ***Hot spot:*** The local areas with high light intensity compared to the rest of the layer. These hotspots areas indicate the highest probability of defect occurrence in the final part.
- ***Drift:*** The drift layer is an indication of the presence of the hotspots in a particular layer. If the local hotspots are present in the layer, the whole layer is termed as the "Drift layer."

- ***No-Drift:*** In contrast to the "Drift layer," the non-existing of hot spots in a particular layer, termed as "No-drift layer."
- ***Feature:*** In machine learning, a feature is a measurable property that can be quantified and recorded. Features are extracted from the input data in a way to simplify the classification task. In our application, the mean and median are the feature extracted from each layer's image.
- ***Feature matrix:*** A $n \times m$ dimensional matrix, where n is the number of data, and m is the number of features. In other words, a feature matrix contains n number of $1 \times m$ dimensional feature row vectors.
- ***Balanced dataset:*** A dataset having an equal number of layers for each label ("drift" and "no-drift").

3. Background and methods

This section confers the theoretical explanation for the used ML algorithms and in-situ monitoring systems used for monitoring the process. In later sections, the link with our work will be established.

3.1. *In-situ monitoring techniques*

Despite many technological advantages in AM over conventional manufacturing techniques like casting, machining, etc., the process capability still represents a major limitation to their industrial breakthrough. The certification standards for AM parts are still under development, and process stability is a major concern for many risk-averse industries [21]. Therefore, in recent years, both researchers and machine manufacturers have shifted their focus on developing more reliable monitoring systems to monitor the process stability and detect the possible drifts during the process. Thus, it led to the development of different monitoring systems such as layer control system, which is mainly dedicated to powder bed spreading, laser power control, melt pool monitoring, acoustic methods, and ultrasonic methods [22-23]. In this study, we will focus on melt pool monitoring systems, which mainly rely on the detection of electromagnetic radiation from the melt pool. In-situ monitoring systems can be categorized into two classes principally based on their layout in the machine: Coaxial/On-axis systems and Off-axis systems.

The co-axial monitoring systems monitor the optical path during printing using high precision sensors such as photodiodes, high-speed cameras, and pyrometers, [23]. The spectral sensitivity of these systems varies from manufacturer to manufacturer, but most of the systems have spectral detection range in the near-infrared region. A bandpass filter is used to eliminate the detection of radiation from the visible spectrum and laser backscattering from

the build chamber. Pavlov *et al.* [24] used on-axis two InGaAs photodiodes with transmission spectrum of 1.26 μm and 1.4 μm, respectively, to monitor the temperature of the laser spot while printing. The relation between photodiode signal response and process parameters such as layer thickness, hatch distance, and scan strategies established. Berumen *et al.* [25] studied the heat accumulation correlation with the varying layer thickness, sharp contours, and overhang structures using an on-axis photodiode with a spectral range of 780 nm up to 900 nm.

In contrast to on-axis systems, the off-axis systems are spatially stationary to laser and do not follow the optical path. For example, Kleszczynski *et al.* [26-27] studied variation on volumetric energy during printing using off-axis high-resolution CCD camera. The sensors are placed outside the optical path and have stationary field of view i.e. the whole build plate. Therefore, the geometric correction factor is always applied to the acquired images [28-30]. In our study, we will be post processing the data from the off-axis sCMOS camera.

It is imperative to illustrate that the sensitivity of these systems depends on various factors, such as process parameters, material, and other unwanted factors. For a more detailed review of various in-situ monitoring systems, interested readers can refer to reference [23]. Some of the commercially available melt pool monitoring systems are listed in Table 1.

Table 1: List of commercially available monitoring systems.

Manufacturer	Product	On/Off axis	Sensors
SLM Solutions	Melt pool monitoring (MPM)	On-axis	Two photodiodes
EOS GmbH	EOSTATE	On and Off-axis	Photodiode and high-speed sCMOS camera
Renishaw	MeltVIEW	On-axis	Photodiode
Concept GE	QM melt pool	On-axis	Infra-red camera and photodiode
Stratonics	ThermaVIZ	Off-axis	CMOS camera
3D Systems	DMP monitoring	Off-axis	Infrared camera and photodiode
Plasmo	-	On-axis	Infrared camera and photodiodes

3.2. *Machine learning*

The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience. In this section, we will define through mathematical modelling the machine learning algorithms utilized in this work.

3.2.1. *K-means clustering*

K-means clustering or Lloyd's algorithm [31] is one of the most widely studied unsupervised machine learning algorithms in the literature. It is an iterative partitional clustering algorithm that divides *n* data/feature samples into *K* disjoint clusters that minimize the squared error criterion, and each cluster is characterized through a centroid. It is important to note that the initial centroid seeds play an important role in *K* - means clustering algorithm, as different initial centroids can provide different results [32]. We have utilized K-means++ algorithm [33] to choose initial centroid seeds which achieve faster convergence to a lower sum-of-squares point-to-cluster-centroid distance than Lloyd's algorithm but offers no guarantees of optimality. The parameter *K* (i.e., number of clusters) is *apriori* selected, which in our case is *2*. The iterative steps for *K*-means ++ algorithm are mentioned in the flow graph shown in Figure 1.

Let us define a $n \times m$ dimensional dataset/feature matrix $D = [x_1, x_2, \dots, x_n]^T$ with n samples, and the set of K centroids be $C = [c_1, c_2, \dots, c_K]^T$, where x and c are necessarily represented points on an m-dimensional plane. The steps used to implement the *K*-means++ algorithm are as follows:

1. Select an observation x_a uniformly at random from the dataset/feature matrix, D. The chosen observation is the first centroid denoted as c_1.
2. Denoting the distance from an i^{th} data point x_i to j^{th} centroid c_j as $d(x_i, c_j)$. Now we compute distances from each observation to c_1, i.e., $d(x_i, c_i)$.
3. Now we choose the next centroid c_2 with a probability of $\frac{d(x_t,c_1)^2}{\sum_{j=1}^{n} d(x_j,c_1)^2}$. This step implies if the data point x_t is near to centroid $c_1$the likelihood of x_t to become centroid $c_2$would be negligible.
4. We now repeat step 3 until all initial *K* centroids get assigned, i.e., $c_3, \dots, c_K$.
5. Now for each $l \in \{1,2, \dots, K\}$, set the cluster C_l to be the set of data/feature points in D that are closer to C_l then they are to C_p, $\forall l \neq p$.

6. Now for each $l \in \{1,2,\ldots,K\}$, update the centroid c_l to be the center of mass of all data points in C_l, i.e., $c_l = \frac{1}{|\mathsf{C}_l|} \sum_{\{x_i \in \mathsf{C}_l\}} x_i$. This process is equivalent to calculate a mean, and hence the name *K-means clustering*.

7. Now iteratively repeat steps 5 and 6 until cluster assignments do not change.

In *K*-means clustering, the distance metric d plays an important role, and in the literature, various distance metric has been suggested, e.g., Euclidean, correlation, cosine, and city block distance. In our work, we have experimented with various distance metrics to find the most suitable one for our application.

