



# Uncovering acoustic signatures of pore formation in laser powder bed fusion

Joshua R. Tempelman<sup>1,2</sup> · Maruti K. Mudunuru<sup>3</sup> · Satish Karra<sup>4</sup> · Adam J. Wachtor<sup>1</sup> · Bulbul Ahmmed<sup>5</sup> · Eric B. Flynn<sup>6</sup> · Jean-Baptiste Forien<sup>7</sup> · Gabe M. Guss<sup>8</sup> · Nicholas P. Calta<sup>7</sup> · Phillip J. DePond<sup>7</sup> · Manyalibo J. Matthews<sup>7</sup>

Received: 11 August 2023 / Accepted: 24 November 2023 / Published online: 28 December 2023  
This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023

## Abstract

We present a machine learning workflow to discover signatures in acoustic measurements that can be utilized to create a low-dimensional model to accurately predict the location of keyhole pores formed during additive manufacturing processes. Acoustic measurements were sampled at 100 kHz during single-layer laser powder bed fusion (LPBF) experiments, and spatio-temporal registration of pore locations was obtained from post-build radiography. Power spectral density (PSD) estimates of the acoustic data were then decomposed using non-negative matrix factorization with custom  $k$ -means clustering (NMF $k$ ) to learn the underlying spectral patterns associated with pore formation. NMF $k$  returned a library of basis signals and matching coefficients to *blindly* construct a feature space based on the PSD estimates in an *optimized* fashion. Moreover, the NMF $k$  decomposition led to the development of computationally inexpensive machine learning models which are capable of quickly and accurately identifying pore formation with classification accuracy of supervised and unsupervised label learning greater than 95% and 90%, respectively. The intrinsic data compression of NMF $k$ , the relatively light computational cost of the machine learning workflow, and the high classification accuracy makes the proposed workflow an attractive candidate for edge computing toward in-situ keyhole pore prediction in LPBF.

**Keywords** Additive manufacturing · Laser powder bed fusion · Acoustic monitoring · Unsupervised learning · Non-negative matrix factorization · Dimensionality reduction

## 1 Introduction

Additive manufacturing (AM) technology has been successfully utilized to rapidly fabricate both prototypes and functional production parts in applications throughout the aerospace, biomedical, and tooling industries [1–6]. Advances in these application areas require parts with geometric complexity that can be impossible to produce using traditional manufacturing methods [7–12]. For this reason, AM processes such as laser powder bed fusion (LPBF), have drawn great interest in a wide variety of industries. However, despite the advantages offered by this and other AM technologies, process variability and lack of quality assurance

standards continue to challenge qualification and certification efforts for low-volume AM part production in safety- or performance-critical applications. Defects, such as porosity, spatter, balling, geometric deviations, microstructural heterogeneities, and impurities, are formed during the build process and implicitly dependent on the thermal history of the build, process parameter settings, and the stochastic nature of the LPBF processes themselves [6, 13, 14]. Such defects can vary in quantity and location, resulting in inconsistencies in the mechanical performance of parts produced through LPBF. Specifically, pores formed within the bulk material may lead to stress concentrations within the part which can reduce the mechanical integrity, specifically with respect to ductility and fatigue life, of produced parts when put into service [15, 16].

Current quality control inspections for LPBF-based AM parts are primarily performed post-production, resulting in wasted resources should critical defects have instead been detected during fabrication [17]. However, LPBF processes are highly suited for in-situ monitoring schemes due to the

✉ Joshua R. Tempelman  
tempelman@lanl.gov

✉ Maruti K. Mudunuru  
maruti@pnnl.gov

Extended author information available on the last page of the article

abundance of optical, thermal, and acoustic information generated during the build processes [18]. Accordingly, the development of in-situ measurement techniques for LPBF has seen rapid advancements over the past several years [19, 20]. However, many of the developed sensing methods require specialized and costly imaging technologies to accurately measure the thermal and optical emissions. Moreover, these imaging methods pose a major hurdle with respect to acquired data. That is, the data collected during imaging of an AM specimen and/or process is of huge volume (e.g., terabyte scale), costly, time-consuming, missing information, ill-structured (e.g., misalignment among different melt pools), and has a low signal-to-noise ratio. Hence, there is a need for a fast, reliable, and cost-effective in-situ quality monitoring technology and a real-time algorithm to process the acquired data near the sensor edge. Such an algorithm could enable potential corrective actions during the fabrication to reduce the overall AM cost [21]. To this end, acoustic emission monitoring combined with recent advances in machine learning (ML) shows promise to achieve this grand challenge [22, 23].

Acoustic measurements have proven effective for identifying process abnormalities in a variety of lasing operations [24–27]. Recently, acoustic monitoring has been extended to the domain of LPBF [28–32]. Typically, one utilizes signal processing tools to *featurize* the acoustic recordings and subsequently uses pattern recognition tools such as neural networks [28, 33–35], clustering [36], or statistical classifiers [29] to interpret the underlying LPBF process. Spectral representations of the acoustic data have been demonstrated to be useful in predicting both LPBF processing parameters over full-builds [28, 33, 35, 36] and localized keyhole pores of spatio-temporally registered data [29]. The physical connection between acoustic spectral features and pore formation is not fully understood, although it has been computationally demonstrated that surface oscillations of the melt pool precedes pore formation in the vicinity of 20–50 kHz [37]. To this end, Ren et al. developed a combined computational and experimental approach to accurately register and predict pores based on thermal oscillations near the range of 40 kHz which were shown to be induced by melt pool instabilities at the onset of pore formation [38].

Despite the promising recent research involving acoustic monitoring, there are several notable challenges that limit acoustic measurements from being used to their fullest potential in practical in-situ applications. Namely, while several studies have used acoustic information for supervised learning, the automated identification of the *most informative* spectral patterns for detecting the formation of keyhole pores has yet to be achieved. The intelligent identification of these spectral patterns could have meaningful implications to relate the acoustic measurements to melt pool physics. Moreover, there is the challenge of data storage for practical in-situ

monitoring schemes. For shorter duration builds (e.g., couple of hours), the acquired data from of AM process monitoring is not overwhelming in size and a clear advantage in real-time processing practicality is achieved as compared to X-ray or high-speed camera imaging methods. However, for longer duration builds (e.g., days), the acoustic emission data acquired with sampling frequencies of  $O(10^5 - 10^6)$  Hz is large even for 16-bit data (e.g., terabyte scale) and is comparable to the above state-of-the-art imaging techniques. Typically, acoustic signals can be collected and processed in near real-time using a reasonably lower-cost computing hardware such as a Raspberry Pi or FPGA [39] (provided the sampling rate does not exceed the microprocessors abilities). To this end, developing an efficient and effective means of compressing the full acoustic signals into actionable information at sensor edges (in a smart and automated fashion) is of great interest for developing practical in-situ technologies.

We aim to address these challenges directly by employing the experimental framework to register acoustic signals to pore locations presented in [40]. In prior works, this data was featurized into massive feature banks which were later reduced in a *fully supervised* feature selection and classification framework [29, 30]. In the current paper, we present a ML model that leverages the same spatio-temporal registration scheme of [40] with the goal to *discover* hidden signals in the acoustic data, *assess* the dominant signal response and associated frequencies for pore formation, and then *classify* the state of the LPBF process (pore or non-pore formed at a given location and time) with an *unsupervised* featurization routine. Hence, the underlying patterns in the acoustic spectra can be *automatically* determined rather than manually identified based on user-defined feature spaces. Moreover, our proposed method provides a compressed representation of the data, greatly alleviating data storage concerns in LPBF monitoring.

