



Explainable AI for layer-wise emission prediction in laser fusion

Weihong “Grace” Guo^{a,b}, Vidita Gawade^{a,b}, Bi Zhang^c, Yuebin Guo^{b,d,*}

^a Department of Industrial and Systems Engineering, Rutgers University-New Brunswick, Piscataway, NJ 08854, United States of America

^b New Jersey Advanced Manufacturing Institute, Rutgers University-New Brunswick, Piscataway, NJ 08854, United States of America

^c Department of Mechanical & Energy Engineering, Southern University of Science and Technology, Shenzhen 518055, China

^d Department of Mechanical and Aerospace Engineering, Rutgers University-New Brunswick, Piscataway, NJ 08854, United States of America

ARTICLE INFO

Article history:

Available online 5 April 2023

Keywords:

Additive manufacturing
Machine learning
Explainable AI

ABSTRACT

The dynamic behavior of melt pools in powder bed-based laser fusion is very challenging to model using physics-based models and conventional black-box data-driven models. Explainable Artificial Intelligence is developed in this work to advance the understanding of convoluted links of non-sequential process physics, online time series sensing data, and process anomaly (e.g., overheating in the melt pool). A Shapley Additive Explanations (SHAP)-enabled Deep Neural Network-Long Short-Term Memory (DNN-LSTM) model has been developed as a mechanism to integrate process parameter knowledge with process history information through online sensing data while providing local and global model interpretation and transparency.

© 2023 CIRP. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Selective laser melting (SLM) is characterized by a unique thermal cycle: rapid heating, fast melting and solidification, as well as melt-back [1] when compared with conventional manufacturing processes such as machining, casting, and forging. The complex thermal cycle is further coupled with part geometrical complexity, nonuniform heat dissipation, and various scanning patterns. The combination of these process phenomena often causes extreme melt pool dynamics, e.g., overheating, which leads to defects (e.g., dross, distortion, and porosity) [2]. Hence, process stability is key for industrial applications of SLM.

Among the in-situ monitoring techniques developed for the volatile SLM processes, pyrometry stands out as a robust technology to quantify the evolving melt pool temperatures via thermal radiation with high spatial-temporal resolutions during the laser consolidation process [3] and capture the process dynamics of melt pools in real-time [4,5]. The prediction of melt pool behaviors, however, remains a key challenge for printing high-quality metal parts. Physics-based modeling methods have inherent issues of exceedingly high computational cost, modeling uncertainty, and oversimplification. On the other hand, the measured emission data of the melt pool can be viewed as a process signature for heat dissipation and melt pool behavior, allowing the development of data-driven methods, e.g., machine learning (ML) or deep learning (DL), to track [6] and predict [7] overheating in SLM.

There is an increasing interest in data-driven methods to model the emission data and predict emissions for future layers. Such black-box methods include statistical models such as regression that use process inputs to estimate emissions, time series models to capture correlation in neighboring layers' emissions, and ML/DL models that can handle

more complex patterns [7]. There is a trade-off between model performance and interpretability as the black-box models have poorer interpretability than the white-box physics-based models.

Explainable DL can be exploited to harness knowledge from different data sources such as process history, which is most suitable for SLM emission prediction. Specifically, process history information (layer number and varying process conditions) can be leveraged to improve DL model interpretability using a deep multi-layer perceptron. The correlation of emission across neighboring layers can be modeled using Recurrent Neural Networks (RNNs) such as Long Short-Term Memory Networks. A deeper structure of explainable DL models helps generalize better and obtain better results than shallower counterparts [8]. A hybrid DL model for predicting emission using process parameters and layer-wise emission occurred only on a fixed set of process parameters (not varying process conditions) and, therefore, lacked interpretability [7].

Recent methodologies have focused on Explainable Artificial Intelligence (XAI). In the rather well-performing hybrid DL models, the explainability or interpretation still significantly lacks. While the terminology of XAI has differed in different domains, we will use explainability in this paper to mean the same thing: the ability for readers to understand how a model arrives at predictions considering the input information. Existing studies for XAI in manufacturing, let alone additive manufacturing, has been severely limited (a recent one is focused on XAI for surface roughness using a lesser black-box machine learning model) [9,10]. An XAI technique is any explanation approach to the model's prediction. Driven by current limitations and opportunities, the objective of this paper is two-fold. First, an additive feature attribution method namely Shapley Additive Explanations (SHAP) inspired by game theory is explored [11] to provide interpretability of an otherwise black-box model. Second, an explainable Deep Neural Network (DNN)-Long Short-Term Memory (LSTM) network model (DNN-LSTM) is developed as a mechanism to integrate the

* Corresponding author.