3.2.2. k-Nearest Neighbor *Classifier*

The k-nearest neighbor (k-NN) is a supervised learning classification algorithm, which is one of the most studied classifiers in the literature, and even with a simplistic formulation, it has a performance at par the most complex classifier available [34]. The k-NN learning algorithm requires labeled training data, and a predefined value of the

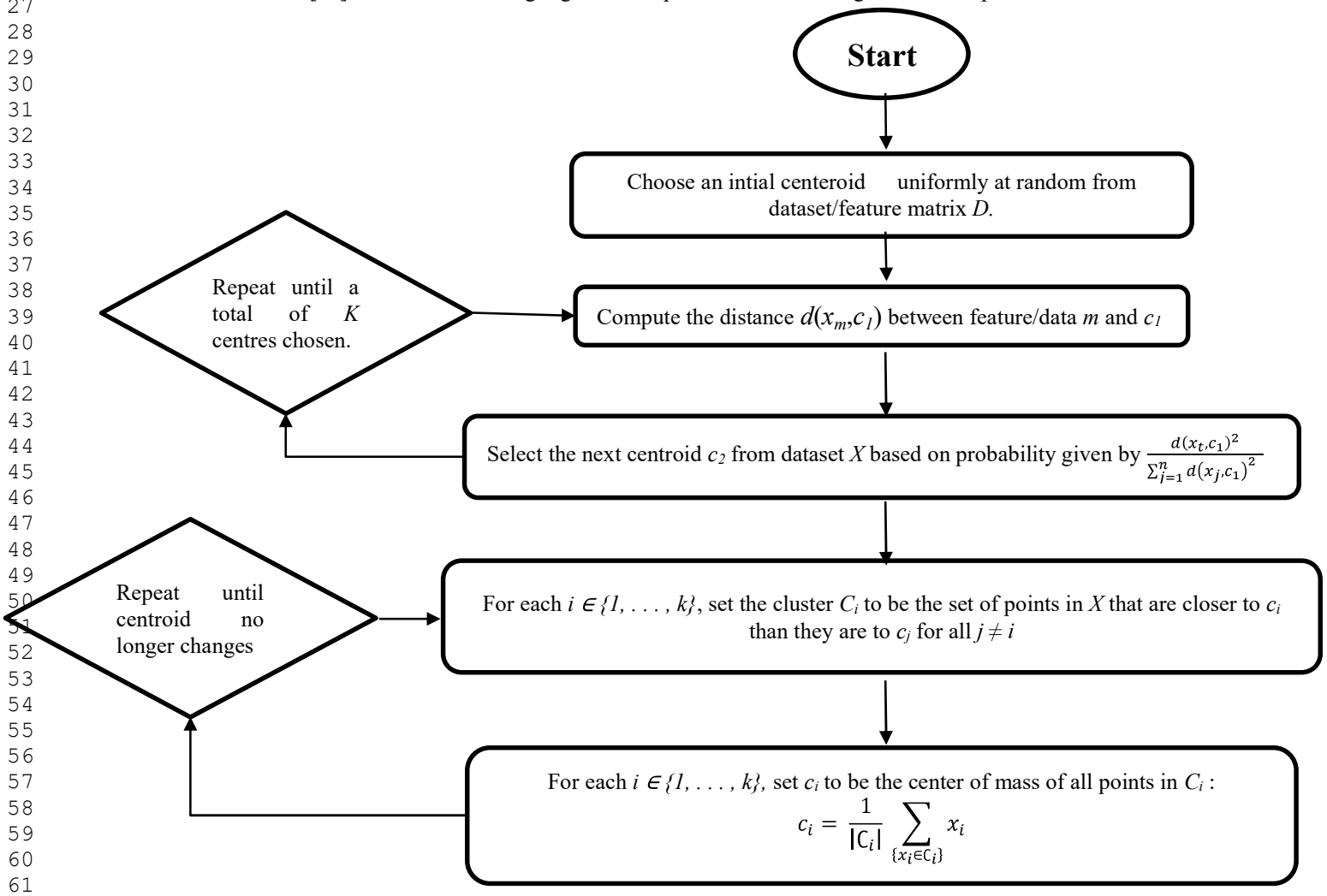


Figure 1: Working principle of *K*-means ++ algorithm.

number of nearest neighbors parameter k, which is used by the classifier to find k-nearest neighbors to a query data (unclassified data), based on a distance metric. The k-nearest neighbors can have different classes, and the algorithm predicts the class of the query data as the majority class of nearest neighbors. Let $T = (x_{i'}, L_{i'})\ \forall\ i' = 1, 2, \cdots, N$ denote the training set, with N samples and $x_{i'} \in R^m$ is a m-dimensional training feature vector have a known class label $L_{i'}$.

A query data, $x'_{i'}$ be an m-dimensional vector of data or features, to which a label $L'_{i'}$ will be assigned. Now let $T' = \left(x^k_{i'}, L^k_{i'}\right)$ for $i' = 1,2, \ldots, k$, denotes the set of k-nearest neighbors based on a distance metric and based on the majority class of T', $d'_{i'}$ will be assigned a label, i.e.,

$$L'_{i'} = \arg max_L \sum_{\left(x^k_{i'}, L^k_{i'}\right) \in T'} \delta_{L,\, L^k_{i'}},$$

where $\delta_{L, L^k_{i'}}$ is a Kronecker delta function.

3.2.3. *Distance Metrics*

As both the supervised (*k-NN*), as well as unsupervised (*K*-means) learning algorithms, rely on a distance metric, in this section, we mathematically define them. For any two points $x'_{i'}$ & $x^k_{i'}$ in an m-dimensional space, the *Euclidean distance* can be defined as,

$$d\left(x'_{i'}, x^k_{i'}\right) = \sqrt{\left(\left(x'_{i'} - x^k_{i'}\right)\left(x'_{i'} - x^k_{i'}\right)^T\right)},$$

City Block distance as

$$d\left(x'_{i'}, x^k_{i'}\right) = \sum_{n=1}^{\mathrm{m}} \left|x'_{i'}[n] - x^k_{i'}[n]\right|,$$

Cosine distance as

$$d\left(x'_{i'}, x^k_{i'}\right) = \left(1 - \frac{x'_{i'}\, x^{k^T}_{i'}}{\sqrt{\left(x'_{i'} x'^T_{i'}\right)\left(\, x^k_{i'}\, x^{k^T}_{i'}\right)}}\right),$$

and *Correlation distance as*

$$d(x'_{i'}, x^k_{i'}) = \left(1 - \frac{(x'_{i'} - \underline{x}'_{i'})\,(x^k_{i'} - \underline{x}^k_{i'})^T}{\sqrt{(x'_{i'} - \underline{x}'_{i'})\,(x'_{i'} - \underline{x}'_{i'})^T}\sqrt{(x^k_{i'} - \underline{x}^k_{i'})(x^k_{i'} - \underline{x}^k_{i'})^T}} \right),$$

where $\underline{x}'_{i'} = \frac{1}{m}\sum_{n=1}^{m} x'_{i'}[n]$ and $\underline{x}^k_{i'} = \frac{1}{m}\sum_{n=1}^{m} x^k_{i'}[n]$.

4. Experimental procedure

4.1. *In-situ monitoring module*

For our study, we used the EOSTATE melt pool monitoring module installed on EOS M290 for which the schematic diagram is shown in Figure 2. The optical tomography (OT) comprises an off-axis scientific Complementary metal-oxide-semiconductor (sCMOS) camera with a spectral detection range in the near infrared region (NIR). Usually, the radiation from the build chamber consists of three components, i.e., backscattered laser (1064 nm), plasma radiations due to evaporation and ionization of gases (400-600 nm) and thermal emissions which range from visible (380-780 nm) to near-infrared (~1400 nm) spectrum [35]. Since the part quality mainly depends