The innovative aspect of our ML model is that it combines recent advances in unsupervised and supervised learning to quickly and accurately predict pore formation with minimal training data. To achieve the discovery of hidden signals or patterns in the data, we make use of an unsupervised matrix factorization method called NMF $k$ . NMF $k$  is based on non-negative matrix factorization (NMF) [41] coupled with a semi-supervised  $k$ -means clustering algorithm [42, 43]. NMF uncovers latent representations of the data while  $k$ -means clustering partitions the latent space to discover underlying patterns [44]. In contrast with traditional NMF [41, 45], NMF $k$  allows for automatic identification of the optimal number of signatures present in the data [42, 43]. Hence, the feature space recovered for the acoustic signals is not only recovered blindly but it is done so in an intelligent and optimal manner for the sake of identifying binary differences in the underlying operating state (e.g., pore forming or non-pore forming). Moreover, NMF $k$  produces a low-

dimensional feature space representation of the frequency content contained in the acoustic measurements allowing the data to be compressed with minimal loss of information (e.g., storing only decomposed matrices), perform blind source separation and gain insights into the melt pool dynamics (e.g., dominant frequencies that characterize pore formation).

Accordingly, the rest of the paper is organized as follows: Section 2 provides the details of the experimental data sampled for ML model development. This data is acquired during an LPBF-based AM experiment, and the ground truth to label the acoustic emission signals is also described in this section. Section 3 proposes the ML workflow to develop the proposed NMFk-based ML model. Details on the procedure to extract dominant pore formation signatures, supervised and unsupervised data labeling strategies, and pore-state classification are also described. Section 4 provides details on the classification accuracy of the trained ML model in the identification of pores and offers a brief discussion regarding classifier selection and computational cost. Finally, Section 5 discusses the conclusions and the potential directions for future ML research for in-situ acoustic monitoring for AM processes.

## 2 Experimental measurements

Keyhole pore formation is a common defect type in LPBF and is the focus of this study. We outline the basic mechanisms of LPBF and keyhole pore formation here and direct readers to Reference [46] for a detailed discussion. A melt pool is formed as the laser melts the metal powder that was spread on a solid metal substrate. Depending on the operating conditions, strong vaporization can occur due to laser heating which drives a vapor recoil pressure and vapor depression in front of the melt pool. As the depression collapses, vapor may become trapped within the build material as the material rapidly solidifies; this results in the formation of keyhole pores, which are most commonly found at turn points where the laser changes speed.

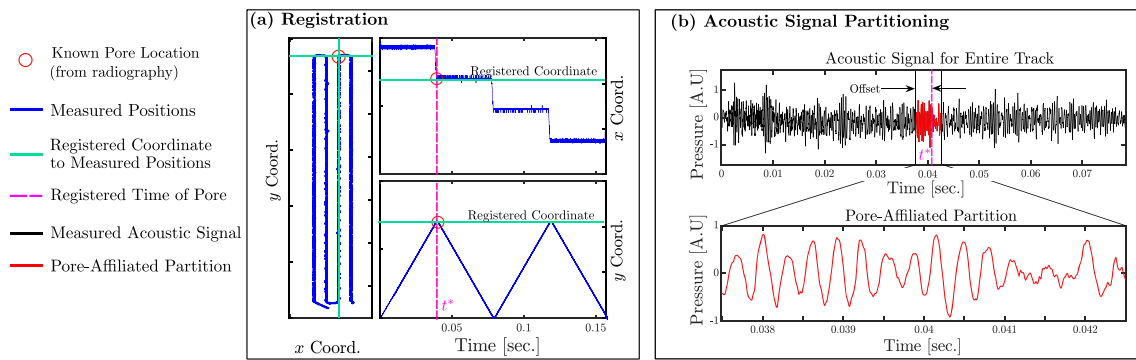
To study the formation of pores, single-layer experiments were conducted using an Aconity3D (Aachen, Germany) open-source platform at Lawrence Livermore National Laboratory using the same optical set up described in References [40, 47]. A 1070-nm continuous wave (CW), 400 W Yb-fiber laser was used in the LPBF system producing an approximate beam diameter of  $100\ \mu\text{m}$  ( $D4\sigma$ ) at its focal point. All experiments were conducted with 316L stainless steel (SS) powder—produced using gas atomization by Additive Metal Alloys (Holland, OH, USA), and sorted to contain particle sized in the range of 15 to  $45\ \mu\text{m}$ —atop 316L SS substrates. The scans were performed in an argon atmosphere with a continuous gas recirculation to redirect the vaporization plume. Single layer patches were melted with dimensions  $2\ \text{mm} \times$

$5\ \text{mm}$  and  $1\ \text{mm} \times 5\ \text{mm}$ . In some instances, the laser powder was modulated near laser turn points to reduce the likelihood of pore formation. Five experimental trials were conducted to provide 5 data sets with the same operating parameters detailed in [29]. The laser power was varied between 150 and 375 W and laser speed between 100 and 400 mm/s respectively with a  $100\ \mu\text{m}$  hatch spacing. The laser power and speeds were varied randomly throughout the plate with a majority corresponding to high volumetric energy densities (VEDs) which are known to cause keyhole pore formation.

Acoustic data was collected by the same procedure described in [29]. An electric microphone was secured to the side of the build chamber above the build surface, approximately 25 cm from the center of the build plate. The microphone was sampled at 100 kHz, coupled with a low pass filter of 6 dB, and collected at a gain of  $10\times$  using a Stanford Research System pre-amplifier. The data was recorded in-sync with measurements of the scan mirror positions, allowing for spatio-temporal registration of the acoustic signal with the location of the laser on the build plate. Ex-situ X-ray imaging of the build plates with the single-layer depositions revealed the pore locations providing ground truth labels for the acoustic data. The X-ray data was acquired at beamline 8.3.2 of the Advanced Light Source housed at Lawrence Berkeley National Laboratory with parallel polychromatic X-ray beams and a 0.5-mm thick copper plate for removal of low-energy X-rays. The transmission images were collected at a resolution of  $2560 \times 2160$  pixels and a 1.5 s exposure time. A PCO Edge sCMOS camera with a  $5\times$  lens was spaced approximately 100 mm away from the plates resulting in an average effective pixel size of approximately  $1.3\ \mu\text{m}$ .

The acoustic measurements were partitioned into windows of 10 ms in duration and labeled accordingly as pore or non-pore classes (Fig. 1). The labeling was based on the spatio-temporal registration algorithm described in Reference [40]. For segments that include a pore, a random offset was applied so that the pores appeared at random locations in the window partitions to ensure that the ML model developed can be generalized. The remainder of the track was associated with non-pores and the data was partitioned into overlapping windows of 10 ms with up to a 0.25-ms overlap. This resulted in a set of pore- and non-pore-affiliated acoustic data, which can be used to train and test our ML models. The data partitioning scheme provided an ensemble of time-series signals stored in a matrix  $\mathbf{X}$  for discovering pore formation signatures in the acoustic data for each experiment which are summarized in Table 1.

The recirculation mechanism for the cover gas in the LPBF machine operates in the 0–5 kHz frequency range which dominates the acoustic power spectra for both pore and non-pore signals and is present even when the laser is off. Therefore, this frequency range may be regarded as



**Fig. 1** The spatial-temporal registration scheme: **a** Pore location(s) found in radiography are registered to the measured laser locations on a  $x$ - $y$  coordinate grid. The time(s) when the measured  $x$  and  $y$  coordi-

nates coincide with a registered pore are marked by  $t^*$  as the registration time(s). **b** The partitioning of an acoustic signal into a pore-affiliated segment based on the registered pore time and random offset

process-irrelevant noise, and frequencies below 2.5 kHz were neglected.