E-mail address: yuebin.guo@rutgers.edu (Y. Guo).

knowledge of varying process conditions by incorporating historical emission data online collected during SLM. Data-driven layer-wise emission prediction is feasible as the high-speed co-axial pyrometers can collect emission data in real time, which provides massive quality data to train a DL model to predict layer-wise emission. This paper aims to provide interpretability of a DL model otherwise black-box, integrate the process knowledge into the DL model, and predict layer-wise emission, which lays a foundation for DL-based process control to prevent melt pool overheating.

2. Experiment setup and measurement

SLM experiments were performed to fabricate 16 samples (numbered Box 1–16) of $10 \times 10 \times 5$ mm using an AconityMINI machine under printing conditions as shown in Fig. 1. The samples were made from stainless steel 316 L powders (wt.%: 69.5 Fe, 16.5 Cr, 10.4 Ni, 2.02 Mo, 1.27 M) with particle size distributions: $d_{10}=18.00 \mu\text{m}$, $d_{50}=29.66 \mu\text{m}$, and $d_{90}=45.78 \mu\text{m}$. SS 316 L offers high resistance to corrosion and withstands high temperatures, while thermal conductivity is influenced by the specific pressure and composition of the gas used in the process [12]. The scanning strategy comprises a raster scan to fill in the interior of the sample, followed by a frame scan. Real-time and in-situ pyrometry was conducted with two high-speed co-axial pyrometers which measure in the ranges of 1450–1700 nm and 2000–2200 nm, respectively, at a 100 kHz acquisition frequency, which allows one measurement every $10 \mu\text{s}$. The emissions were collected in a circular field of view (FOV) with a $\sim 500 \mu\text{m}$ diameter and a spatial accuracy within $\pm 10 \mu\text{m}$, which is larger than the melt pool size. This ensures enough spatial measurement resolution and the collected emissions within the whole melt pool along the laser scanning path. The measured raw emission data is stored as [X-position, Y-position, Emission] for each layer and the data can be further processed for data analytics.

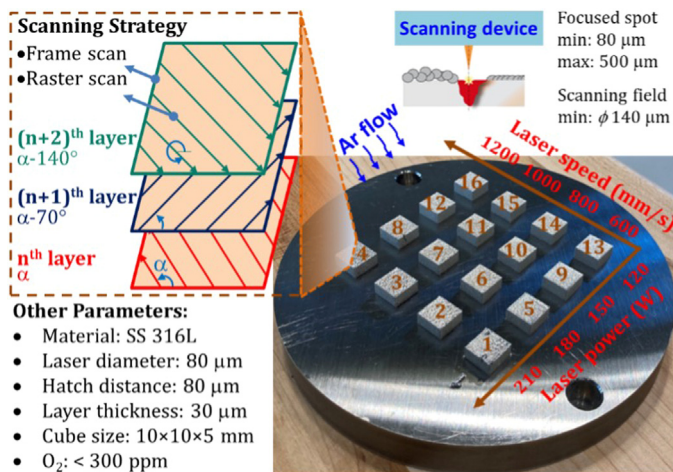


Fig. 1. SLM experiment with in-situ pyrometry.

3. SHAP-enabled explainable DNN-LSTM model development

3.1. Overview

As shown in Fig. 2, the methodology developed provides a mechanism to integrate the process history information and online

measured data within a hybrid DNN-LSTM deep learning framework while endowing local and global interpretability and transparency of the black-box model using Shapley Additive Explanations (SHAP). The DL model harnesses the behavior of layer-wise average emission (propensity to go up and down) while treating each layer as a time step via the LSTM network and harnesses the process history knowledge via the DNN. The inputs' impact on the output is described by SHAP.

3.2. DNN-LSTM model framework

As shown in Fig. 3, the inputs of the proposed DNN-LSTM model are comprised of box-wise process knowledge (laser power w_b , laser speed v_b , and layer number l) and average emissions observed in the previous 5 layers ($y_{b,l-j}$ where $j = 1, 2, \dots, 5$). The training and test datasets were split at the same random point in the layer number for all boxes $b = 1, 2, \dots, 16$, forming the training dataset to be composed of all layers $l = 15, 16, \dots, 120$ and the test dataset to be composed of all layers $l = 120, 121, \dots, 166$. The LSTM network is a type of RNN that harnesses the temporal dependencies of observations. There are 16 LSTM networks developed (one for each box) that train in parallel. Fig. 3 shows the LSTM network for a given box b that incorporates previous layers' average emission in two stacked hidden LSTM layers (both with 50 units) and a fully connected (FC) layer for predicting a subsequent layer's average emission $\hat{y}_{b,l}$. The LSTM model runs on 1000 epochs with early stopping criteria of the validation loss not decreasing for more than 50 epochs.