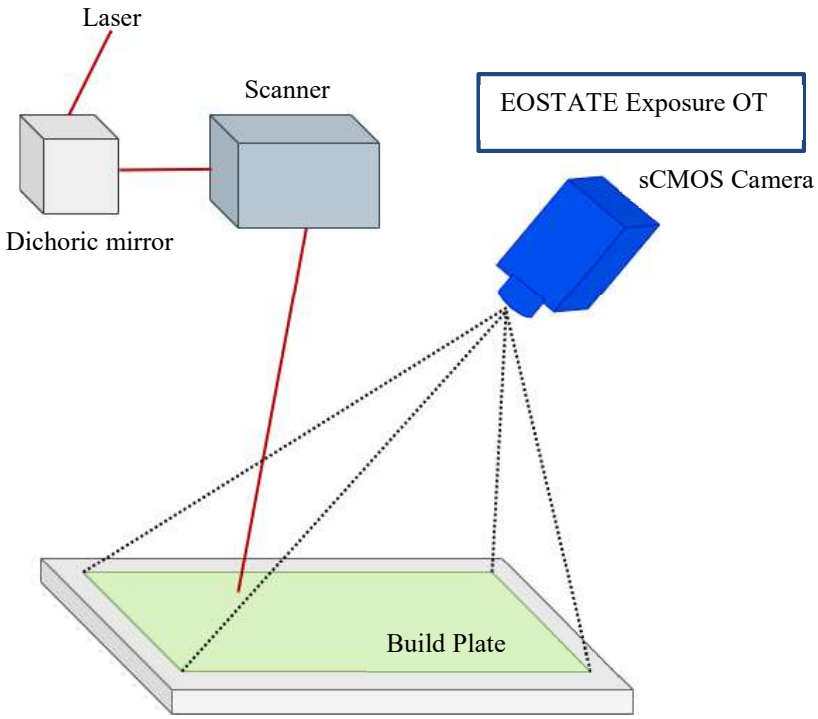


Figure 2: Schematic diagram of the EOSTATE exposure module.

on the thermal emissions from the melt pool, so other wavelength radiations are eliminated by installing a bandpass filter (the type of bandpass filter cannot be relieved due to the machine provider confidentiality clause) in front of the camera. The OT system has a camera resolution of 2560 x 2160 pixels, which allows achieving a spatial resolution of 125 µm/pixel across the entire build platform. The OT system has a frame rate of 10 frames per second. At the end of every scanned layer, all the images are superimposed, and a holistic picture for the whole layer is saved in a 16 bit TIFF single image. The final image of the particular layer represents the process map that can be correlated with the emitted light intensity by the process. Due to non-centralizing position of the camera, the geometric correction was applied to the images via software.

4.2. *Materials and methods*

In our study, a total of 18 Stainless Steel CX cylindrical samples with a diameter of 10 mm and a height of 15 mm were printed on EOS M290. The input volumetric energy density was varied over a range to prepare a dataset of certified samples that would be used further for training and verification of the model. The deliberately varied process parameters induce drift during the process which results in varied porosity level in the parts. The optimized printing process parameters are summarized in Table 2. The laser power and laser scan speed were varied ± 30 % from the optimized processed parameters tabulated in Table 2. Also, the layer thickness was varied to 30, 60 and 90 µm. It is worth noting that the possibility of having drift due to bad powder layer spreading is not taken into account for this study. To prepare a labeled dataset for training, computed tomography was performed on printed samples for analysis using an X-ray inspection system for determining pores with a 180 kV microfocus tube and an area detector with a voxel size of 19 or 22.5 microns. The smallest evaluated pore has a volume of 5 voxels. For testing the supervised learning classifier, two case studies were chosen, which are: an automotive part called “Knuckle” and a benchmark part called “Overhang structure” (Figure 3). The automotive part called “knuckle” is printed with AlSi10Mg, and benchmark part is printed with Maraging steel. The “knuckle” and “overhang part” were chosen as a case study because the location of the overheating failure in these parts was detectable via visual inspection.Thus we made the choice not to carry out CT scan for this first study. The analysis tool from EOS can indeed detect cold and hotspots, which are treated as areas of having the highest probability of finding a defect in the final part. EOSTATE Exposure OT analysis tool is then used for cross-validation of results obtained from the supervised learning approach for the presented case studies.

Table 2: Process parameters for the printing.

Material	Power (W)	Scanning speed (mm/s)	Hatch distance (µm)	Layer thickness (µm)	Energy density (J/mm^3)
Stainless Steel CX	260	1000	100	30	80

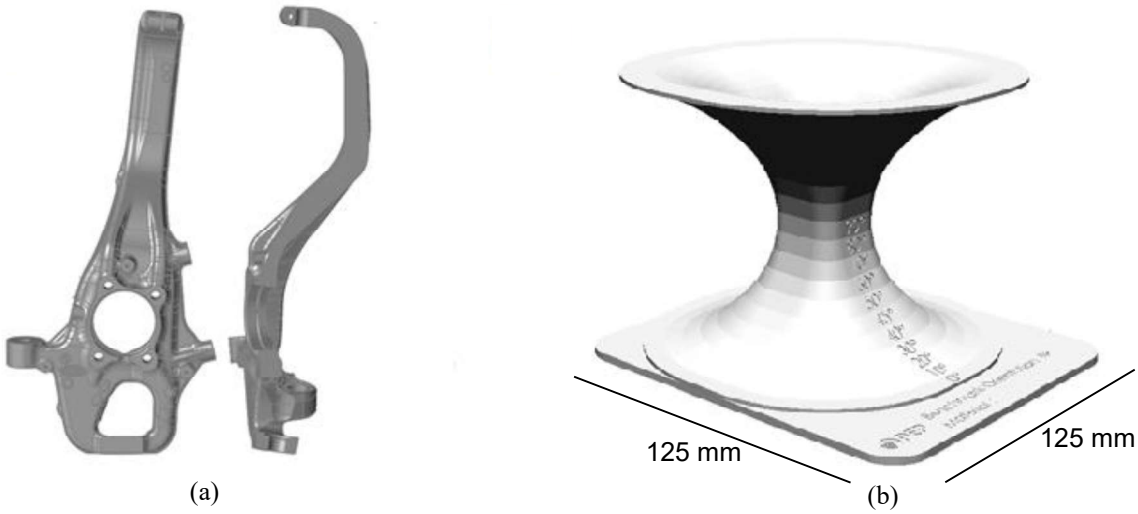


Figure 3. The CAD design of (a) An automotive steering knuckle (3d printed within the framework of the European project Maestro [36]), the length of this part is approximately 500 mm, only a section was 3D printed, (b) benchmark used to determine the critical overhanging angle in additive manufacturing processes.

4.3. Image analysis

Clijsters *et al.* [37] used a co-axial setup (CMOS camera and photodiode in NIR range) to capture the melt pool signatures such as melt pool area and intensity to monitor the quality of the SLMed parts. Clijsters *et al.* prepared a dataset based on the melt pool area and intensity for different scan vectors such as fill scan and contour scan. It is presented that the heat transfer depends mainly on the environment of the melt pool: the heat flux is higher when the metal pool is surrounded by printed material than when it is surrounded by powder. For each of these classes (fill scan and contour scan), a confidence interval was defined, and errors were detected based on the defined confidence interval. The proposed method was studied for small cube structures**.** However, in our study, we also noticed such a confidence window based on mean and standard deviation could not be defined for complex and real case parts. To prove this, the mean and variance of each layer for the complex part called "Knuckle" (refer Figure 3 (a) for part geometry) are plotted in Figure 4. It can be observed that the mean and standard deviation are dependent on the printed area of each layer and changes with the shape of the complex part. Therefore, the global threshold limit based on the

mean and standard deviation cannot be applied for the complex geometries, and thus the ML is suitable approach for post processing of in-situ data. For training our ML classifier, it is then required to select suitable features from the given dataset.