### 3 Acoustic data decomposition and ML workflow

This section presents the proposed pore detection framework. Figure 2 shows the overall ML workflow to discover hidden signatures in acoustic data and then predict the formation of pores during the AM process. The workflow is divided into two main steps. In the first step, unsupervised learning based on NMF $k$  is used for extracting hidden signatures/features and dimensionality reduction of the acoustic data. The data collected during each LPBF experiment was assembled into a data matrix  $\mathbf{X}$  whereby each column corresponds to an acoustic time-series window labeled as either containing a pore or

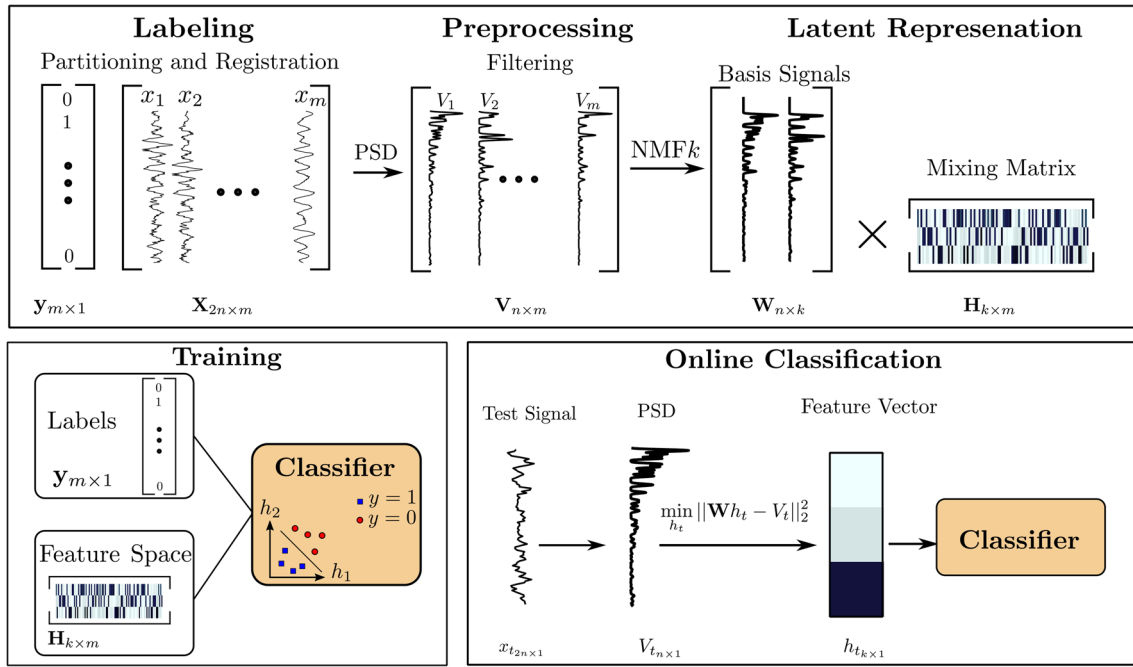
not, as described in Sect. 2. The time-series matrix  $\mathbf{X}$  was converted into a PSD matrix  $\mathbf{V}$  using Welch's method [48]. Next,  $\mathbf{V}$  was decomposed into two non-negative matrices  $\mathbf{W}$  and  $\mathbf{H}$ .  $\mathbf{W}$  represents a dictionary of spectral patterns or hidden signals in the acoustic data. In other words,  $\mathbf{W}$  forms as a basis to reconstruct  $\mathbf{V}$  based on the linear combination of basis vectors weighted by the coefficients  $\mathbf{H}$  (which is commonly referred to as the mixing matrix). The matrix  $\mathbf{H}$  provides information on how the discovered signals are mixed, which is useful for test samples in the second step of the ML workflow. The values in  $\mathbf{H}$  also provide information on the individual importance of each discovered signal to reconstruct the acoustic training data.

An obvious concern is whether or not a linear decomposition (NMF $k$ ) sufficiently captures an intrinsically nonlinear dataset (since the PSD is a nonlinear operator). However, despite the nonlinearity introduced by the PSD, NMF $k$  still works for this case study since the basis vectors characterizing pores and their non-pore counterparts are well separated in the encoded latent space. Moreover, if the measured signals are assumed to be a linear combination of sufficiently separable sources, then their PSDs may be reconstructed as a linear combination of source PSD signals [49]. Hence, the NMF $k$  decomposition can sufficiently capture the  $\mathbf{V}$  matrix.

Once a latent space is learned, a test acoustic signal  $x_{t_{2n \times 1}}$ , can be used to classify its PSD vector  $V_{t_{n \times 1}}$ . Namely, the dictionary of spectral patterns  $\mathbf{W}$  (generated from the training data) provides a basis to generate the coefficients vector  $h_{t_{k \times 1}}$  to reconstruct  $V_{t_{n \times 1}}$  using the least-squares method. The coefficient vector  $h_{t_{k \times 1}}$  can be treated as a feature vector to classify the test data sample as pore or non-pore based on the patterns of the  $\mathbf{H}$  matrix learned from the training data. This allows us to store important features in the high-throughput data and their coefficients, thus leading to dimensionality reduction and data compression. Classification can be performed by either unsupervised learning (e.g.,  $k$ -means clustering) or

**Table 1** Details of training and testing sets separated by experimental trial

		Training data			
Set	Samples	Pores	No pore	Ratio	
1	370	73	297	4.1	
2	222	42	180	4.3	
3	703	249	454	1.8	
4	415	98	317	3.2	
5	806	360	446	1.2	
		Testing data			
Set	Samples	Pores	No pore	Ratio	
1	2146	749	1397	1.9	
2	2295	780	1514	1.9	
3	1813	573	1240	2.0	
4	2101	724	1377	1.9	
5	1720	462	1258	2.7	



**Fig. 2** ML workflow: schematic representation of our ML workflow to discover hidden signatures in acoustic data and predict pore formation during the AM process. The PSD of the signal windows is stored in matrix  $\mathbf{V}$ , which is decomposed into a basis  $\mathbf{W}$  and coefficient matrix

$\mathbf{H}$ . The label vector  $\mathbf{y}$  and coefficient matrix are used to train a classifier. The coefficients for a test signal,  $h_t$ , are computed with the PSD of the test signal and the basis  $\mathbf{W}$ , which can then be used for classification

using a supervised learning method (e.g., decision trees, random forests, discriminant analysis, support vector machines).

### 3.1 Non-negative matrix factorization with custom $k$ -means clustering (NMF $k$ )

We have not yet addressed the dimensionality of  $\mathbf{W}$  which is typically a user-defined input for NMF decomposition. Given an  $(n, m)$  matrix  $\mathbf{V}$ , a  $k$ -dimensional basis will factorize  $\mathbf{V}$  into matrices  $\mathbf{W}$  and  $\mathbf{H}$  of sizes  $(n \times k)$  and  $(k \times m)$ , respectively as  $\mathbf{V} = \mathbf{W} \times \mathbf{H}$ . The variable  $k$  corresponds to an unknown number of dominant basis signals present in the data. To this end, NMF $k$  estimates the optimal number of hidden signals  $k_{\text{opt}}$  by performing a series of NMF operations for different values of  $k$ ;  $k = 1, 2, 3, \dots, d$ , where the maximum value  $d$  cannot be expected to exceed  $n$  or  $m$ . This is achieved by minimizing the following objective function,  $\mathcal{L}$ , based on the Frobenius norm for all possible values of  $k$ .