The DNN network regresses the process history knowledge of each box to extract features using a hidden (H) feedforward layer of 15 units with 'ReLU' activation and an FC layer with 'linear' activation. The DNN is trained on 2000 epochs with the same early stopping criteria as LSTM. Features extracted from each of the 16 LSTMs and the one DNN network are concatenated via a concatenate (C) layer. The output of the C layer is obtained in an H layer and an FC layer to output the predicted average emission ($\hat{y}_{b,l}$) for all boxes. This hybrid model is fitted using the mean squared error loss function with the Adam optimizer. The loss function calculates the mean of the summation of squared differences of true value $y_{b,l}$ and predicted value $\hat{y}_{b,l}$. This hybrid model has robust performance for emission prediction, yet it remains a black-box until the employment of the SHAP approach.

3.3. Shapley additive explanations (SHAP)

Although high accuracy can be achieved by the hybrid DL models, the challenge remains on interpretability due to the complexity of the model. SHAP tasks each feature variable a SHAP value (an importance value) for each instance of predictions enabling the interpretability of an inherent black-box model. Inspired by game theory, the model functions as the game and the feature input variables function as players of the game.

Computed in Eq. (1), a SHAP value ϕ_i for a specific feature i and an observation x is the difference between the expected model output and the expected model output conditioning the feature's value x_i . SHAP computes a deviation from the base value $E[f(x)]$ (expected model output given no feature values are provided) to predicted output value $f(x)$. The summation of the SHAP values $\phi_i(f, x)$ for all i and given x will be the model's predicted output value, $f(x)$.

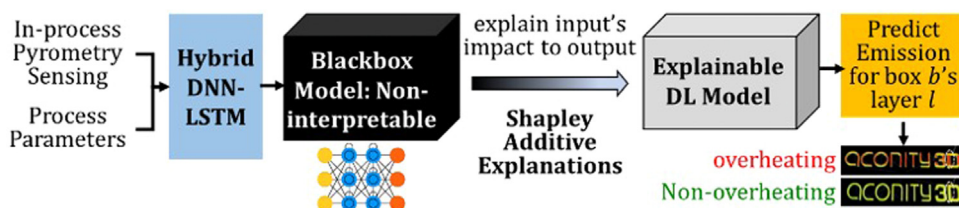


Fig. 2. Proposed hybrid deep learning (DL) with Explainable AI.

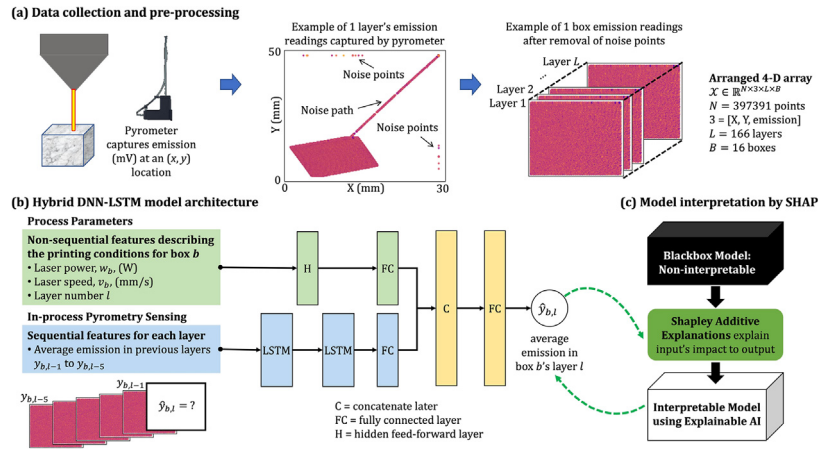


Fig. 3. Methodology architecture for layer-wise emission prediction with explainable AI.

$$\phi_i(f, x) = \sum_{z \subseteq x} \left(\left| z \right| \times \binom{|x|}{|z|} \right)^{-1} \times [f_x(z) - f_x(z_i)] \quad (1)$$

In this case study, the hybrid DNN-LSTM model f is depicted in Fig. 3, where an observation $x = [y_{b,l-1}, y_{b,l-2}, y_{b,l-3}, y_{b,l-4}, y_{b,l-5}, v_{b,l}, w_{b,l}, l]$, $|x| = 8$, $z \subseteq x$ represents all z vectors which are a subset of x , $|z|$ will be the number of features in z which ranges from 1 to 8, $f_x(z)$ represents the output of the model f including feature i , and $f_x(z_i)$ represents the output of the model f not including feature i . The weight given to marginal contributions of a m -feature model is calculated by $\left(|z| \times \binom{|x|}{|z|} \right)^{-1}$ and the marginal contribution of the model composed of subset z is calculated by $[f_x(z) - f_x(z_i)]$. Once the SHAP values are integrated into the model, its performance and interpretation can be evaluated. In this work, root-mean-square error (RMSE) is used to evaluate model performance, and process knowledge is used to evaluate the SHAP technique.

4. Model predictions and interpretability

4.1. Emission prediction results

The predictions are compared to the true average emissions using the RMSE. Table 1 summarizes how well the model performed on each box b on the training layers and test layers. The model performs with higher RMSEs for the boxes (1, 2, 3, 5, 7) with greater variance in the observed emissions and with lower RMSEs for the boxes with lower variance in the observed emissions. The interpretation in Sections 4.2 and 4.3 will focus on Boxes 6 and 10 and non-uniform layers with a significant variation in emissions to allow interpretation of how process parameters and layer-wise emission history impact the prediction.

Table 1
Hybrid DNN-LSTM model performance on each box.

Box #	1	2	3	4	5	6	7	8
Training RMSE	2.32	1.21	2.05	1.10	2.10	1.05	1.49	1.13
Test RMSE	2.19	1.25	2.09	1.03	2.03	1.05	1.42	1.20
Box #	9	10	11	12	13	14	15	16
Training RMSE	1.00	0.90	1.05	0.99	1.10	0.72	0.78	0.75
Test RMSE	0.96	0.93	0.99	1.05	1.17	0.67	0.66	0.71

Note: Lower metric values indicating better performance (unit in mV).

4.2. Global interpretability

Traditionally, a model is evaluated using performance metrics such as accuracy to indicate a reliable model, but the model could have good accuracy for the wrong reasons, and it is important to unveil reasonings for an intrinsically black-box model. Prior to the

integration of SHAP interpretation, one has no knowledge of how the model arrived at the prediction. Fig. 4 shows the overall average impact of each feature on the output $y_{b,l}$. The values in Fig. 4 represent each feature i 's mean absolute SHAP value ($|\phi_i|$) over all instances (layers) of the dataset. It can be seen from Fig. 4 that the average emission in layers $(l-1)$ and $(l-5)$ have a greater overall impact on the prediction than in layers $(l-2)$, $(l-3)$, and $(l-4)$. Laser power and laser speed are the top two features with the greatest average impact while layer number has the lowest average impact on the hybrid DNN-LSTM output magnitude.

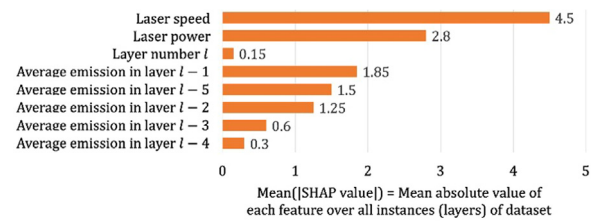


Fig. 4. Plot of average impact on DNN-LSTM output magnitude.

While Fig. 4 unveils the overall impact considering all boxes and all layers, detailed explanations of the features' impact on each layer's prediction will be elaborated in Figs. 5 and 6. The upper portions of Figs. 5 and 6 provide a global interpretation of the model to unveil the black-box model's thought process.

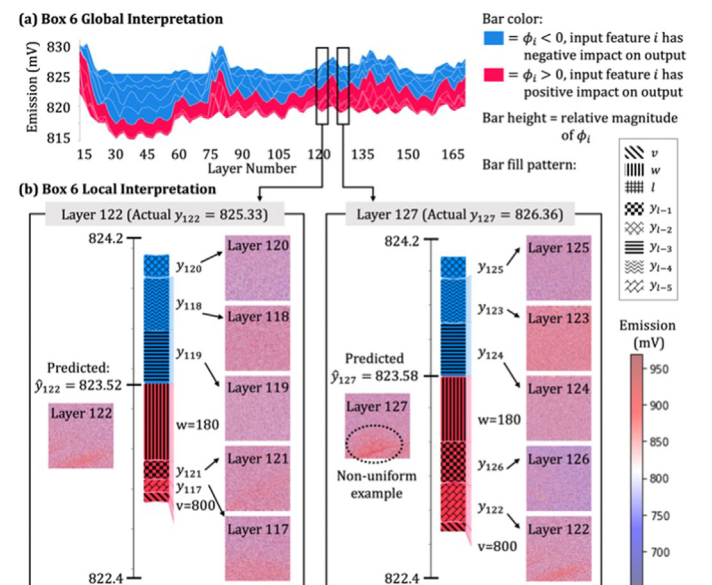


Fig. 5. Global and local force plots in Box 6.

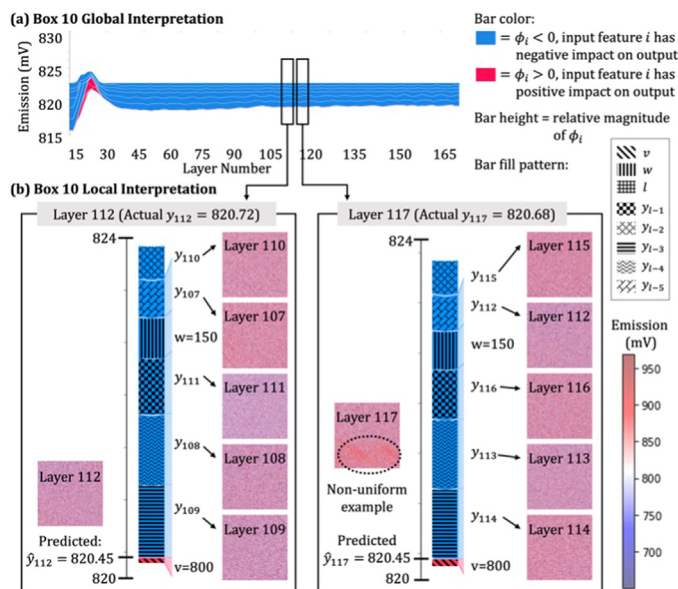


Fig. 6. Global and local force plots in Box 10.

The global force plot can be used to see what a model has learned and if the model has learned what naturally occurs to the box across layers. For Box 6 in Fig. 5(a), the global force plot shows both positive (red color) and negative (blue color) impacts from the different factors. Quantitatively, the attribution values retrieved indicate a positive impact by laser power 180 W and laser speed 800 mm/s and both positive and negative impacts by the lag features y_{l-j} . For Box 10 in Fig. 6(a), the global force plot shows more negative than positive impacts from different factors indicated by the predominately blue color. Quantitatively, the attribution values retrieved indicate that the model employs lag features while process parameters have little impact ($\phi_i \approx 0$).

When energy density is used to calculate the laser energy into powder materials, energy density can be correlated positively with emission readings. This model may learn what happens in the SLM process as the model is indicating a positive influence from power on emission readings while having “up and down” layer-wise average emission behaviors mimicked by $\phi_i < 0$ and $\phi_i > 0$.

4.3. Local interpretability

Local interpretability is enabled in local force plots (lower portion of Figs. 5 and 6) for each observation x of each input feature i 's impact on output $f(x)$ ($\hat{y}_{b,l}$). The individual force plots illustrate how the model computes a prediction ($\hat{y}_{b,l}$) for an instance of an observation (layer l). Each $\phi_i(f, x)$ is represented by a pattern, red bars indicate positive ϕ_i values and blue bars indicate negative ϕ_i values. The height of the bar indicates the relative impact of feature i on the prediction $\hat{y}_{b,l}$.

In Fig. 5(b), the individual force plots explain the predictions for Box 6's layers 122 and 127. These layers are selected due to their non-uniform emission reading and can indicate non-uniform heat dissipation resulting in possible defects. For each of those layers, the laser power 180 W, the laser speed 800 mm/s, and the previous lag 1 ($y_{6,l-1}$) and lag 5 ($y_{6,l-5}$) have a positive impact indicated by the red bars, lags 2, 3, and 4 ($y_{6,l-2}$, $y_{6,l-3}$, $y_{6,l-4}$) have a negative impact indicated by the blue bars, and layer number has no impact. Lag 1 and lag 5 impact on the current layer's emission prediction is supported by the fact that previous lag 1 and lag 5 are more correlated with the current layer's emission. The true values ($y_{6,l}$) and the predicted values ($\hat{y}_{b,l}$) are comparable.

In Fig. 6(b), individual force plots explain the predictions by the hybrid DNN-LSTM model for Box 10's layers 112 and 117. Layer 117 was selected due to its non-uniform emission reading, and layer 112 was selected to see to interpret the previous subsets of five lag layer

emission history. For each of these layers, laser speed 800 mm/s has a positive impact indicated by the red bars, meanwhile, all 5 lag layers and laser power 150 W have a negative impact indicated by the blue bars. In fact, lags 1, 3, and 4 have a greater negative impact than laser power 150 W, but lags 2 and 5 have a less negative impact than laser power 150 W.

Local force plots can be used at the process design and print stage to control the behavior of layers. The plots can confirm if the model is learning the SLM process history correctly by indicating the relative importance of features. These local force plots provide a means for a reader to make these interpretations from domain knowledge. For example, if the model indicates a high SHAP value for a domain-confirmed non-important feature, that could mean that the model may not be useful for future predictions. But this can also indicate new insight into the dynamic SLM process as a high SHAP value on a non-intuitive feature, especially locally, can be revealed. Each force plot allows local interpretation, but the overall story can change for a box at the global level. While local interpretation shows that each prediction is impacted differently by the features, the global force plots can be helpful to show what the black-box model has generally learned.

5. Conclusion

A SHAP-enabled hybrid DNN-LSTM model with transparency has been developed. The model integrates process parameters knowledge with layer-wise emission history for robust prediction performance of layers' average emission. Local and global interpretation endows an understanding of a black-box model's learning process. The methodology sheds light on the complicated links of process parameters, sensor data, and process anomaly. Future work includes other quantitative evaluations of SHAP. The methodology can be transferred to different types of part design, additive manufacturing printers, and powder materials for monitoring a time series attribute of the process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Herzog D, Seyda V, Wycisk E, Emmelmann C (2016) Additive Manufacturing of Metals. *Acta Mater* 117:371–392.
- [2] Fox JC, Moylan SP, Lane BM (2016) Effect of Process Parameters on the Surface Roughness of Overhanging Structures In Laser Powder Bed Fusion Additive Manufacturing. *Procedia CIRP* 45:131–134.
- [3] Furumoto T, Ueda T, Alkahari MR, Hosokawa A (2013) Investigation of Laser Consolidation Process for Metal Powder by Two-Color Pyrometer And High-Speed Video Camera. *CIRP Annals* 62/1:223–226.
- [4] Galkin G, Gawade V, Guo W, Guo YB (2022) In-situ and Real-Time 3D Pyrometry for Thermal History Diagnosis in Laser Fusion Process. *Manufacturing Letters* 33:862–871.
- [5] Gutknecht K, Haferkamp L, Cloots M, Wegener K (2020) Determining Process Stability of Laser Powder Bed Fusion Using Pyrometry. *Procedia CIRP* 95:127–132.
- [6] Mahato V, Obeidi MA, Brabazon D, Cunningham P (2020) An Evaluation of Classification Methods for 3D Printing Time-Series Data. *IFAC-PapersOnLine* 5/2:8211–8216.
- [7] Gawade V, Guo YB, Guo W (2022) Layer-wise Emissivity Prediction in Powder Bed Fusion using Time Series and Deep Learning Models. In: *Proceedings of the IISE Annual Conference & Expo 2022*.
- [8] Lago J, De Ridder F, De Schutter B (2018) Forecasting Spot Electricity Prices: deep Learning Approaches and Empirical Comparison of Traditional Algorithms. *Appl Energy* 221:386–405.
- [9] Guo W, Tian Q, Guo S, Guo YB (2020) A Physics-Driven Deep Learning Model for Process-Porosity Causal Relationship and Porosity Prediction with Interpretability in Laser Metal Deposition. *CIRP Annals* 69/1:205–208.
- [10] Tiensuu H, Tamminen S, Puukko E, Rönning J (2021) Evidence-Based and Explainable Smart Decision Support for Quality Improvement in Stainless Steel Manufacturing. *Applied Sciences* 11/22:1–19.
- [11] Lundberg S, Lee S (2017) A unified approach to interpreting model predictions. *31st Conference on Neural Information Processing Systems*, 1–10.
- [12] Wei LC, Ehrlich LE, Poweril-Palm PJ, Montgomery C, Beuth J, Malen JA (2018) Thermal Conductivity of Metal Powders for Powder Bed Additive Manufacturing. *Additive Manufacturing* 21:201–208.