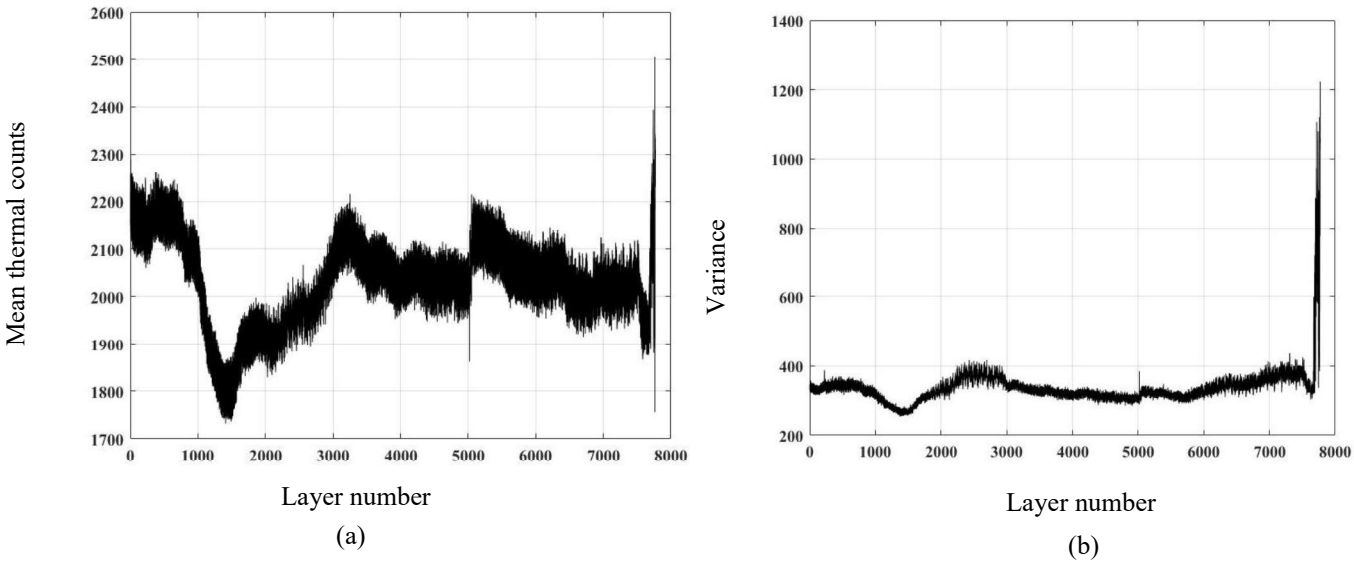


Figure 4: (a) Mean, and (b) variance per layer for the case study part "Knuckle".

Preprocessing of the OT images is a necessary step before feature extraction and further ML processing. As the OT image captures the whole build plate, the specific part region is cropped. The removal of the background from the cropped part images is performed based on a suitably selected intensity threshold. Through a preliminary statistical analysis performed on individual image layers, it was observed that the mean and the median of the corresponding images increase significantly due to the presence of hotspots in the thermal images when compared to the cases of absence (Figure 5). The presence of localized hotspots in the layer is due to local variation in the melt pool shape and size. The localized variation in melt pool signatures can be influenced by the localized variation in powder bed spreading or process parameters. The more hotspots there are in the images, the highest probability of drift. For example, Zenzinger *et al.* [35] demonstrated the link between the hotspots in OT images to the defect in the µCT scan of the layer. Recently, Mohr *et al.* [38] also studied the OT images to detect a defect in the final part and compared the OT images with the µCT images. It is also verified that the hotspots in the OT images link to defect in the final part.

For example, Figure 5(a), represents the layer image without any hotspots, verified with CT image as well (Figure. 5(b)) and, another image with hotspots (marked by red circles) is shown in Figure 5(c). The corresponding CT image with the porosity (marked by red circles) is also shown in Figure 5(d). On comparing the histogram of images in both

scenarios, the histogram of the image, which represents a probable drift layer (layer with hotspots is termed as drift layer), shows a right shift when compared to the image of the layer with no possible drift (layer without hotspots). The right shift in the histograms is due to the presence of hotspots, which leads to higher mean in the layer compared to no drift layer (see Figure 5(c)).

As the relative value of the mean and median of the image pixels can describe the degree of right/left shiftiness of the histogram, based on these observations, we decided on the mean and median being the features of choice for training and testing the classifier. The flowchart of the feature extraction procedure is shown in Figure 6.

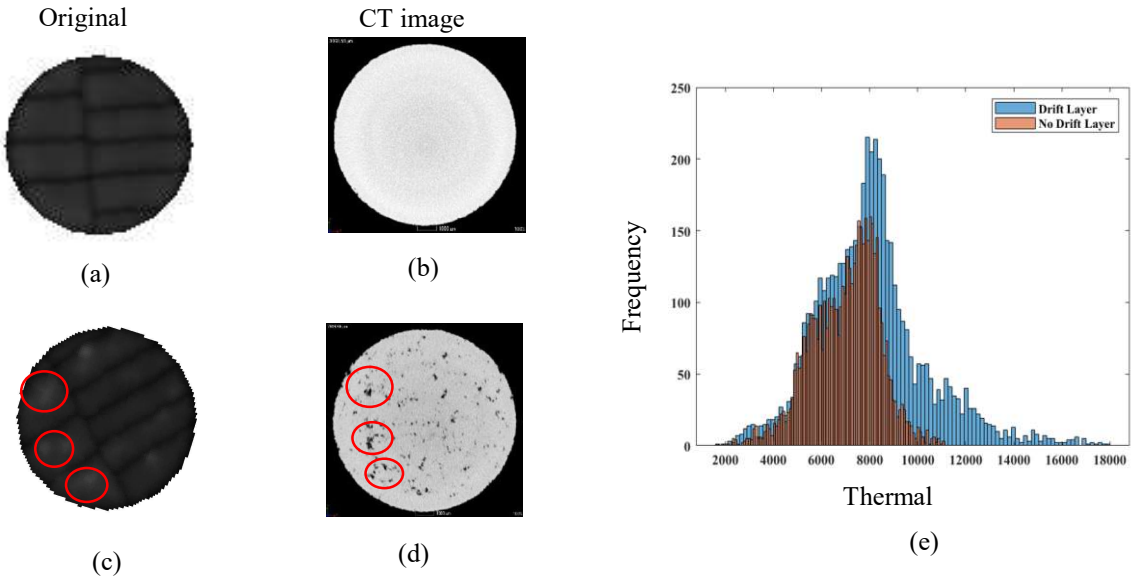


Figure 5: Example picture of (a) OT image without hotspots , (b) CT image of corresponding layer (a) without hotspots, (c) OT image with hotspots , (d) CT image of corresponding layer (c) with hotspots, and (e) histogram comparison between (a) and (c).

4.4. *Semi-supervised Learning Approach*

4.4.1. *Initial data labelling and Suitable Distance Metric Selection through K-means clustering algorithm*

Another essential aspect of ML algorithm like K-means and k-NN is the distance metric. As discussed previously in Section 2.2.3, there are many distance metrics already proposed in the literature, but for a particular application, only a few will suit. In this section, we discuss the initial labelling of sample layers as "drift" or "no-drift" and choose the suitable distance metric for our application based on the experimentation done on the *certified dataset*.

Due to the complex nature of the DMLS process, obtaining the labeled dataset in an automated fashion is a challenging task. Therefore initially, a labeled dataset of 40 layers (20 layers of each label, i.e., "drift" and "no-drift") from a set of 14 Stainless Steel CX cylindrical samples were prepared manually. The categorization of specific layers as "drift" and "no-drift" is based on the visual comparison between of particular OT image with the corresponding CT image (see Figure 5). However, the resolution difference between OT and CT images hinders a direct visual comparison of all the images. Thus, it is not possible to label and prepare a large dataset manually. Therefore, in this case the unsupervised algorithm K-means clustering come to the rescue and is used to automate the task of labelling. As seen in comparison with CT data, it can be concluded that a higher number of OT hotspots results in a higher probability of having a real defect in that layer. Based on this hypothesis, a dataset of another 200 unlabelled layers is prepared such that it comprises an approximately equal number of "drift" and "no-drift" labels. Nevertheless, it shall be noted that it does not mean that every image with an OT indication will lead to a real defect in the printed layer due to the repetitive nature of printing of the DMLS process, as the existing defect in the previous printed layer may be resolved in the next printing scan.