$$\mathcal{L} = \|\mathbf{V} - \mathbf{W} \times \mathbf{H}\|_F \tag{1}$$

$\mathbf{W}, \mathbf{H} \geq 0 \quad \forall n, k, m$

For each  $k$  in  $1 \leq k \leq d$ , NMF is performed for 10 random initializations of  $\mathbf{W}$  and  $\mathbf{H}$ , respectively. The converged value for  $\mathcal{L}$  is returned for each of the 10 iterations at a fixed dimension  $k$ . The smallest value of  $\mathcal{L}$  over the 10 iterations

is considered the reconstruction error for the  $k$ -dimensional factorization  $\epsilon(k)$ . The resulting  $(k \times m)$ -sized  $\mathbf{H}$  matrices are clustered using customized  $k$ -means clustering. During clustering, we enforce the condition that each of the  $k$  clusters contain an equal number of members, which is equal to the number of performed multiple random runs (e.g., 10 solutions). Next, the average Silhouette width  $S(k)$  and Akaike Information Criterion (AIC) score is computed.  $S(k)$  quantifies how similar an object is to its own cluster compared to other clusters; a low  $S$  score indicates redundancy in the latent space whereas high values indicate the clusters are well separated. The combination of  $\epsilon(k)$ ,  $S(k)$ , and the AIC score determines the optimal number of hidden signals,  $k_{\text{opt}}$  [50]. Typically,  $S(k)$  declines sharply after an optimal number,  $k_{\text{opt}}$ , is reached. If  $k$  is low, the Silhouette width will be high, but so may be  $\epsilon(k)$  because of under-fitting, whereas a high  $k$  may lead to redundant basis signals. So, the best estimate for  $k$  is a number that optimizes both  $\epsilon(k)$  and the Silhouette width,  $S(k)$ . The resulting  $\mathbf{W}$  at  $k_{\text{opt}}$  provides information on the hidden signals in the acoustic PSDs.

### 3.2 NMF $k$ -based pore detection ML model

The coefficient vectors  $h_{t \times 1}$  extracted from a test sample  $x_{t \times 1}$  were compared against the library of coefficients in  $\mathbf{H}$  by using ML classifiers that were trained, tested, and val-

idated using the  $\mathbf{H}$  matrix and the ground truth (pore or non-pore labels). The coefficient vector  $h_{t_{k \times 1}}$  from an unseen test sample served as the input to the trained ML classifiers and is representative of the so-called feature vector. The data to train, test, and validate our NMF $k$ -based pore detection ML model was divided into 5 separate sets, which coincide with 5 data acquisition trials of the same AM experiment (see Sect. 2). Our training and testing schemes were set up such that the ML models were trained on the  $\mathbf{H}$  matrix estimated for one of the 5 data sets, while the remaining 4 datasets were used for testing. Table 1 provides a summary of the number of pore samples, number of non-pore samples, and class imbalance ratio for the training and testing data of each respective set. To account for the imbalance in the training data, we resample the training set through under-sampling and 10-fold cross-validation during model development.

Two different ML methodologies were implemented in this study: supervised and unsupervised data labeling. For supervised data labeling, the ground truth labels generated from correlating X-ray radiography were used to label each training and testing sample (acoustic signal) as either pore or no pore. The alternative to this approach is unsupervised or *blind* data labeling. Demonstrating the efficacy of blind labeling is a major advantage when limited ground truth is available, as is commonly the case for complex AM experiments whereby proper registration is highly nontrivial. Note that the NMF $k$  algorithm is inherently a blind feature extraction technique and is, thus, a blind ML technique on its own. By introducing blind labeling, the ML workflow becomes fully autonomous. The blind labeling was achieved by blindly assigning class labels in the feature space  $\mathbf{H}$  by partitioning the latent space into two groups with an unsupervised  $k$ -means clustering algorithm. The blindly labeled feature space  $\mathbf{H}$  was then used to train ML classifiers. The resulting ML model was tested on hold-out data which was not used for blind labeling or training. Finally, the accuracy of the models trained and tested on blindly labeled latent spaces was computed by comparing the classifier predictions to the ground truth of each observation to the classification results. In this framework, the classifier predicts a test sample as belonging to cluster 1 or cluster 2. These clusters were associated as being pore or non-pore affiliated in an ad-hoc fashion.

Classification was performed using decision trees (DTR),  $k$ -nearest neighbors (KNN), Gaussian discriminant analysis (DSC), nonlinear support vector machine (SVM), adaptive boosting (ADB), gentle adaptive boosting (GTB), logistic regression boosting (LTB), robust boosting (ROB), and random undersampled boosting (ROB). The performance of all algorithms was compared for both accuracy and computational cost in order to resolve the best suited for the proposed ML workflow. Details pertaining to the mathemat-

ical machinery and implementation of these models can be found in foundational pattern recognition texts [51].

## 4 Results

This section presents the results obtained by applying the algorithmic workflow described in Sect. 3 to the acoustic LPBF data. We first show that the NMF $k$  routine produces well-suited latent spaces for pore identification and spectrum interpretation. We then demonstrate that this methodology results in highly accurate and efficient ML models.

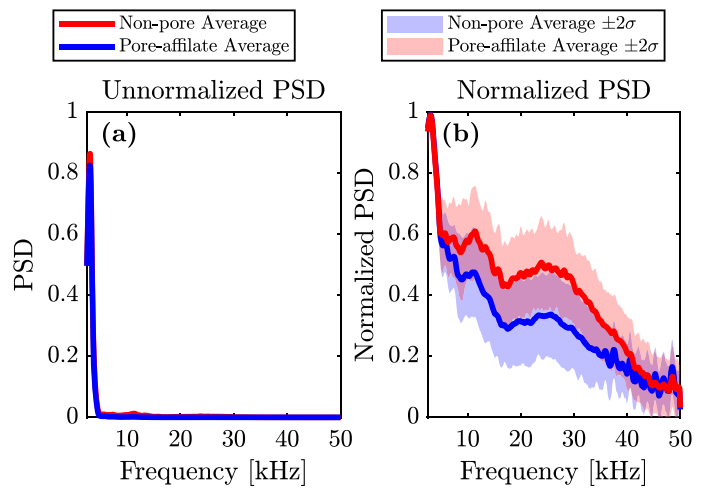
### 4.1 NMF $k$ decomposition of the AM data

The PSD estimate of each analysis window was computed for all data sets and was then split into separate training and testing sets per Table 1. Prior to NMF $k$  featurization, the PSD estimates  $\mathbf{V}$  were normalized per  $[\ln \mathbf{V} - \min[\ln \mathbf{V}]] / \max[\ln \mathbf{V} - \min[\ln \mathbf{V}]]$  to enhance the sensitivity over the frequency range associated with the LPBF process. This normalization produces a logarithmic scaling of  $\mathbf{V}$  (which is a standard scale for PSD estimates) normalized between 0 and 1 (which is a standard normalization for ML tasks). Figure 3a shows the unnormalized signals at pore and non-pore locations and Fig. 3b shows the normalized signals. The solid line in Fig. 3b represents the average PSD across all observations in training set 1, and the color band shows the associated variance.

The NMF $k$  algorithm was applied to the normalized PSD estimates to return a  $\mathbf{W}$  basis dimension  $k$ . Figure 4a shows this decomposition of data set 1, where  $k = 3$  is indicated as the optimal basis dimension based on the Silhouette width  $S$ . Figure 4b shows the discovered basis signals  $\mathbf{W}$  that minimize  $\epsilon(k)$  over the 10 independent NMF iterations computed at  $k = 3$ . From this, it is apparent that a broadband signal (e.g., hidden signal 1) is discovered in the data with dominant frequencies between 10 and 50 kHz. Lastly, Fig. 4c and d show the coefficients from the  $\mathbf{H}$  matrix, which represents the training coefficients, and testing coefficients  $h_{t_{k \times 1}}$  which are obtained for test PSDs  $V_t$  by solving the least squares problem,  $\min_{h_t} (\mathbf{W}h_t - V_t)$ , Fig. 4c and d depict a clear demarcation of signal groups in both training and test sets.

Figure 5 depicts the results of the NMF $k$  routine applied to all 5 training and test sets (Table 1). The left two columns show the AIC and silhouette scores used to determine the optimal number of signals (denoted by a green bar) as determined by NMF $k$ . The middle-right column shows the discovered PSD basis signals in the training datasets with line thickness indicating *importance* of the basis signal for pore prediction as determined by the normalized impurity

**Fig. 3** Normalization of acoustic data: PSDs for training set 1 of the ML workflow. The average PSD value of **a** unnormalized and **b** normalized to enhance the separation on a linear scale between the pore-affiliated and non-pore-affiliated classes

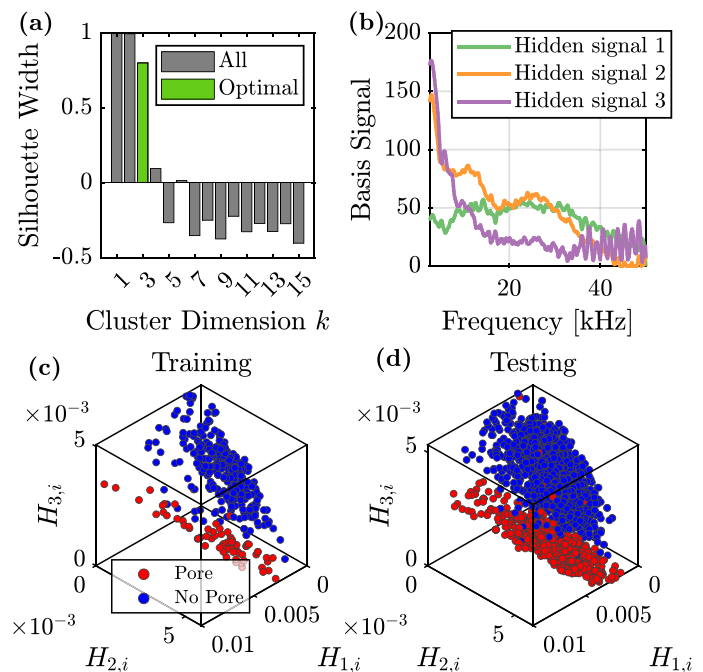


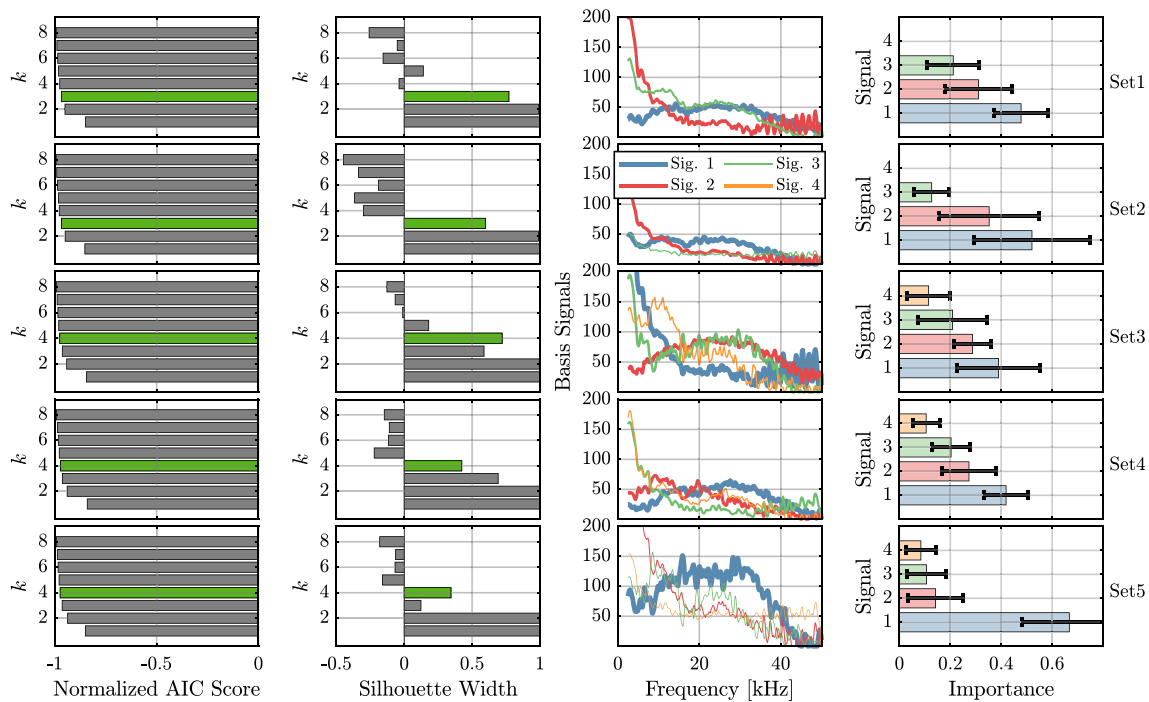
reduction for each feature in random forest emulators using 50 trees and bootstrap samples. Lastly, the right column of Fig. 5 shows the feature importance with error bars of depicting the standard deviation of impurity reduction across the bootstrap samples.

Several common trends arise between the data sets. Namely, the discovered basis signals always return a signal with elevated amplitude in the 10–40 kHz range, and another with high amplitude in the low-frequency range. Since all PSDs in Fig. 3 have elevated energy in the low-frequency range, the NMF $k$  decomposition returns a signal with low-frequency energy that all signals match with. However, pore-affiliated spectra possess a prominent elevation in the high-frequency ranges as compared to their non-pore counterparts. Hence, a high-frequency and broad-band basis

signal is returned from NMF $k$  that correlates strongly to pore-affiliated data (i.e., has a high matching coefficient). To this end, the broad-band signal provides the contribution to the reconstructed spectra that signifies pore formation, which is further supported by the fact that the broad-band signal was found to be of high importance from the random forest emulators. This finding is in good agreement with the identified frequency ranges reported in [29, 30, 36, 37] which heuristically found frequencies between 10 and 40 kHz to correspond to pore formation. We have shown here that the same frequency range can be recovered in an automated and optimized fashion via NMF $k$  and have returned the underlying basis of signals which allow for such distinctions in the acoustic spectra to be possible.

**Fig. 4** NMF $k$ -based feature analysis of acoustic data: **a** Silhouette width revealing the optimal dimension  $k$  to be 3 based on clustering, **b** basis signals recovered from NMF $k$ , **c** shows the feature space **H** constructed from the labeled training data and basis signals **W**, and **d** shows the testing data plotted in this same **H** space based on the training basis **W**