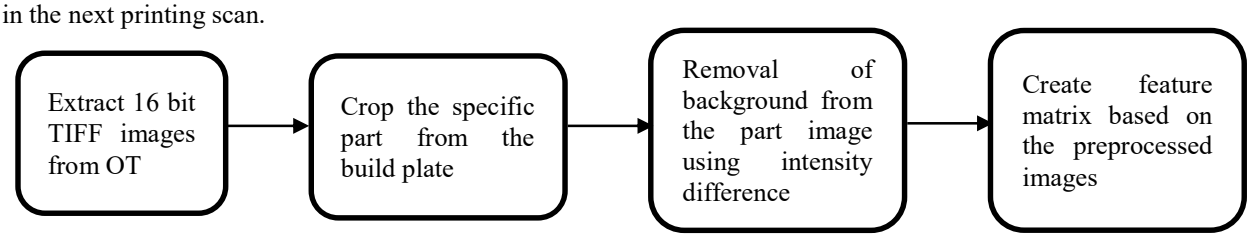


Figure 6: Flowchart for extraction of features from the OT pictures.

Next, the 40 labeled layers and 200 unlabelled layers were mixed up to perform a cluster analysis. We extract $m = 2$ features (mean and median) of each layer image, and a total of $n = 240$ feature vectors were formed, resulting in a feature matrix D of dimension 240×2. The clustering parameter $K = 2$ (as the data has to be divided into two clusters) and maximum iterations to 1000 was set. The clustering process is repeated 5 times using new initial cluster

centroid positions/seeds. The final clustering solution will be the one having the lowest sum of points to centroid distance (ref. Section 3.2.1). The accuracy of K-means algorithm cluster assignments is cross-validated with the initially labeled 40 labeled samples utilizing multiple distance metrics, and the *Correlation* distance metric shows the maximum validation accuracy. The high validation accuracy also indicates that the selection of mean and median as features are sufficient to differentiate between "drift" and "no-drift" conditions. The cluster assignment using different distance metrics on the expanded dataset is shown in Figure 7. The validation accuracies of the different distance metrics are listed in Table 3. Based on the highest percentage accuracy achieved through the K-means method, we choose the correlation distance as the suitable distance metric for our application and finalize the resultant labels for training further a supervised learning-based classifier, i.e., k-NN classifier.

Table 3: Performance of the different distance metrics on certified data.

S. No.	Distance Algorithm	Accuracy (%)
1.	Cosine	85.00
2.	Squared Euclidean	82.50
3.	City Block	85.00
4.	Correlation	100.00

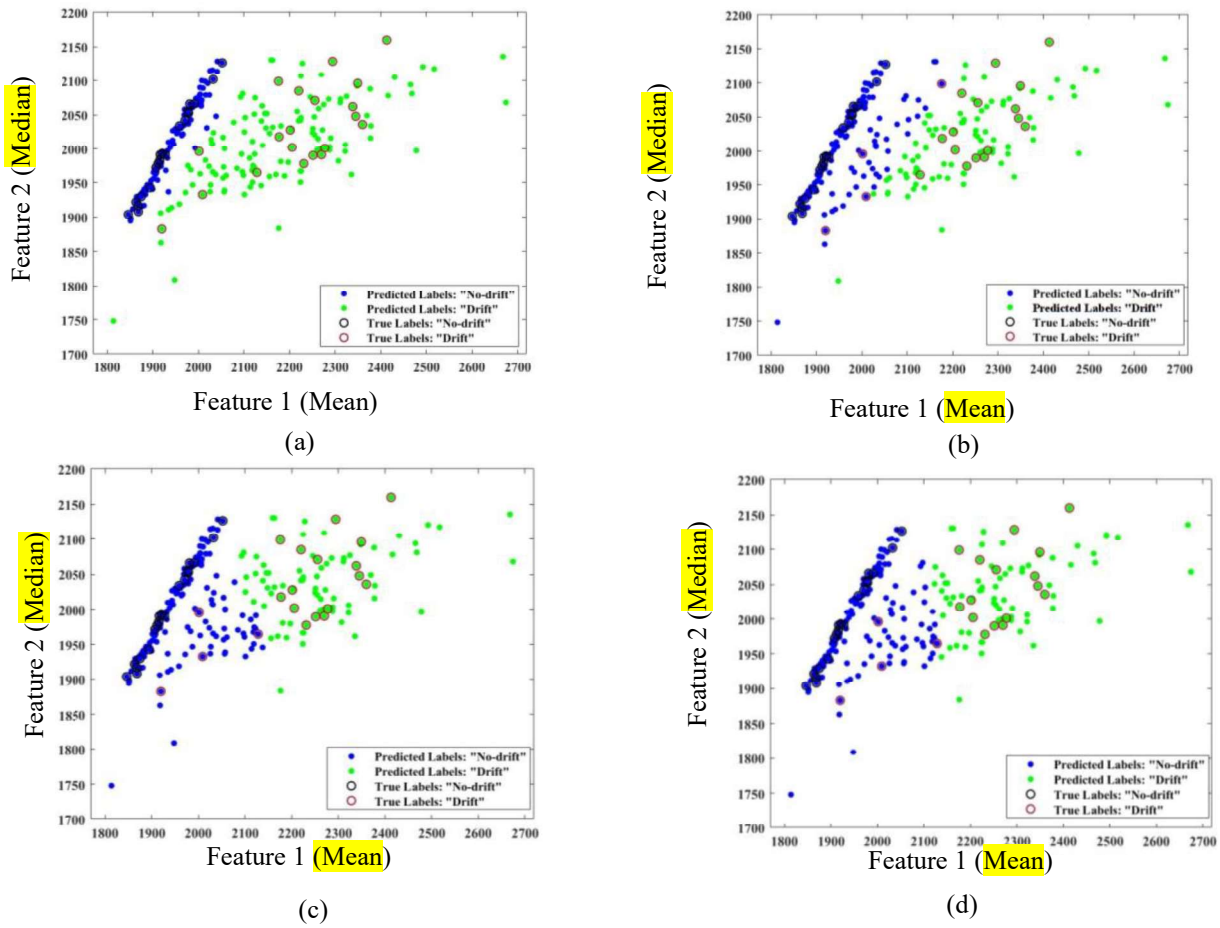


Figure 7: Clustering using (a) correlation, (b) cosine, (c) city block, and (d) squared Euclidean distance metrices.

4.4.2. *Number of nearest neighbour parameter selection for k-NN classifier*

After identifying the suitable distance metric, the next important step is to prepare a balanced training dataset (dataset with an equal number of data of each label) to train a k-NN classifier (a supervised machine learning algorithm). It is imperative to prepare a balanced dataset to avoid the biasing problem. Therefore, 100 data points for each label (total 200) were chosen randomly from the dataset of 240, which resulted from the K-means clustering. Although we prepared a balanced dataset for training, one can apply methods like "*Class confidence weighting*" in case only an imbalanced dataset is available [39-40]. The training dataset was divided randomly into a 70:30 ratio for preparing training and validation dataset. To find the best value for the number of nearest neighbors, i.e., k, the accuracy of the trained k-NN classifier with varying k was tested on the validation set, which is also shown in Figure 8. The accuracy vs k graph plot suggests that a value of $k \in (9, 23)$, can be chosen for the maximum classification accuracy. As choosing a large value of k results in the increase of computational complexity of k-NN classifier, we choose the minimum indicative value $k = 9$. The overall working pipeline of our semi-supervised learning model is depicted in Figure 9.

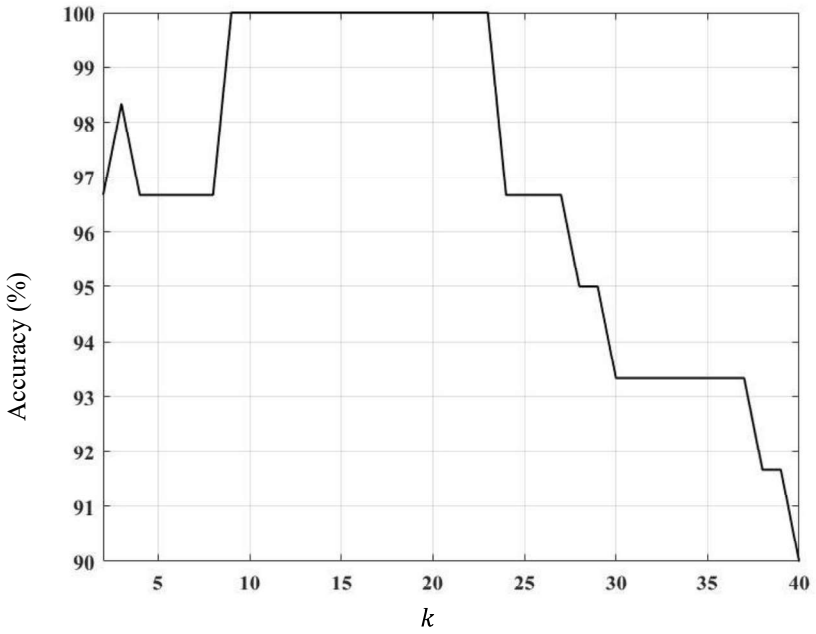


Figure 8: Accuracy versus k plot for the k-NN classifier.

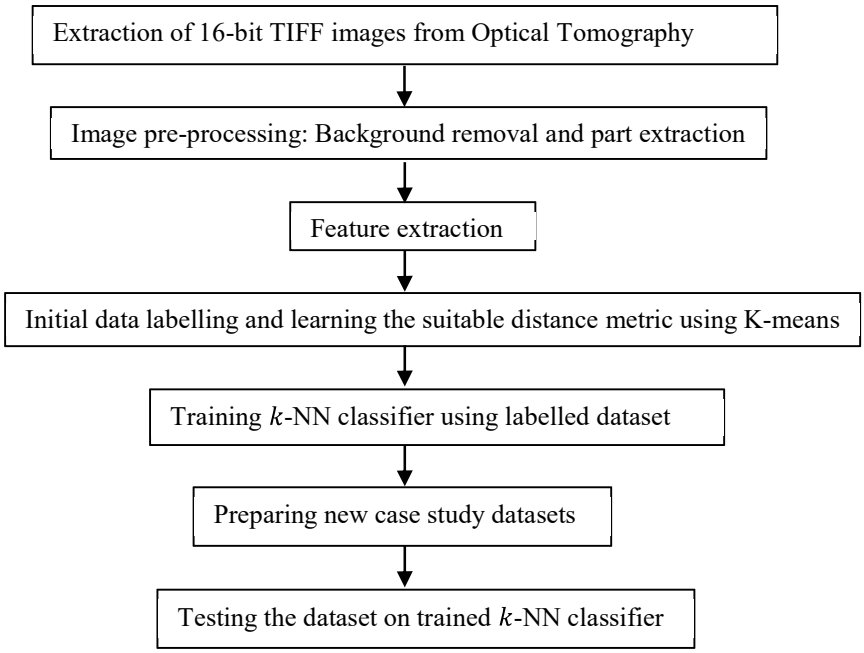


Figure 9: Overall workflow model of the proposed semi-supervised model.

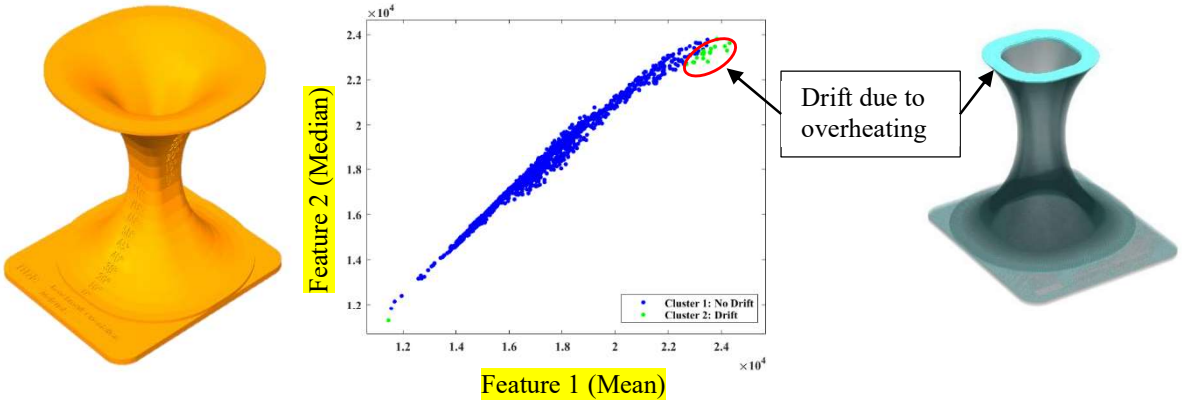
5. Results and discussions

Testing on certified data

In the experimental section, the stainless steel CX cylindrical samples were printed with different process parameters such as layer thickness, power, scanning speed, and hatch distance to generate a certified dataset for training the algorithm. The certification of the samples was done through a CT scan. A total of four certified cylindrical samples with CT porosities of 0.0033 %, 0.0054 %, 4.3967 %, and 0.1436 % were chosen for testing the trained k-NN classifier. It should be noted that the data of the four chosen cylinders were not used to train the classifier. The semi-supervised model predicted 20.82 %, 1.9543 %, 1.1363 %, and 0.4862 % layers as "drift" for the sample with CT scan porosities of 4.3967 %, 0.1436 %, 0.0054 %, and 0.0033 %, respectively. The k-NN classifier-based test results indicate that the sample with CT scan porosity of 4.3967 % has the highest number of layers with hotspots, which is due to non-optimized processing parameters. As the dataset size is limited, further comments on an analogy between porosity and drift detected (particularly at lower porosity levels) are out of the scope of this work.

5.1. *Case study: Benchmark part*

We now apply our trained *k*-NN classifier on the data obtained through real complex parts, and for that, we choose benchmark parts that are often used for material development in the industry. Out of different benchmark parts, "Overhang" is one of the critical parts which is shown in Figure 10.(a), which aims to find the critical overhang angle for a particular material that can be printed without the support structures. As we know, above a critical overhang angle, the support structures are used as an anchor to prevent the failure of the part. Therefore, it is vital to know the overhang angle for the optimization of process parameters for new material development in the DMLS process. The surety of the failure of the part at a particular overhang angle makes this part as a suitable candidate for our case study. As the location of the failure of this part is known and can be used as ground truth label. The OT images of the part exported and pre-processed as described in the previous sections and the feature matrix is prepared. As mentioned in the previous section, the *k*-NN classifier is already trained with labeled data of the certified part, and the predicted labels are shown in Figure 10.(b). The location of the layers labeled as "drift" is shown in Figure 10.(c), the poor thermal conductivity due to a large overhang angle has led to the overheating drift in the layers numbered from1535 to 1564. This hypothesis is validated by the EOSTATE Exposure analysis tool, which shows the presence of hotspots in the "drift" labeled layers. Thus, the labels predicted by the ML model are in correlation with the results from the analysis tool as well.



(a) (b) (c)

Figure 10: (a) CAD geometry, (b) Labels predicted by the *k*-NN model, and (c) printed part geometry with

5.2. *Case study: Industrial case automotive steering "Knuckle."*

For another case study, we choose an automotive part called "Knuckle." Figure 11.(a) contains the scatter plot of resultant predicted labels, and the last few layers of the part are classified as "drift" layers. This classification is in correlation with the visual inspection of the part quality as well. As can be seen in Figure 11.(b), the topmost layers from layer number 7682 until 7764 of the part failed due to overheating. The reason for the failure of these layers is due to the lack of heat dissipation. This part acts as a good example to show the need for design optimization in AM as poor support structures lead to part failure. It was also be verified by the EOSTATE Exposure analysis tool that the overheating leads to hotspots in the captured images.

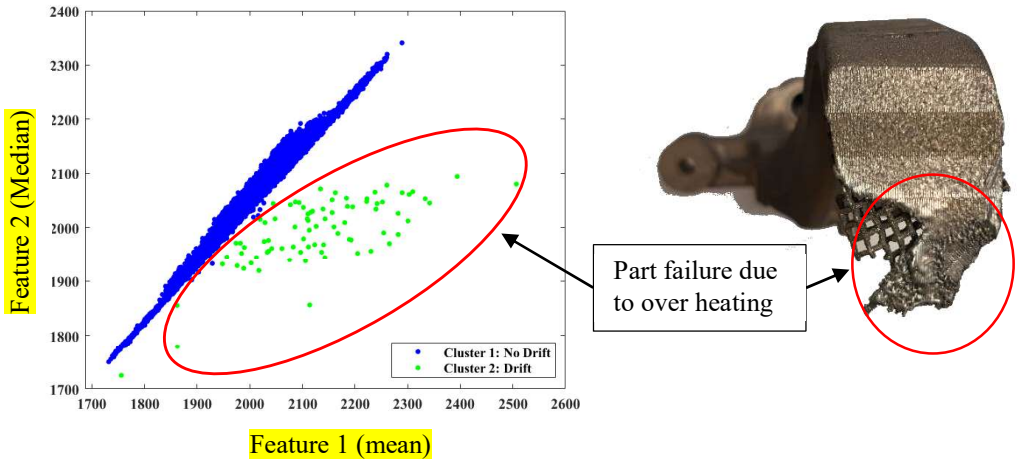


Figure 11: (a) Labels predicted by the k-NN model, and (b) printed part geometry with overheated region.

Another optimized design knuckle was printed with optimized support structures for better heat dissipation. Figure 12.(a) represents the predicted labels, and it shows that only 8 layers classified as "drift" layers, which is due to presence of hotspots in the layers numbered from 6150 to 6158. This hypothesis is cross validated with the EOSTATE Exposure analysis tool, as shown in Figure 12.(c) which shows the location of the hotspot in the layer number 6150. It shall be note that the hotspot continues to be present at the same location for next 8 layers (from layer number 6150 to 6158) which is also predicted by the proposed K-NN approach. The automotive part called "Knuckle" shown in Figure 12.(b) is printed within the framework of the European Union project titled "MAESTRO" [36].

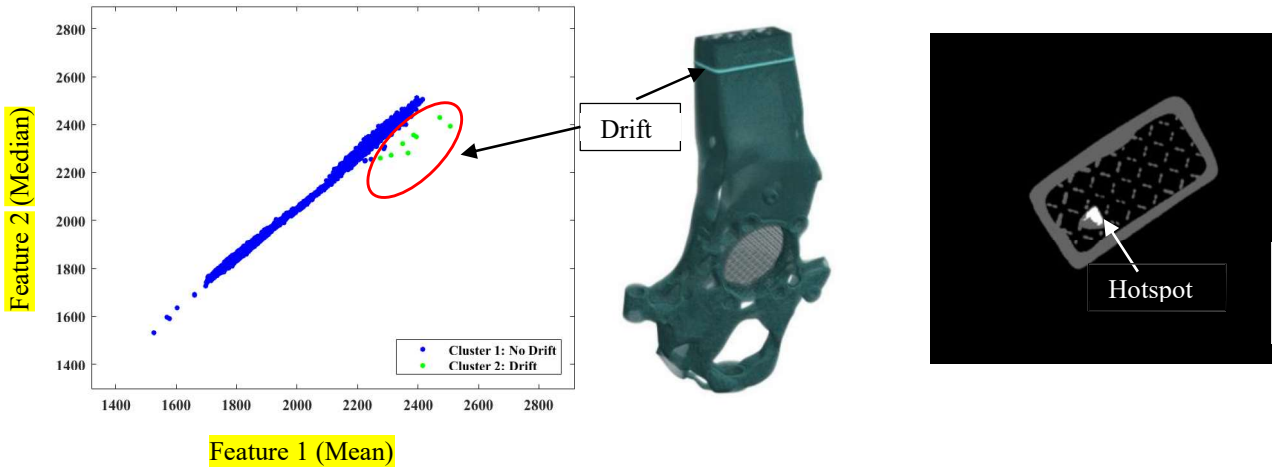


Figure 12: (a) Labels predicted by the k-NN model, (b) part geometry with drift layers (in bright color), and (c) location of the hot spot (marked in white color) in the drift layer.

The initial tests utilizing the proposed semi-supervised ML model is performed for three different parts, which were printed with different process parameters and different material (Benchmark part- Maraging Steel, Knuckle-AlSi10Mg). Therefore, it can be concluded that the proposed algorithm and feature extraction is independent of process parameters and material of the printed part.

It is also to be noted that the OT is not sensitive to all kinds of defects that could occur during the printing process. As the DMLS process is very complex in terms of printing, there are hundreds of factors that could affect the quality of the final part. For example, Galy *et al.* listed all the different parameters that can influence the quality of the Al alloy parts during printing [41]. Therefore, it is challenging to monitor all the possible factors with optical tomography, and certification for quality assurance requires expensive techniques like a CT scan. However, the CT scan is a costly technique, and part size limitation makes it more difficult to use it for every part certification. Therefore, the proposed model based on ML can be used to select good parts for the further expensive post-processing techniques.

6. Conclusion

In-situ monitoring of the DMLS process can significantly improve the reliability of the whole process for quality assurance of the product. But there are few issues with the monitoring systems which need to be solved for fast and easy anomaly detection. Firstly, processing the enormous amount of unlabelled data obtained from these monitoring

systems is a huge challenge. Secondly, it is very laborious to detect the drift in the final part of the in-situ monitored data. Therefore, the use of ML algorithms for the treatment of in-situ monitoring data has merit. This paper is a continuation of the research going into the field of data treatment in additive manufacturing.

The key contributions of this work are summarized as follows:

- An unsupervised K-means clustering algorithm was used to label the unlabelled data with the help of small set of labelled data and helps in choosing the most suitable distance matric. The accuracy of the algorithm was verified using the computer tomography-based certified data.
- The labeled dataset was utilized to train the k-NN classifier. The optimum value of number of nearest neighbors k was judged on the basis of accuracy vs k plot.
- The semi-supervised approach successfully classifies the layers into either "drift" and "no-drift" for presented case studies and the cross-validated through the EOSTATE Exposure OT analysis tool.

The future aim of this work is to study the possibility to detect the exact location of the drift in the specific layers and to check the reliability of this model for other complex parts/materials as well. The coupling with data from other sensors, such as photodiodes, must to be studied. The semi-supervised machine learning shows that machine learning in additive manufacturing can be a robust method to improve the post processability of the in-situ data. The final aim of this approach could be achieved by real-time monitoring of the whole process in a closed feedback control manner: if drift in the process is detected, different printing parameters could be used to avoid part failure.

Acknowledgments

This work was conducted as part of the "ENABLE" project funded by the European Union's Marie Sklodowska-Curie Actions (MSCA) Innovative Training Networks (ITN) H2020-MSCA-ITN-2017 under the grant agreement Number 764979.

References

1. Astm, I. ASTM52900-15 Standard Terminology for Additive Manufacturing—General Principles—Terminology. ASTM International, West Conshohocken, PA 2015, 3, 5.

2. Taminger, K.M.; Hafley, R.A. Electron beam freeform fabrication for cost effective near-net shape manufacturing. 2006.

3. Bourell, D.L.; Leu, M.C.; Rosen, D.W. Roadmap for additive manufacturing: identifying the future of freeform processing. The University of Texas at Austin, Austin, TX 2009, 11-15.

4. Gibson, I.; Rosen, D.W.; Stucker, B. Additive manufacturing technologies; Springer: 2014; Vol. 17.

5. Mani, M.; Feng, S.; Lane, B.; Donmez, A.; Moylan, S.; Fesperman, R. Measurement science needs for real-time control of additive manufacturing powder bed fusion processes. 2015.

6. Tapia, G.; Elwany, A. A review on process monitoring and control in metal-based additive manufacturing. Journal of Manufacturing Science and Engineering 2014, 136.

7. Renken, V.; Albinger, S.; Goch, G.; Neef, A.; Emmelmann, C. Development of an adaptive, self-learning control concept for an additive manufacturing process. CIRP Journal of Manufacturing Science and Technology 2017, 19, 57-61.

8. Purtonen, T.; Kalliosaari, A.; Salminen, A. Monitoring and adaptive control of laser processes. Physics Procedia 2014, 56, 1218-1231.

9. Grasso, M.; Laguzza, V.; Semeraro, Q.; Colosimo, B.M. In-process monitoring of selective laser melting: spatial detection of defects via image data analysis. Journal of Manufacturing Science and Engineering 2017, 139.

10. Grasso, M.; Demir, A.; Previtali, B.; Colosimo, B. In situ monitoring of selective laser melting of zinc powder via infrared imaging of the process plume. Robotics and Computer-Integrated Manufacturing 2018, 49, 229-239.

11. Repossini, G.; Laguzza, V.; Grasso, M.; Colosimo, B.M. On the use of spatter signature for in-situ monitoring of Laser Powder Bed Fusion. Additive Manufacturing 2017, 16, 35-48.

12. Khanzadeh, M.; Chowdhury, S.; Marufuzzaman, M.; Tschopp, M.A.; Bian, L. Porosity prediction: Supervised-learning of thermal history for direct laser deposition. Journal of manufacturing systems 2018, 47, 69-82.

13. Okaro, I.A.; Jayasinghe, S.; Sutcliffe, C.; Black, K.; Paoletti, P.; Green, P.L. Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. Additive Manufacturing 2019, 27, 42-53.

14. Scime, L.; Beuth, J. Anomaly detection and classification in a laser powder bed additive manufacturing process using a trained computer vision algorithm. Additive Manufacturing 2018, 19, 114-126.

15. Zhu, X.; Ghahramani, Z.; Lafferty, J.D. Semi-supervised learning using gaussian fields and harmonic functions. In Proceedings of Proceedings of the 20th International conference on Machine learning (ICML-03); pp.

912-919.

16. Lieber, D.; Stolpe, M.; Konrad, B.; Deuse, J.; Morik, K. Quality prediction in interlinked manufacturing processes based on supervised & unsupervised machine learning. Procedia Cirp 2013, 7, 193-198.

17. Zhu, X.; Goldberg, A.B. Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning 2009, 3, 1-130.

18. Aamir, M.; Zaidi, S.M.A. Clustering based semi-supervised machine learning for DDoS attack classification. Journal of King Saud University-Computer and Information Sciences 2019.

19. Peikari, M.; Salama, S.; Nofech-Mozes, S.; Martel, A.L. A cluster-then-label semi-supervised learning approach for pathology image classification. Scientific reports 2018, 8, 1-13.

20. Gu, Y.; Wang, Y.; Yang, Z.; Xiong, F.; Gao, Y. Multiple-features-based semisupervised clustering DDoS detection method. Mathematical Problems in Engineering 2017, 2017.

21. "Committee F42 on Additive Manufacturing Technologies.". Available: https://www.astm.org/COMMITTEE/F42.htm. [Accessed: 31-Jan-2020].

22. Ye, D.; Hong, G.S.; Zhang, Y.; Zhu, K.; Fuh, J.Y.H. Defect detection in selective laser melting technology by acoustic signals with deep belief networks. The International Journal of Advanced Manufacturing Technology 2018, 96, 2791-2801.

23. Grasso, M.; Colosimo, B.M. Process defects and in situ monitoring methods in metal powder bed fusion: a review. Measurement Science and Technology 2017, 28, 044005.

24. Pavlov, M.; Doubenskaia, M.; Smurov, I. Pyrometric analysis of thermal processes in SLM technology. Physics Procedia 2010, 5, 523-531.

25. Berumen, S.; Bechmann, F.; Lindner, S.; Kruth, J.-P.; Craeghs, T. Quality control of laser-and powder bed-based Additive Manufacturing (AM) technologies. Physics procedia 2010, 5, 617-622.

26. Kleszczynski, S.; zur Jacobsmühlen, J.; Reinarz, B.; Sehrt, J.T.; Witt, G.; Merhof, D. Improving process stability of laser beam melting systems. In Proceedings of Fraunhofer Direct Digital Manufacturing Conference.

27. Kleszczynski, S.; Zur Jacobsmühlen, J.; Sehrt, J.; Witt, G. Error detection in laser beam melting systems by high resolution imaging. In Proceedings of Proceedings of the twenty third annual international solid freeform fabrication symposium.

28. Zhang, B.; Ziegert, J.; Farahi, F.; Davies, A. In situ surface topography of laser powder bed fusion using

fringe projection. Additive Manufacturing 2016, 12, 100-107.

29. Li, Z.; Liu, X.; Wen, S.; He, P.; Zhong, K.; Wei, Q.; Shi, Y.; Liu, S. In situ 3D monitoring of geometric signatures in the powder-bed-fusion additive manufacturing process via vision sensing methods. Sensors 2018, 18, 1180.

30. Caltanissetta, F.; Grasso, M.; Petro, S.; Colosimo, B.M. Characterization of in-situ measurements based on layerwise imaging in laser powder bed fusion. Additive Manufacturing 2018, 24, 183-199.

31. Lloyd, S. Least squares quantization in PCM. IEEE transactions on information theory 1982, 28, 129-137.

32. Taheri, H.; Koester, L.W.; Bigelow, T.A.; Faierson, E.J.; Bond, L.J. In situ additive manufacturing process monitoring with an acoustic technique: clustering performance evaluation using K-means algorithm. Journal of Manufacturing Science and Engineering 2019, 141.

33. Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding; Stanford: 2006.

34. Abu Alfeilat, H.A.; Hassanat, A.B.; Lasassmeh, O.; Tarawneh, A.S.; Alhasanat, M.B.; Eyal Salman, H.S.; Prasath, V.S. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. Big data 2019, 7, 221-248.

35. Zenzinger, G.; Bamberg, J.; Ladewig, A.; Hess, T.; Henkel, B.; Satzger, W. Process monitoring of additive manufacturing by using optical tomography. In Proceedings of AIP Conference Proceedings; pp. 164-170.

36. "Maestro project – About the project Horizon 2020." [Online]. Available: https://www.maestro-project.eu/projects/about-the-project/. [Accessed: 27-Jan-2020].

37. Clijsters, S.; Craeghs, T.; Buls, S.; Kempen, K.; Kruth, J.-P. In situ quality control of the selective laser melting process using a high-speed, real-time melt pool monitoring system. The International Journal of Advanced Manufacturing Technology 2014, 75, 1089-1101.

38. Mohr, G.; Altenburg, S.J.; Ulbricht, A.; Heinrich, P.; Baum, D.; Maierhofer, C.; Hilgenberg, K. In-Situ Defect Detection in Laser Powder Bed Fusion by Using Thermography and Optical Tomography—Comparison to Computed Tomography. Metals 2020, 10, 103.

39. Dubey, H.; Pudi, V. Class based weighted k-nearest neighbor over imbalance dataset. In Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining; pp. 305-316.

40. Liu, W.; Chawla, S. Class confidence weighted knn algorithms for imbalanced data sets. In Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining; pp. 345-356.

41. Galy, C.; Le Guen, E.; Lacoste, E.; Arvieu, C. Main defects observed in aluminum alloy parts produced by SLM: From causes to consequences. Additive Manufacturing 2018, 22, 165-175.