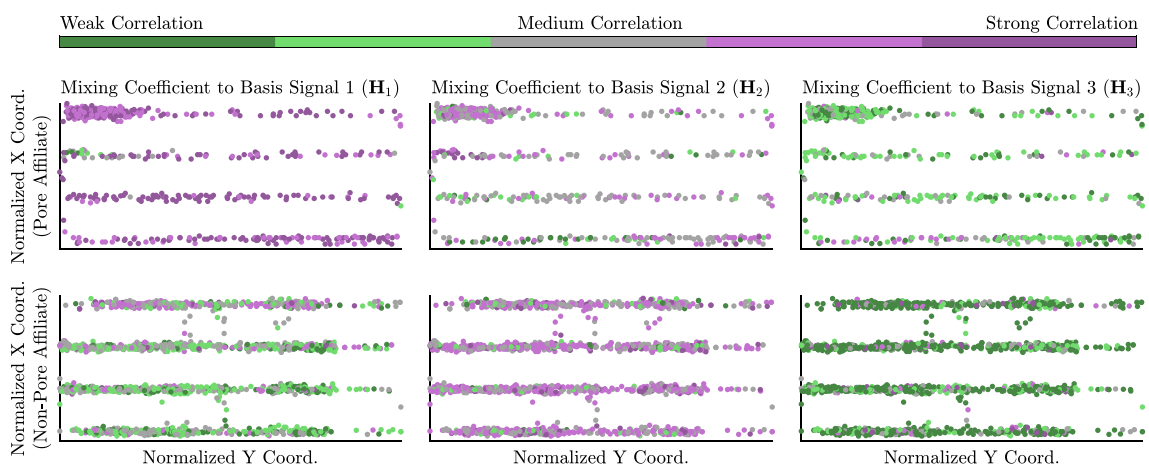


**Fig. 5** NMFk-based discovered signals and associated feature importance: details on optimal numbers of signals  $k$ , basis signals discovered from the training datasets, and associated feature importance of each discovered signal with respect to pore or non-pore classification

### 4.2 Spatial mappings of H-space

Prior to performing classification experiments, we first explored possible spatial correlations. This is an important step in order to ensure that the clustering of  $\mathbf{H}$  coefficients is not due to an artifact of the experiment, such as laser start-up or turn points, but is rather associated with the existence of a pore formation. This was achieved by mapping the first 3

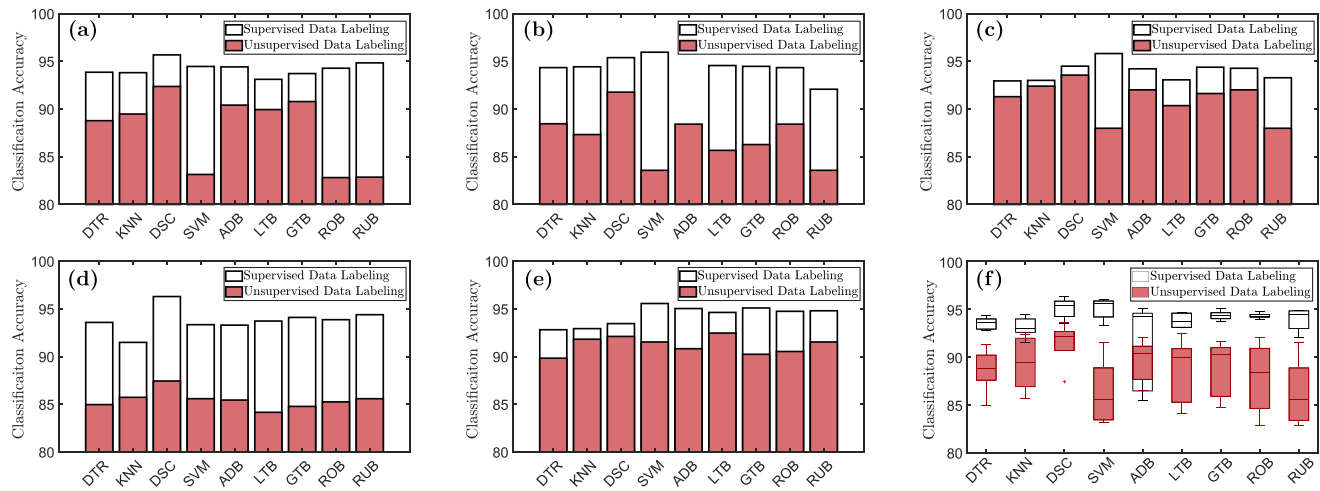
coefficients of  $\mathbf{H}$  matrix to the spatial location from which the acoustic data is collected and comparing their values for pore and non-pore labeled samples. Each dot represents a sample of a test acoustic signal and a color value is assigned to represent the magnitude of the  $\mathbf{H}$  coefficient. For this case, training data from set 2 is used to generate the  $\mathbf{W}$  matrix, and the coefficients of all five experiments are superimposed in Fig. 6. Figure 6 shows that the first mixing coefficient value



**Fig. 6** Spatial mappings of  $\mathbf{H}$  coefficients: spatial correlations of the coefficient matrix  $\mathbf{H}$  for all pore and non-pore observations based on the basis signals  $\mathbf{W}$  recovered from training set 2. All tracks from the data set are overlaid and colors are assigned based on the values of  $\mathbf{H}$

coefficients to indicate the correlation to the basis signals in  $\mathbf{W}$ . Note the  $x$  and  $y$  axes are flipped from Fig. 1 for presentation purposes. This figure further confirms our hypothesis that the  $\mathbf{H}$  coefficients are able to distinguish pore and non-pore signals





**Fig. 7** Supervised and unsupervised ML model classification accuracy: classification accuracy for a suite of ML algorithms for supervised and unsupervised ML models. Subplot a–e correspond to sets 1–5 defined in Table 1, respectively, and f shows the box-plots for all training and testing permutations together

( $H_1$ ) is on average far higher for pore data when compared to non-pore data. Similarly, the second mixing coefficient ( $H_2$ ) is higher for non-pore data than pore data. This is expected from the results presented in Fig. 5. Moreover, the values of the  $H$  coefficients are close to uniform across the tracks and are largely determined by the underlying pore versus no pore state. Hence, it is unlikely that the  $H$  clusters are formed due to experimental artifacts. To summarize, this supports the hypothesis that the spectral features of basis signals contain process-relevant information for pore state detection.

### 4.3 Classification experiments

Each data set of Table 1 was evaluated using the ML classifiers listed in Sect. 3.2 whereby the training was used to recover  $H$  and  $W$ , and  $h_t$  was computed to featurize each test sample. The classification accuracy was computed based on the number of correct predictions versus the total number of observations, and the results for each classifier and test set are summarized in Fig. 7. The x-axis indicates which classifier was used, and the y-axis indicates the classification accuracy for both supervised and unsupervised data labeling. Among various ML algorithms investigated here, non-linear SVM and discriminant analysis consistently return the highest accuracy with the scores of approximately 95% returned for the majority of data sets when supervised labeling is employed. When unsupervised labeling is considered, the performance of SVM drops substantially, and discriminant analysis is the clear front-runner. Figure 7 depicts this clearly with a boxplot comparing the performance across all data sets. Moreover, the prediction accuracy across all classifiers based on supervised labeling is greater 90% across the data sets, whereas the unsupervised labeling is typically greater

85%. To summarize, all the investigated ML algorithms perform well, showing that the information extracted from the NMFk decomposition leads to an effective feature space for classification.

Lastly, it is important to note that since the NMFk procedure begins with a random initialization, one should check to what effect the initialization has on the optimal basis selection and classification performance. Moreover, one should ensure that the number of random initializations (in our case 10) is sufficient for consistent results. We report that these parameters did not modify the findings of basis patterns, nor their ability to provide informative predictions.

## 5 Conclusions

In this work, we presented an automated workflow to recover the spectral signatures of keyhole pore formation during LPBF processes that lead to fast, accurate, and reliable models for high-throughput acoustic monitoring data. Using five training and testing datasets of acoustic measurements, which were labeled as pore and non-pore classes, the PSD estimates of each signal class were decomposed using non-negative matrix factorization with  $k$ -means clustering, called NMFk. This matrix decomposition provided a low-dimensional latent representation of the spectral data (i.e., the coefficients of the  $H$  matrix) corresponding to pore and non-pore affiliated spectral patterns (i.e., the basis signals  $W$ ). Moreover, the decomposed matrices  $H$  and  $W$  for each training set (Table 1) provided us with a dictionary of spectral patterns that could in turn be used to develop predictive ML models to classify new test samples. Classification

of our test datasets was performed based on a supervised (>95% accuracy) and unsupervised (>90% accuracy) training labeling schemes and using a suite of ML classifiers. Furthermore, we discovered that simple ML classifiers such as Gaussian Linear Discriminate Analysis were able to accurately perform these predictions. This means that we can use relatively few training instances and have good accuracy in comparison to deep learning methods (e.g., CNN-LSTMs), which require large amounts of training data. Furthermore, the computational cost to train, test, and predict is very low, which makes our NMFk-based ML model attractive for in-situ AM process monitoring and the recovered basis signals also allow us to represent the acoustic measurements in a greatly compressed representation, alleviating data storage demands.

As opposed to part-level classification, our approach delivers binary pore versus non-pore statues for a localized patch of build material similar to the approach in [29]. Moreover, we have *blindly* recovered the spectral patterns necessary to perform classification using novel matrix decomposition methods, and our recovered basis signals were shown to corroborate the importance of the broadband spectral ranges of [29, 36, 37] in a statistically *blind* fashion. As such, our methodology elucidates the relevant spectral content associated with pore formation, which can be leveraged in future works to link these acoustic signals to the physics of the pore formation process.

To conclude, our ML workflow demonstrated the potential to identify the formation of keyhole pores with high accuracy throughout multiple single-layer LPBF experiments. This capability is a step toward online monitoring of full 3-dimensional builds where the location and density of keyhole pores may critically affect the performance of the produced part. By making this identification in-situ via the acoustic PSD, manufacturers can make decisions about the criticality of the density and location of keyhole pores. This may in turn bolster the quality assurance of LPBF build while also deterring unnecessary scrapping of parts and reducing post-build evaluations. Moving forward, this work could be augmented by evaluating multi-layer build experiments and investigating multi-sensor datastreams [30] (e.g., thermal, optical) using recent advances in tensor decompositions [21, 50]. Lastly, our pore detection scheme could be further bolstered with the use of secondary models constructed to explicitly localize pores once the primary detection algorithm has detected a pore, thus furthering the ability to spatially resolve pore location in a build.

**Author contribution** JRT: conceptualization, formal analysis, investigation, software, writing—original draft, writing—review and editing; MKM: conceptualization, investigation, writing—original draft, writing—review and editing; SK: conceptualization, supervision, writing

— review and editing; AJW: supervision, investigation, writing—review and editing; BA: writing—review and editing; EBF: supervision, writing—review and editing; J-BF: investigation, formal analysis, data collection, writing—review and editing; GMG: conceptualization, investigation, writing—review and editing; NPC: conceptualization, investigation, writing—review and editing; PJD: conceptualization, investigation, writing—review and editing; MJM: conceptualization, writing—review and editing, investigation.

**Funding** Los Alamos National Laboratory (LANL) co-authors were funded by the LANL Laboratory Directed Research and Development (LDRD) Director's Initiative AI@Sensor Project #20200669DI. LANL is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (Contract No. 89233218CNA000001). JRT is partially supported by a National Science Foundation Graduate Research Fellowship under Grant No. DGE - 1746047 during the paper writing process. SK thanks the Environmental Molecular Sciences Laboratory for its support. Environmental Molecular Sciences Laboratory is a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830. SK thanks the Environmental Molecular Sciences Laboratory for its support. Environmental Molecular Sciences Laboratory is a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Portions of this work were performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The Advanced Light Source is supported by the US Department of Energy under contract no. DE-AC02-05CH11231.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Do A-V, Khorsand B, Geary SM, Salem AK (2015) 3d printing of scaffolds for tissue regeneration applications. *Adv Healthc Mater* 4(12):1742–1762
2. Gross BC, Erkal JL, Lockwood S Y, Chen C, Spence DM (2014) Evaluation of 3d printing and its potential impact on biotechnology and the chemical sciences. *Anal Chem* 86(7):3240–3253
3. Bak D (2003) Rapid prototyping or rapid production? 3d printing processes move industry towards the latter. *Assem Autom* 23(4):340–345

4. Lee J-Y, Jia An CK, Chua CK (2017) Fundamentals and applications of 3d printing for novel materials. *Appl Mater Today* 7:120–133
5. Duda T, Raghavan LV (2016) 3d metal printing technology. *IFAC-PapersOnLine* 49(29):103–110
6. DebRoy T, Wei HL, Zuback JS, Mukherjee T, Elmer JW, Milewski JO, Beese AM, Wilson-Heid A, De A, Zhang W (2018) Additive manufacturing of metallic components – process, structure and properties. *Prog Mater Sci* 92:112–224
7. Vandenbroucke B, Kruth J-P (2007) Selective laser melting of bio-compatible metals for rapid manufacturing of medical parts. *Rapid Prototyp J* 13(4):196–203
8. Ventola CL (2014) Medical applications for 3d printing: current and projected uses. *Pharm Ther* 39(10):704
9. Rengier F, Mehndiratta A, von Tengg-Kobligk H, Zechmann CM, Unterhinninghofen R, Kauczor H-U, Giesel FL (2010) 3d printing based on imaging data: review of medical applications. *Int J Comput Assist Radiol Surg* 5(4):335–341
10. Liu R, Wang Z, Sparks T, Liou F, Newkirk J (2017) Aerospace applications of laser additive manufacturing. In *Laser Additive Manufacturing*, pages 351–371. Elsevier
11. Dey NK (2014) Additive manufacturing laser deposition of ti-6al-4v for aerospace repair application
12. Gao J, Folkes J, Yilmaz O, Gindy N (2005) Investigation of a 3d non-contact measurement based blade repair integration system. *Aircr Eng Aerosp Technol* 77(1):34–41
13. Olakanmi EO, Cochrane RF, Dalgarno KW (2011) Densification mechanism and microstructural evolution in selective laser sintering of al-12si powders. *J Mater Process Technol* 211(1):113–121
14. Grasso M, Colosimo BM (2017) Process defects and in situ monitoring methods in metal powder bed fusion: a review. *Meas Sci Technol* 28(4):044005
15. Attar H, Calin M, Zhang LC, Scudino S, Eckert J (2014) Manufacture by selective laser melting and mechanical behavior of commercially pure titanium. *Mater Sci Eng A* 593:170–177
16. Thijs L, Verhaeghe F, Craeghs T, Van Humbeeck J, Kruth J-P (2010) A study of the microstructural evolution during selective laser melting of ti-6al-4v. *Acta Mater* 58(9):3303–3312
17. Berumen S, Bechmann F, Lindner S, Kruth J-P, Craeghs T (2010) Quality control of laser- and powder bed-based additive manufacturing (AM) technologies. *Phys Procedia* 5:617–622
18. McCann R, Obeidi MA, Hughes C, McCarthy Éanna, Egan DS, Vijayaraghavan Rajani K, Joshi AM, Garzon VA, Dowling DP, McNally PJ, Brabazon D (2021) In-situ sensing, process monitoring and machine control in laser powder bed fusion: a review. *Addit Manuf* 45:102058
19. Grasso M, Remani A, Dickins A, Colosimo BM, Leach RK (2021) In-situ measurement and monitoring methods for metal powder bed fusion: an updated review. *Meas Sci Technol* 32(11):112001
20. Yang H-C, Huang C-H, Adnan M, Hsu C-H, Lin C-H, Cheng F-T (2021) An online AM quality estimation architecture from pool to layer. *IEEE Trans Autom Sci Eng* 18(1):269–281
21. Khanzadeh M, Tian W, Yadollahi A, Doude HR, Tschopp MA, Bian L (2018) Dual process monitoring of metal-based additive manufacturing using tensor decomposition of thermal image streams. *Addit Manuf* 23:443–456
22. Md Shahjahan H, Taheri H (2020) In situ process monitoring for additive manufacturing through acoustic techniques. *J Mater Eng Perform* 29(10):6249–6262
23. Masinelli G, Shevchik SA, Pandiyan V, Quang-Le T, Wasmer K (2020) Artificial intelligence for monitoring and control of metal additive manufacturing. In *Industrializing additive manufacturing*, pages 205–220. Springer International Publishing
24. Duley WW, Mao YL (1994) The effect of surface condition on acoustic emission during welding of aluminium with CO<sub>2</sub> laser radiation. *J Phys D: Appl Phys* 27(7):1379–1383
25. Wang F, Mao H, Zhang D, Zhao X, Shen Y (2008) Online study of cracks during laser cladding process based on acoustic emission technique and finite element analysis. *Appl Surf Sci* 255(5):3267–3275
26. Lee S, Ahn S, Park C (2013) Analysis of acoustic emission signals during laser spot welding of SS304 stainless steel. *J Mater Eng Perform* 23(3):700–707
27. Koester LW, Taheri H, Bond LJ, Faierson EJ (2019) Acoustic monitoring of additive manufacturing for damage and process condition determination. In *AIP Conference proceedings*, volume 2102, page 020005. AIP Publishing LLC
28. Wasmer K, Kenel C, Leinenbach C, Shevchik SA (2017) In situ and real-time monitoring of powder-bed AM by combining acoustic emission and artificial intelligence. In *Industrializing additive manufacturing - proceedings of additive manufacturing in products and applications - AMPA2017*, pages 200–209. Springer International Publishing
29. Tempelman JR, Wachtor AJ, Flynn EB, Depond PJ, Forien JB, Guss GM, Calta NP, Matthews MJ (2022) Detection of keyhole pore formations in laser powder-bed fusion using acoustic process monitoring measurements. *Addit Manuf* 55:102735
30. Tempelman JR, Wachtor AJ, Flynn EB, Depond PJ, Forien J-B, Guss GM, Calta NP, Matthews MJ (2022) Sensor fusion of pyrometry and acoustic measurements for localized keyhole pore identification in laser powder bed fusion. *J Mater Process Technol* 308:117656
31. Seleznev M, Gustmann T, Friebe JM, Peuker UA, Kühn U, Hufenbach JK, Biermann H, Weidner A (2022) In situ detection of cracks during laser powder bed fusion using acoustic emission monitoring. *Addit Manuf Lett* 3:100099
32. Fang Q, Xiong G, Zhou M, Tamir TS, Yan C-B, Wu H, Shen Z, Wang F-Y (2022) Process monitoring, diagnosis and control of additive manufacturing. *IEEE Trans Autom Sci Eng* 1–27
33. Shevchik SA, Kenel C, Leinenbach C, Wasmer K (2018) Acoustic emission for in situ quality monitoring in additive manufacturing using spectral convolutional neural networks. *Addit Manuf* 21:598–604
34. Shevchik SA, Masinelli G, Kenel C, Leinenbach C, Wasmer K (2019) Deep learning for in situ and real-time quality monitoring in additive manufacturing using acoustic emission. *IEEE Trans Ind Inform* 15(9):5194–5203
35. Pandiyan V, Drissi-Daoudi R, Shevchik S, Masinelli G, Le-Quang T, Logé R, Wasmer K (2022) Deep transfer learning of additive manufacturing mechanisms across materials in metal-based laser powder bed fusion process. *J Mater Process Technol* 303:117531
36. Pandiyan V, Drissi-Daoudi R, Shevchik S, Masinelli G, Logé R, Wasmer K (2020) Analysis of time, frequency and time-frequency domain features from acoustic emissions during laser powder-bed fusion process. *Procedia CIRP* 94:392–397
37. Khairallah SA, Sun T, Simonds BJ (2021) Onset of periodic oscillations as a precursor of a transition to pore-generating turbulence in laser melting. *Addit Manuf Lett* 1:100002
38. Ren Z, Gao L, Clark SJ, Fezzaa K, Shevchenko P, Choi A, Everhart W, Rollett AD, Chen L, Sun T (2023) Machine learning-aided real-time detection of keyhole pore generation in laser powder bed fusion. *Science* 379(6627):89–94
39. Wirtz SF, Cunha A, Labusch M, Marzun G, Barcikowski S, Söffker D (2018) Development of a low-cost FPGA-based measurement system for real-time processing of acoustic emission data: Proof of concept using control of pulsed laser ablation in liquids. *Sensors* 18(6):1775
40. Forien J-B, Calta NP, DePond PJ, Guss GM, Roehling TT, Matthews MJ (2020) Detecting keyhole pore defects and monitoring process signatures during laser powder bed fusion: a correlation

- between in situ pyrometry and ex situ x-ray radiography. *Addit Manuf* 35:101336
41. Cichocki A, Zdunek R, Phan AH, Amari SI (2009) *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons
  42. Alexandrov BS, Vesselinov VV (2014) Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization. *Water Resour Res* 50:7332–7347
  43. Vesselinov VV, Alexandrov BS, O'Malley D (2018) Contaminant source identification using semi-supervised machine learning. *J Contam Hydrol* 212:134–142
  44. Wagstaff K, Cardie C, Rogers S, Schrödl S (2001) Constrained *k*-means clustering with background knowledge. In *Icml*, volume 1, pages 577–584
  45. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401:788–791
  46. Martin AA, Calta NP, Khairallah SA, Wang J, Depond PJ, Fong AY, Thampy V, Guss GM, Kiss AM, Stone KH, Tassone CJ, Weker JN, Toney MF, van Buuren T, Matthews MJ (2019) Dynamics of pore formation during laser powder bed fusion additive manufacturing. *Nat Commun* 10:1–10
  47. Forien J-B, Depond PJ, Guss GM, Jared BH, Madison JD, Matthews MJ (2019) Effect of laser power on roughness and porosity in laser powder bed fusion of stainless steel 316L alloys measured by x-ray tomography. Technical report, Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States)
  48. Welch P (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoust* 15(2):70–73
  49. Grais EM, Erdogan H (2013) Spectro-temporal post-enhancement using MMSE estimation in NMF based single-channel source separation. In *Interspeech 2013. ISCA*
  50. Vesselinov VV, Mudunuru MK, Karra S, O'Malley D, Alexandrov BS (2019) Unsupervised machine learning based on non-negative tensor factorization for analyzing reactive-mixing. *J Comput Phys*
  51. Bishop CM (2011) *Pattern recognition and machine learning*. New York Inc., Springer-Verlag

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Joshua R. Tempelman<sup>1,2</sup> · Maruti K. Mudunuru<sup>3</sup> · Satish Karra<sup>4</sup> · Adam J. Wachtor<sup>1</sup> · Bulbul Ahmed<sup>5</sup> · Eric B. Flynn<sup>6</sup> · Jean-Baptiste Forien<sup>7</sup> · Gabe M. Guss<sup>8</sup> · Nicholas P. Calta<sup>7</sup> · Phillip J. DePond<sup>7</sup> · Manyalibo J. Matthews<sup>7</sup>

<sup>1</sup> Engineering Institute, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>2</sup> Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

<sup>3</sup> Watershed & Ecosystem Science, Pacific Northwest National Laboratory, Richland, WA 99352, USA

<sup>4</sup> Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352, USA

<sup>5</sup> Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>6</sup> Space Remote Sensing and Data Science, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>7</sup> Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

<sup>8</sup> Engineering Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA