NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Detecting the extent of co-existing anomalies in additively manufactured metal matrix composites through explainable selection and fusion of multi-camera deep learning features

Safdar, Mutahar; Wood, Gentry; Zimmermann, Max; Lamouche, Guy; Wanjara, Priti; Zhao, Yaoyao Fiona

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1080/17452759.2025.2515240
Virtual and Physical Prototyping, 20, 1, 2025-06-18

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=55d63171-bfff-4aec-915b-c5349cada065
https://publications-cnrc.canada.ca/fra/voir/objet/?id=55d63171-bfff-4aec-915b-c5349cada065

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

VIRTUAL AND PHYSICAL PROTOTYPING
2025, VOL. 20, NO. 1, e2515240 (39 pages)
https://doi.org/10.1080/17452759.2025.2515240

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Detecting the extent of co-existing anomalies in additively manufactured metal matrix composites through explainable selection and fusion of multi-camera deep learning features

Mutahar Safdar[a,b], Gentry Wood[c], Max Zimmermann[d], Guy Lamouche[b], Priti Wanjara[b] and Yaoyao Fiona Zhao [a]

[a]Department of Mechanical Engineering, McGill University, Montreal, Canada; [b]National Research Council Canada, Montreal, Canada; [c]Apollo-Clad Laser Cladding, a Division of Apollo Machine and Welding Ltd., Edmonton, Canada; [d]Fraunhofer Institute for Laser Technology ILT, Aachen, Germany

ABSTRACT

Process development for customised additively manufactured materials is challenging and labour-intensive. Advanced in-situ monitoring coupled with modern machine learning (ML) methods can expedite defect detection and qualification of additive manufacturing (AM) parts. Directed energy deposition (DED) processes offer flexibility to deposit material on existing complex parts for hybrid manufacturing and repairs. DED enables custom metal matrix composites (MMCs) like nickel tungsten carbide (Ni-WC) overlays on ferrous mining tools for enhanced wear resistance. However, co-existing anomalies specific to defects in the matrix, reinforcement and their interaction present development challenges. The challenge is compounded since the co-existing anomalies can exist in varying extents (e.g. absent, low, high). This study investigates dual mid-wave infrared (MWIR) cameras (FLIR and CLAMIR) for defect extent detection in Ni-WC MMCs. Deep learning features extracted with a fine-tuned vision transformer outperformed conventional methods by improving anomaly separability and revealing process-regime-aware feature distributions. Explainable artificial intelligence identified key MWIR features detecting six defect categories. Data ablation revealed FLIR's superior accuracy and generalisability under noise, while CLAMIR demonstrated robustness under instability. Explainable fusion enabled effective selection of camera features. Our work provides a foundation for ML-assisted development of AM-based Ni-WC and similar MMCs by facilitating in-situ detection of co-existing anomalies.

ARTICLE HISTORY
Received 1 April 2025
Accepted 29 May 2025

KEYWORDS
Defect extent detection; co-existing anomalies; metal matrix composites; explainable AI; video transformer features

1. Introduction

Additive manufacturing (AM) or 3D printing technologies enable layer-wise fabrication of parts from digital designs [1]. AM offers unique benefits when compared to traditional manufacturing technologies. Some key features include part consolidation, material efficiency, virtual inventory, complex designs, multi-material structures, to name a few [2]. Directed energy deposition or DED is a family of AM technologies that use an energy source (laser, electron beam, arc) to fuse feedstock (powder or wire) materials by melting, thereby producing a material deposit on a substrate [3]. One key feature of DED processing technology is its potential to repair, manufacture shape features on existing geometries (making hybrid products), as well as enhance parts with changing chemistry and/or mechanical properties [4]. This latter capability for tailoring parts can be highly beneficial for a wide range of industrial applications, with the current focus in this study being on a metal matrix composite (MMC) material. MMCs are composed of a metal matrix and one or more reinforcement materials. MMCs combine the desirable properties of metals (e.g. toughness, ductility) with those of reinforcements (e.g. stiffness, strength, wear resistance), but their manufacture can be challenging and costly [5]. For instance, DED-deposited nickel tungsten carbide (Ni-WC) overlays on a steel base/substrate can lead to improved wear resistance of functional components for diverse applications (e.g. agriculture, automotive, oil and gas, mining, and aerospace) [6], but their processing suffers from the intense efforts needed to develop/optimise the operational window and adapt/scale the parametric conditions to part/geometry specific requirements.

CONTACT Yaoyao Fiona Zhao yaoyao.zhao@mcgill.ca Department of Mechanical Engineering, McGill University, Montreal, QC, Canada H3A 0C3

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Advanced in-situ monitoring is an important component of the ongoing research efforts for AM defect detection [7]. In this regard, optical, acoustic, and laser-line scanning based techniques are most common [8,9]. Within the optical or vision-based techniques, monochrome or infrared (IR) sensors have been widely used in co-axial or off-axial configurations to monitor the process. Co-axial charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) cameras (monochrome) have been usually deployed to capture geometric melt pool features associated with generic process states. On the other hand, IR-based thermal cameras can capture intricate process dynamics from the melt pool and surrounding regions of the geometry being printed [10]. Certain AM applications can also benefit from the simplicity and low cost associated with using acoustic sensors. Laser-line scanning based approaches have been used to monitor the quality of the geometry being printed and detect defects in the macrostructure [11]. The existing barriers to AM process monitoring include challenges, such as real-time data processing, sensing capabilities and cost, accurate temperature measurement, and sensor integration. In their recent review on in-situ quality monitoring in laser AM, Chen et al. [8] recommended the use of multiple thermal cameras placed at different positions to overcome some of these challenges.

The use of advanced monitoring techniques coupled with state-of-the-art deep learning (DL) architectures opens the possibility of addressing a range of AM process concerns across different applications. The initial efforts were aimed at applying machine learning (ML) in AM tasks, such as process state prediction, anomaly detection, and parameter classification [12]. Porosity detection and classification in AM processes has been a key focus of several studies integrating in-situ monitoring with ML. Since then, these efforts have pivoted to more sophisticated and complex prediction tasks from both process (state/regime) and structure (macroscale and microscale) perspectives. The DED process is characterised by a complex interplay between the feedstock, the substrate and the heat source (e.g. laser beam in this research). The deposition freedom in DED supports applications involving complex geometries, as well as multi-material structures. Data-driven solutions to in-situ quality predictions have been enhanced through innovations in process monitoring [13], as well as prediction pipelines [14]. The information gathered from process monitoring has been utilised in different downstream analytics tasks [15]. The algorithmic maturity of these techniques ranges from basic statistical methods to more advanced DL techniques. This transition is guided by the capacity of the DL methods to perform the complex prediction tasks, as well as to handle large process datasets. The AM community has justified the use of DL due to the lack of existing physical and empirical models for complex process phenomena [16,17]. For the datasets collected with optical monitoring techniques, the focus has been on the development of sophisticated spatio-temporal learning techniques [18–20]. The progress is being guided by the specific detection task or application at hand, as well as the capacity of the employed in-situ sensing technologies.

In the case of Ni-WC based overlays, the microstructural defects ensuing from the lack of process understanding with varying DED process parameters represent the key challenge. The constituents in the microstructure of the deposited MMCs determine their properties and hence play a key role in their performance [20]. Specifically, a balanced distribution of the nickel matrix and the carbide particles in a defect-free microstructure is required when depositing and repairing/redepositing on existing tools and parts. However, anomalies related to the matrix and reinforcement particles, as well as their interactions in the reinforcement area and with the base material lead to defects that impact this balance and limit the potential of wear-resistant overlays [21]. The ability to detect these defects using advanced monitoring techniques and ML can inform and accelerate process optimisation, whilst minimising the need for destructive evaluation after deposition. These defects can occur simultaneously, which makes it challenging to accurately predict them across a wide range of key process parameters. Moreover, under similar processing conditions, the extent of these anomalies can vary across different defect categories, requiring tailored solutions to detect the different extents of multiple co-existing anomalies.

A brief comparison of defect detection in DED AM through vision-based monitoring and ML is presented in Table 1. Six out of ten identified works focused on porosity detection through different co-axial and off-axial vision setups. The melt pool visual and thermal features were captured by digital and infrared cameras respectively. The prevalent detection approach was found to be based on convolutional neural network (CNN) and its ensemble with classical ML or domain informed techniques. Datasets, ML/non-ML features and model predictions were fused to enhance the performance. A single MWIR camera was deployed with other sensors at best and the number of detected anomalies were limited to two or three microscale defects. The work of Guo et al. [22–24] focused on detecting the extent of porosities by integrating simulated predictions and monitoring data. Overall, these works are restricted in using convolutional DL features with limited explainability and for

Table 1. Comparison of ML based defect detection at micro-level through visual data in laser DED (data from [8]).

Reference	Monitoring setup	Micro-scale detection	ML approach	ML inputs/ Fusion(type)	Year
Yin et al. [25]	Co-axial CCD camera	Local (porosity) detection	Multi-branch Fusion CNN	Melt pool images/ Yes(model)	2023
Kim et al. [26]	IR thermal camera (co-axial)	Layer height, surface unevenness estimation	ANN	Melt pool temperature, geometry/ No(NA)	2023
Xie et al. [27]	IR thermal camera (off-axis)	Location-dependent Mechanical properties	CNN (ResNet)	Thermal history data features/ No(NA)	2021
Knaak et al. [28]	MWIR and NIR cameras (co-axial)	Lack of Fusion and Lack of Penetration etc.	Ensemble CNN-GRU	Melt pool images/ Yes (feature and decision)	2021
Gonzalez-Val et al. [29]	MWIR camera (co-axial)	Dilution estimation, Lack of Fusion pore identification	CNN (ResNet)	Melt pool images / No(NA)	2020
Zhang et al. [30]	High-speed digital camera (co-axial)	Porosity detection	CNN	Melt pool images/ No(NA)	2018
Khanzadeh et al. [31]	Infrared camera (off -axis)	Porosity prediction	KNN, SVM, DT, SOM etc.	Melt pool thermal images/ No(NA)	2018
Chen et al. [13]	Microphone, SWIR camera (off-axis), co-axial CCD camera	Crack and keyhole pore detection	Classic ML	Acoustic, co-axial melt pool, SWIR image features/ Yes(sensor input)	2023
Jamnikar et al. [32,33]	Pyrometer, off-axis melt pool camera	Microstructure and bead geometry prediction	CNN	Melt pool images, temperature features/ Yes(features)	2022
Guo et al. [22–24]	Co-axial pyrometer, off-axis infrared camera	In-situ porosity and pore size prediction	Physics-informed CNN	Pyrometer, infrared thermal images/ Yes(monitoring and simulation)	2020

Notes: ANN: artificial neural network, KNN: k-nearest neighbours, SVM: support vector machines, DT: decision trees, SOM: self-organising maps, MWIR: mid-wave infrared, NIR: near infrared, SWIR: short-wave infrared, NA: not applicable.

fewer detection targets as opposed to anomalies in DED-processed Ni-WC MMC. Coexistence of matrix and reinforcement anomalies are unique to Ni-WC MMCs and detecting their extent requires careful selection of the monitoring features across each defect category.

For ML-assisted deposition of Ni-WC MMCs, this work contributes to:

- A multi-defect and multi-level monitoring dataset
- A data-efficient and small-scale video vision transformer for process-aware feature learning
- An evaluation of the complementary potential of two MWIR cameras (co-axial and off-axial) for microstructure defect detection
- An approach for explainable selection of the most relevant features to minimise noise and redundancy
- A pipeline to detect the extent of six co-existing anomalies
- A multi-camera fusion strategy for robust in-situ defect detection under changing process dynamics

The rest of this paper is arranged as follows. Section 2 introduces the overcall framework including the experiments, feature learning, feature selection, model selection and feature fusion steps and describes the specific methods implemented in the paper. Section 3 presents the results and associated discussions. Section 4 concludes this work by highlighting the main findings, limitations as well as outlining the future research directions.

2. Overall framework and methodology

The proposed framework for ML-based detection of co-existing anomalies in additively manufactured (AMed) MMCs is divided into five phases. These phases are explained in detail below and highlighted in Figure 1.

(1) Process monitoring dataset for machine learning: In the first phase, a process monitoring dataset during the deposition of single-track Ni-WC MMC is collected for training DL models. As a prerequisite, a multi-camera experimental setup is employed in an industrial setting, incorporating both co-axial and off-axial MWIR cameras. A baseline setup with a visible light camera is employed to compare two types of optical monitoring features. Extensive variations of the process parameters are used to deposit eighty different samples. The deposited samples are evaluated to extract the ground truth values from optical metallography and identify the presence of six co-existing anomalies specific to Ni-WC MMCs. The key activities within this step are listed below:
   - *Multi-camera experimental setup*
   - *Process and material parameters*
   - *Sample evaluation*
   - *Co-existing Ni-WC MMC anomalies*

(2) Process-aware deep feature learning: In the second phase, a data-efficient video transformer is fine-tuned to learn process-specific features for subsequent classification of the co-existing anomalies and their extent. Deploying DL at process classification is more appropriate, as it ensures a sufficiently large and balanced dataset. In contrast, defect-specific or defect-extent specific classification drastically reduces the number of samples per class, leading to much smaller imbalanced datasets that are difficult to augment effectively. Therefore, a clip-based vision transformer is fine-tuned to

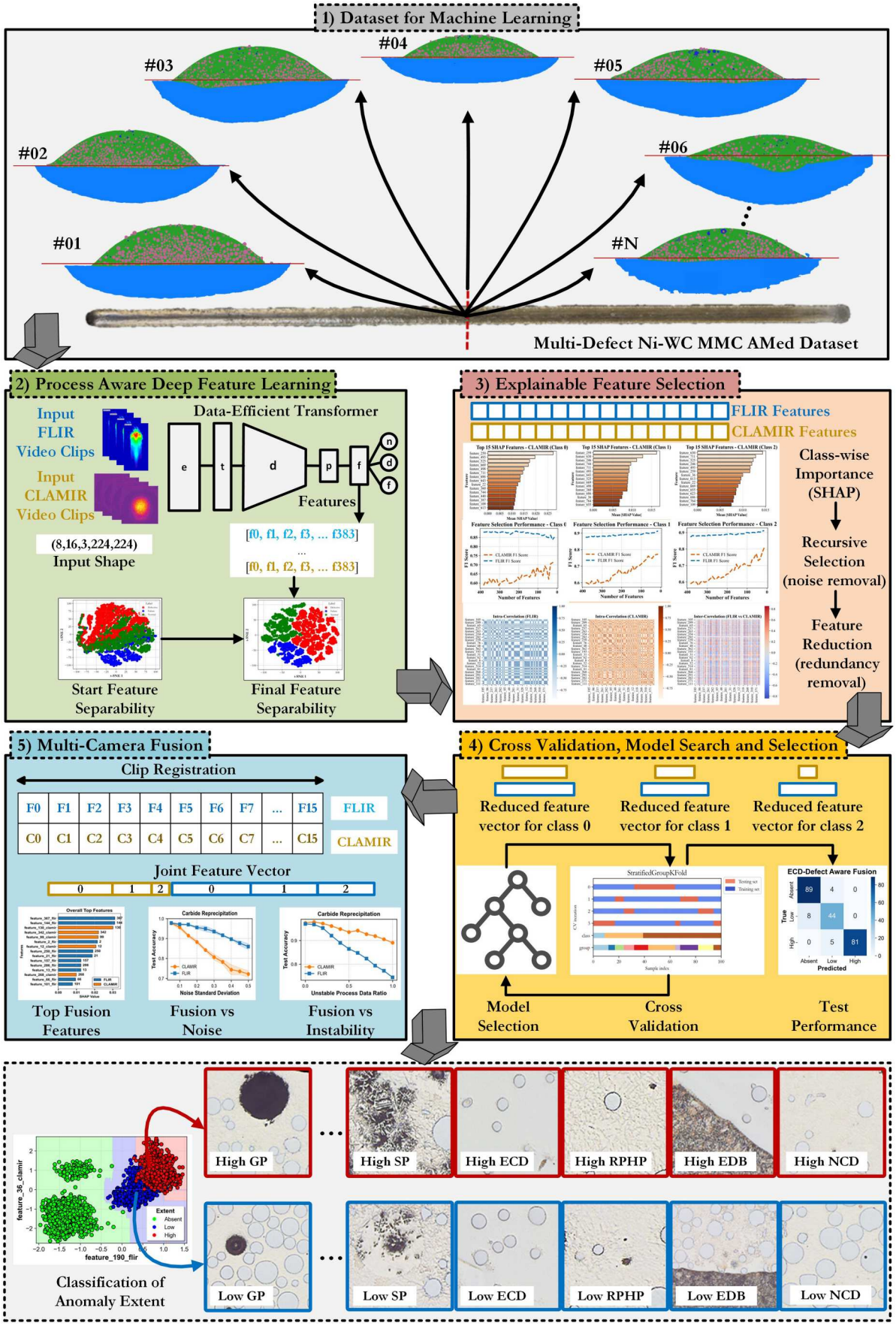

Figure 1. Framework for selecting and fusing multi-defect and multi-camera DL features for detecting extent of co-existing anomalies in AMed Ni-WC MMC.

exploit the spatiotemporal patterns specific to different processing parameters associated with normal, defective, and failure process states. The fine-tuned models are used to extract learned feature representations for subsequent selection and fusion. The fine-tuned features are also compared with the pre-trained image-based features (convolutional (ResNet CNN [34]) and self-attention models (DeiT [35])), and their separability is evaluated using cluster quality metrics. The key steps of the methodology at this phase include:

- *Data efficient video transformer-based feature learning*
- *Feature extraction and defect dataset construction*
- *Comparisons for feature separability*

(3) Explainable feature selection for anomaly extent classification: In the third phase, the trained vision transformer models of each camera dataset are used to extract the features from the video clips. The extracted features are arranged to construct anomaly specific datasets to classify their extent into three levels represented by absent, low, or high classes. Explainable artificial intelligence (AI) based feature importance scores are used to rank features relative to each extent class, leading to three different sets of ranked features. Recursive elimination is carried out to select the best performing features for each class. The selected features for each class are merged into a single feature set. To refine the feature set for each anomaly dataset, redundant and highly correlated features are removed, resulting in final datasets with minimal noise and redundancy. The key methodology steps under this phase are listed below:

- Explainable AI based ranking of feature importance
- Class-wise feature selection to minimise noise
- Feature elimination to remove redundancy

(4) Cross validation, hyperparameter search, and model selection: In the fourth phase, reduced feature datasets corresponding to six anomalies are used in four different classical ML models to detect the extent of each anomaly. An extensive range of key hyperparameter values is explored to identify the optimal candidates for each camera feature set corresponding to the six anomalies. To avoid data leakage from the train set to the validation set, a stratified GroupKFold cross-validation approach is used during the selection of hyperparameters. The best set of parameters for each dataset are selected for subsequent evaluation and usage. The key methodology steps under this phase are listed below:

- Candidate classifiers
- Cross-validation and hyperparameter search
- Model selection

(5) Explainable fusion of multi-camera features: The selected models are evaluated for stable, unstable and perturbed process conditions. This step provides insights specific to the deployed co-axial and off-axial MWIR cameras. The impact of multi-camera fusion to enhance and complement the performance of individual camera models is evaluated. In this regard, direct feature fusion is compared with decision-level and confidence-based fusion techniques. The impact of the features on detecting difficult-to-separate higher extent of anomalies is analysed before and after fusion, leading to insights on the strengths and complementarity of individual camera features. The key methodology steps under this phase are listed below:

- Impact of the process dynamics and environmental noise on camera features
- Multi-camera fusion strategies (feature-based, decision-based, confidence-based)
- Explaining the impact of camera-specific features on fusion

The main steps of the methodology to implement the proposed framework are discussed in the following subsections.

2.1. Experiments

The primary setup (Figure 2 top setup) for DED of Ni-WC MMCs in a production environment was customised with a 6-kW fibre laser, a twin-disk powder feeder, and a modified milling platform for substrate handling. The setup was also integrated with a co-axial CLAMIR camera from New Infrared Technologies [36] and an off-axial A6751sc FLIR camera from Teledyne Technologies Inc. to monitor the melt pool and build interactions [37] during the deposition of the Ni-WC MMC overlays. Before running the experiments, the laser was calibrated with a beam profiler to achieve the desirable spot size (e.g. 5.5 mm) for wear resistant overlay applications of Ni-WC MMCs, and the powder feed rate was calibrated with a tolerance of 1 g/min by adjusting the rotational speed of the hopper. Once the system was calibrated, it was prepared for executing the experiments. Before mounting the steel substrate (base for the DED experiments), these base plates were milled and scribed with reference lines along the deposition direction placed at 0, 25%, 50%, 75%, and 100% of the deposition length (e.g. 200 mm). Following the markings, test base plates were cleaned with acetone,

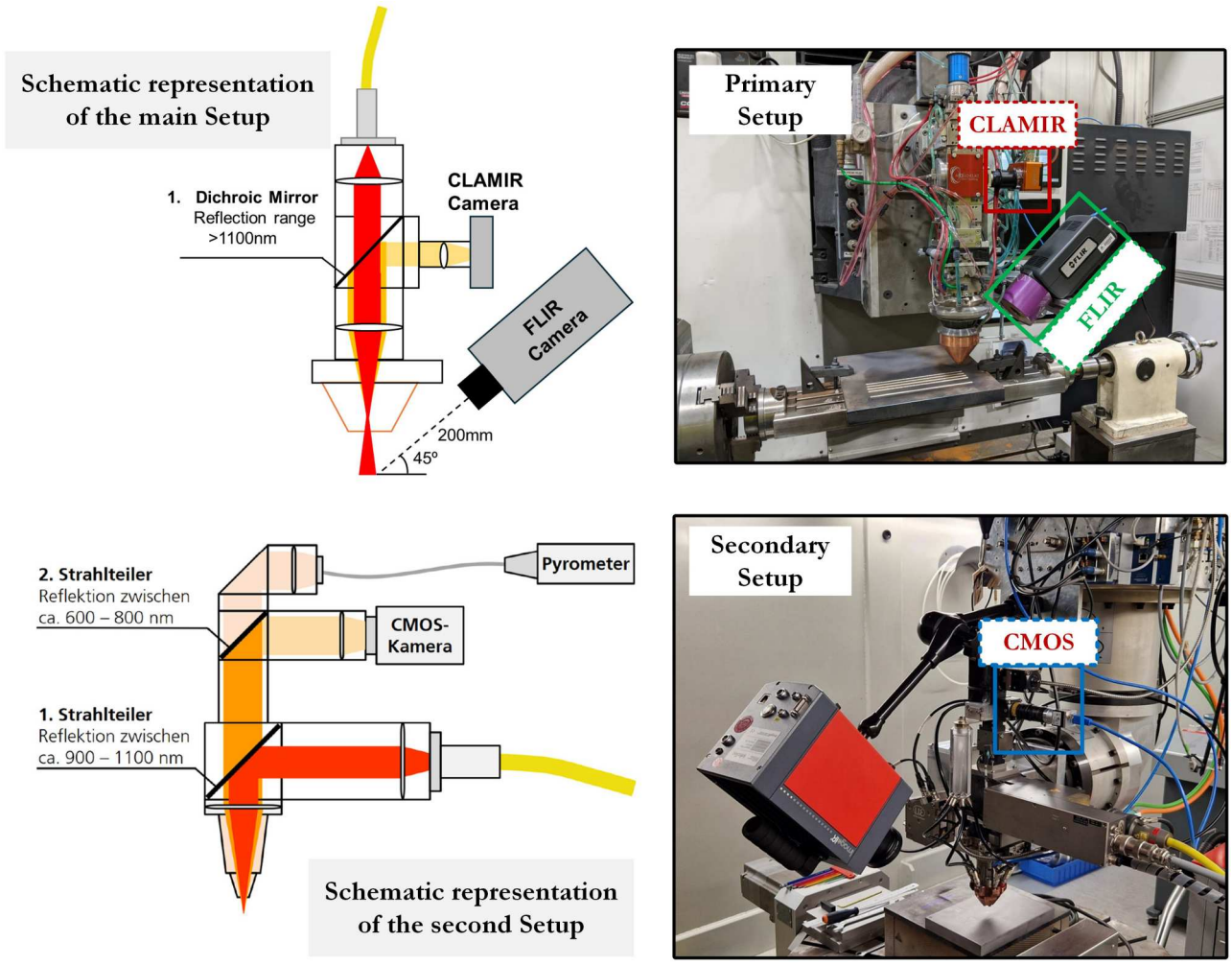

Figure 2. Schematic illustrations and experimental photos of manufacturing and integrated monitoring systems employed in this research at primary setup (top) and secondary setup (bottom).

preheated with a propane torch to > 260°C and then allowed to cool down to ~260°C to achieve better deposition quality. A wire wheel was used to remove the oxides formed on the surface of the plates from the preheating process. The deposition head was then moved to the start position and the desired nozzle spacing was achieved. Finally, data acquisition and powder flow were triggered before running the deposition program for each bead. Data acquisition was stopped after the program completed the bead deposition. The setup also had access to an off-line laser line scanner to measure the geometry of the deposited samples.

The experimental process and material parameters used to deposit samples for the dataset generation are presented in Table 2. An extensive range of laser power for DED AM of the MMCs was explored, varying between 1 and 5 kW, alongside variations of the traverse speed and powder feed rate. To ensure that the test conditions were consistent throughout the experiments, the midpoint of the whole test matrix (e.g. 2800 W, 1524 mm/min, and 80 g/min) was selected to run at the start, in the middle, and at the end of the experiments. The powder feed material used in this analysis was a mixture of cast spherical fused tungsten carbide and a proprietary Ni-Cr-B-Si blend of metals comprising the matrix in the deposited overlay. The chemistry of the tungsten carbide reported by the powder supplier was 3.8 wt.% C and the balance W. The two component powders were mixed together in a roughly 60%–40% weight fraction. The size range, hardness range, measured weight fractions, and densities are listed in Table 2. The substrate/base material was an ASTM A36 hot rolled structural steel plate with the composition shown in Table 3. Two plates of 305 mm by 305 mm by 25 mm (12 in. by 12 in. by 1 in.) were utilised for the experiments with the primary setup. Plates were milled to these dimensions from oversized plate feedstock to obtain a uniform machined surface finish prior to glass beading. Similar material and procedures were

Table 2. Experimental parameters for the process and material.

Experimental Parameters	Values
Power Levels (W)	1000, 1600, 2200, 2800, 3400, 4000, 4600, 5000
Travel Speed Levels (mm/min)	1524, 2286
Powder Feed Rate Levels (g/min)	40,80
Laser Beam Diameter (mm)	5.5
Laser Beam Profile	Gaussian
Deposition Length (mm)	200
Carbide Powder Size Range (µm)	45–106
Carbide Powder Density (kg/m^3)	16,896
Carbide Powder Weight Fraction (wt. %)	62.60
Carbide Powder Hardness Range After Deposition (HVN)	2700–3500
Metal Powder Size Range (µm)	53–150
Metal Powder Density (kg/m^3)	8100
Metal Powder Weight Fraction (wt. %)	37.40
Metal Powder Hardness Range After Deposition (HVN)	425
Powder Feed Materials	Cast spherical fused tungsten carbide with proprietary Ni–Cr–b–Si
Substrate Material	ASTM A36 hot rolled structural steel
Substrate Dimensions (width mm x length mm x height mm)	305 by 305 by 25
Metal-Carbide Powder Fractions	~ 40–60
Energy Density (J/mm^2)	4.77–35.79
Type of Defects	Gas Porosity, Shrinkage Porosity, Excessive Dilution with the Base, Non-Uniform Carbide Distribution, Excessive Carbide Dissolution, Reprecipitated Hard Phases

utilised for the experiments executed on the second setup equipped with the CMOS camera.

The secondary setup in an R&D environment with different hardware specifications was configured to match the deposition conditions of the primary setup (production environment). The 5-axis system was equipped with a 5-kW disc laser and a BEO-D70 optic with motorised collimation, enabling laser spot diameters of up to 6 mm. Integrated into the optical beam path was a Basler 1420–220 um CMOS camera for image-based monitoring of the melt pool, as well as a ratio pyrometer for temperature measurement of the melt pool (Figure 2 bottom setup). Lateral process monitoring was achieved using an InfraTec ImageIR 7350 S (Figure 2 bottom setup). For track measurement, a laser line scanner from Micro-Epsilon Messtechnik GmbH & Co. KG was integrated alongside the optics, allowing for data acquisition within the system. The powder was delivered to the nozzle through a disc feeder, and the substrate was preheated to the required temperature using a heated plate.

In this study, optical process videos from the three camera systems were used as the main sensing modality, since these represent the most effective information to capture the process phenomena for ML applications in metal AM [8]. The first two cameras, CLAMIR and FLIR, were available on the primary system. CLAMIR – a MWIR camera – focused closely on melt pool body dynamics at a high frame rate (1000 Hz), but with a low resolution (64 by 64 pixels) and was developed to be used in closed-loop adaptive control applications for laser DED processes. The FLIR camera captured the laser-powder interactions in addition to the melt pool and its surroundings. The FLIR camera operated in mid IR region, and the frames were captured at a lower rate (400 Hz), but with significantly higher resolution (640 by 512 pixels) compared to CLAMIR. The CLAMIR camera, originally developed for closed-loop laser power control in DED systems, offers high temporal resolution and is optimised for capturing precise intensity variations in thermal emission signals despite its limited spatial resolution (64 × 64 pixels) [36]. In contrast, the FLIR A6750sc thermal camera (640 × 512 pixels, 400 Hz) is widely used in thermal monitoring applications due to its high spatial fidelity and is well suited for capturing fine-grained spatial features of the melt pool [38]. The complementary use of these two sensors enables a balanced configuration that captures both the rapid transient dynamics of the melt pool and its spatial morphology. While the thermal cameras do not directly image microstructural features, the recorded thermal gradients and histories serve as representations that inform the solidification conditions, including the likelihood and distribution of carbide precipitates, re-solidification patterns, and dilution zones. These correlations are supported by established process–structure relationships in welding and additive manufacturing [39], allowing the fused thermal data to support downstream learning and inference of microstructural characteristics.

The third camera on the secondary system was based on a visible-light CMOS sensor. It was co-axially aligned with the system and captured the melt pool geometric features alongside additional process dynamics. This type of camera was expected to face challenges when monitoring a high-temperature application but was added in the experiments as a baseline to evaluate and compare its potential. The CMOS camera captured the process at 200 Hz with the highest resolution of 1200 by 1200 pixels among the experiments. For thermal camera systems, the raw intensity values were

Table 3. Chemical composition (in wt.%) of the A36 steel plate used as substrate in the experiments.

C	Si	Mn	P	S	Cr	Ni	Cu	Nb	Ti	V	Al	Fe
0.16	0.19	1.16	0.016	0.006	0.02	0.01	0.022	0.03	0.001	0.001	0.031	Balance

directly recorded and used in this study. In order to minimise the loss of information that might occur in video conversion, the raw values recorded by each pixel were extracted from the camera files in their proprietary format (.dat for CLAMIR and .ats for FLIR) [40]. Both the CLAMIR and FLIR MWIR cameras recorded intensity as arrays of 16-bit unsigned integers. Once the data was read, it was normalised by recording the lowest and highest raw values across all experiment videos as in Equation (1).

$$Image_{normal} = \left(\left(\frac{Image_{array} - min\ (Golabl)}{max\ (Global) - min\ (Global)}\right) \times 255\right) \quad (1)$$

Subsequently, these were converted and stored as an image (i.e. lossy JPEG or lossless TIFF).

2.2. Co-existing Ni-WC MMC anomalies

Ni-WC single beads were deposited on the steel substrate(s) in lengths of 200 mm each. For multi-track or multi-layer builds, relatively significant process variations are expected across the deposition for the builds with complex geometries. However, for straight single bead depositions with fixed process parameters, it is usually not the case and domain knowledge suggests that representative conditions are achieved rather rapidly and persist for the majority regions of the beads. As a result, the microstructure representative of the processing parameters was extracted from the middle region of the deposited single tracks.

For metallographic analysis of the deposited beads, each single track was sectioned transverse to the travel/deposition direction at the mid-region. The metallographic cross-sections were then mounted in Bakelite and manually ground successively with resin-bonded diamond abrasive pads (80, 180, 220, 500 grit) to produce a flat surface using a Struers LaboPol-2 grinding and polishing machine equipped with a LaboDoser. After grinding, the metallographic samples were polished to a mirror finish using the same Struers machine in progressively finer steps using diamond slurry (9, 6, 3 µm) followed by a finishing step with a 0.04 µm alumina suspension. The heat affected zone (HAZ) resulting in the substrate from each deposited bead was revealed by immersing for 5 s in a Nital etch solution (3 ml nitric acid in ethanol) followed by rinsing and cleaning with high purity ethanol. After preparation, the metallographic samples were examined under a Keyence VHX-7000 microscope at 50X magnification. The Keyence software was then used to automatically stitch the images together, creating a large panorama of the full cross-section with a resolution of 0.20 µm/pixel.

The images were then further post-processed in Adobe Photoshop to improve the image quality and consistency. The procedure was as follows: straightening the image relative to the fusion line between the bead and the base/substrate material, cropping to an appropriate size to include bead reinforcement area alongside the substrate, masking the Bakelite using a black background, increasing the contrast to improve the visual difference between the tungsten carbides and the surrounding nickel-based metal matrix, applying an 'Unsharp Mask' filter, and placing the scale bar in the corner of the image. The processed images were then segmented using a DL-based semantic segmentation tool MicroSegQ+ developed for quantitative metallography [20]. The evaluation sequence was repeated for all samples printed during the experiments.

An example cross-section where the presence of these co-existing defects is highlighted using colours (Figure 3(a)) and zoomed in crops (Figure 3(b)). Six microstructural defects have been identified as relevant to this analysis for their propensity to reduce the wear resistance of Ni-WC depositions. These defects are: (1) gas porosity (GP), (2) shrinkage porosity (SP), (3) excessive dilution with base material (EDB), (4) non-uniform carbide distribution (NCD), (5) excessive carbide dissolution (ECD), and (6) re-precipitation of hard phases (RPHP). The GP defect can evolve from powder exposure to high process heat and/or poor inert shielding gas protection of the molten pool. Gas pores are identifiable in the cross-section from their spherical shape and size as shown in Figure 3(b) (crop 3); these are typically significantly larger than the size of the tungsten carbide reinforcing particles and can thus be differentiated from any voids formed due to lost (pulled out) carbide particles during metallurgical sample preparation. The SP defect is recognisable from its irregular shape resulting from chemical segregation of the matrix microstructure during solidification. Both forms of porosity (Defects 1 and 2) represent the absence of deposited material to resist wear and are therefore undesirable in the final deposit. The EDB defect (Defect 3) is the mixing of the substrate chemistry with that of the deposition, resulting in a blended solidified chemistry having inferior performance. The dilution in this work is quantified as the area of the deposition below the surface level of the substrate as highlighted in blue in Figure 3(a). EDB is indicative of extreme process heat, of the changes in the process heat transfer phenomena from conduction mode to keyhole mode, and of an increased influence of convective heat transfer in the molten pool. These process conditions serve to degrade the tungsten

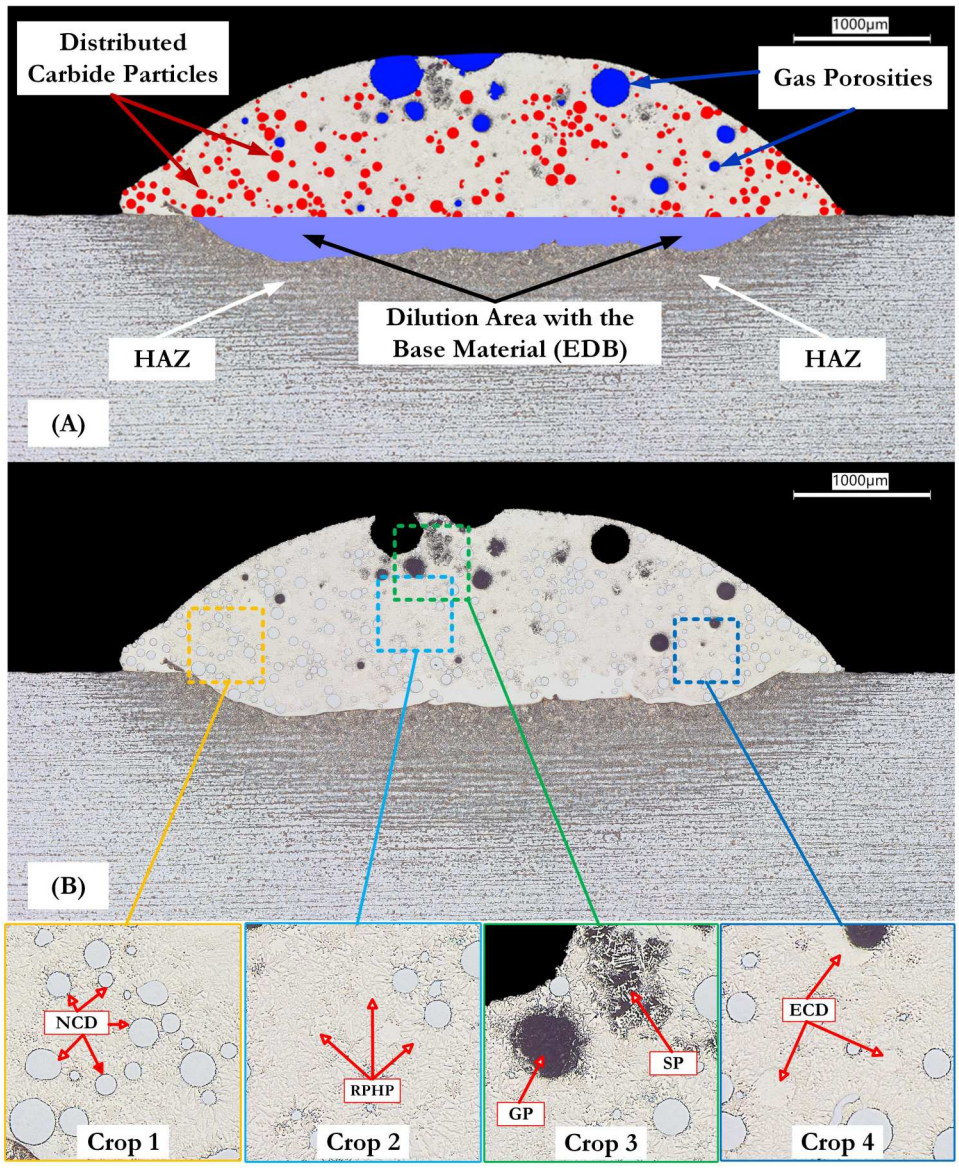

Figure 3. An example bead with microstructural defects of interest. EDB and HAZ are highlighted in (a) while the remaining anomalies are highlighted in (b). The square shapes in the lower metallograph represent 512 by 512 pixel sized crops taken at higher magnification. From left to right, these crops highlight zoomed in regions of non-uniform carbide distribution ('NCD' pointing to poor distribution in crop 1), reprecipitation of hard phases ('RPHP' pointing to texture corresponding to reprecipitated hard phases in crop 2), gas and shrinkage porosity ('GP' & 'SP' in crop 3), and excessively dissolved carbide particles ('ECD' pointing to in-place dissolved particles in crop 4).

carbide particles and must be avoided for optimal deposition performance.

The NCD defect (Defect 4) is visible as regions of the overlay devoid of tungsten carbide particles. In practice, this is observed due to settling of the higher density carbides in the molten pool (vertical variation) or significant melting of the carbides most prominently in the centre of the overlay, where the laser intensity is at its maximum (horizontal variation). Figure 3(b) (crop 1) highlights poor carbide distribution in general. This non-uniformity causes unpredictability in wear across the deposited bead. The ECD defect (Defect 5) is the dissolving of the tungsten carbides into the molten pool of the deposit. Some carbide reinforced depositions (e.g. titanium carbide-reinforced MMCs) rely on precipitation of fine, evenly distributed hard phases during solidification to achieve their performance, but for Ni-WC depositions, it is critical that the carbides remain undissolved during the deposition process to maintain their integrity and corresponding high abrasion resistance properties. Carbide dissolution is most commonly observed as partially dissolved, blocky precipitates

(Figure 3(b) crop 4) on the surface of the tungsten carbides combined with reprecipitated carbides or RPHP (Defect 6), which appear as fine texturing (Figure 3(b) crop 2) in the matrix, within local areas or, in extreme conditions, throughout the entire deposition cross-section. Similar to dilution with the base material, the melting and re-solidification of the Ni-Cr-B-Si matrix with significant amounts of tungsten and carbon leads to substantial structural changes in the morphology of the carbides and, most notably a significant decrease in the abrasion resistance, as well as the impact toughness. The tungsten carbides are the primary wear resistant constituent of the MMC; thus all thermal damage to these carbides must be minimised as the area fraction of undamaged tungsten carbide in the deposit is the largest factor in predicting relative performance of the deposition. It is worth noting that under the highest heat input conditions, re-precipitated carbides disappear in the cross-section and fully dissolve into the matrix that exhibits a uniform appearance in conjunction with elevated levels of dissolved iron from excessive dilution.

All of the defects identified in this analysis and depicted in Figure 3 are related in some capacity to high thermal conditions of the laser DED process induced from a combination of high laser power, low travel speeds, and or low powder feed rates relative to each other. The defects do not tend to appear in isolation and can often appear either all together or not at all. Despite this all or nothing observation, it has been difficult with the infinite parameter adjustment between these major process variables to identify a stable process operating window, a strong incentive for the development of a systematic ML approach to defect extent identification and corresponding parameter selection.

2.3. Machine learning datasets

The extensive Ni-WC MMC dataset generated for defect detection is divided into three versions with different microstructure labels as shown in Figure 4. The first dataset corresponds to process regimes or states, namely defective, normal, or failure. The defective state corresponds to depositions with one or more anomalies of the matrix, reinforcement, or both. The normal state represents the absence of these anomalies, leading to desired microstructures of the deposited MMCs. The failure state corresponds to cases where the process parameters fail to deposit a viable track, leading to its detachment from the substrate. The second dataset classifies the presence of six co-existing anomalies, including GP, SP, ECD, NCD, EDB, and RPHP. The third dataset corresponds to the extent of each anomaly. The extent of each anomaly is classified as either absent, low, or high based on microstructure quantification and expert labelling. The distribution of CLAMIR and FLIR frames across the classes of each dataset is also presented in Figure 4. The following steps were consistently applied to all the cross-sections to determine the extent of co-existing anomalies:

(1) quantify the presence of carbide particles, porosities, matrix, dilution area, and reprecipitated hard phases using DL-based semantic segmentation models tailored to the Ni-WC MMC dataset from optical and scanning electron microscopy
(2) apply expert industrial thresholds on anomalies reflected as relative area fractions**,** which included porosities, carbide particles, and reprecipitated phases in the reinforcement area, leading to GP, SP, ECD, and RPHP extent labels as well as dilution area below the original substrate surface, leading to EDB extent labels
(3) apply expert analysis on anomalies reflected as a distribution such as uniform or non-uniform distribution of the carbide particles, leading to NCD extent labels

2.4. Small-scale data-efficient video transformer

It has been established in data-driven process defect or anomaly detection tasks of AM that critical temporal information is lost while working with only image-based process models [19,41]. Similarly, melt pool-related process signatures, whether directly or indirectly derived, have been shown in multiple studies to effectively capture correlations with process parameters [42–44]. Moreover, the process rates and states associated with complex phenomena are best captured using a spatial–temporal structure of monitored camera data [28]. Building upon these findings of ML-assisted AM research, the present research considered state-of-the-art process video models to capture the melt pool variations and relate them to defects of interest. The video models were used to capture high-level process variations across normal, defective, and failure regimes, enabling accurate classification of video clips into these categories.

The recent success of self-attention models on long sequences of textual data has been successfully translated to spatial (images) and spatiotemporal (videos) data structures [45]. Transformer models act on the vision data by dividing it into patches and applying the attention mechanism across the patches. Each patch of pixels is flattened and linearly transformed to

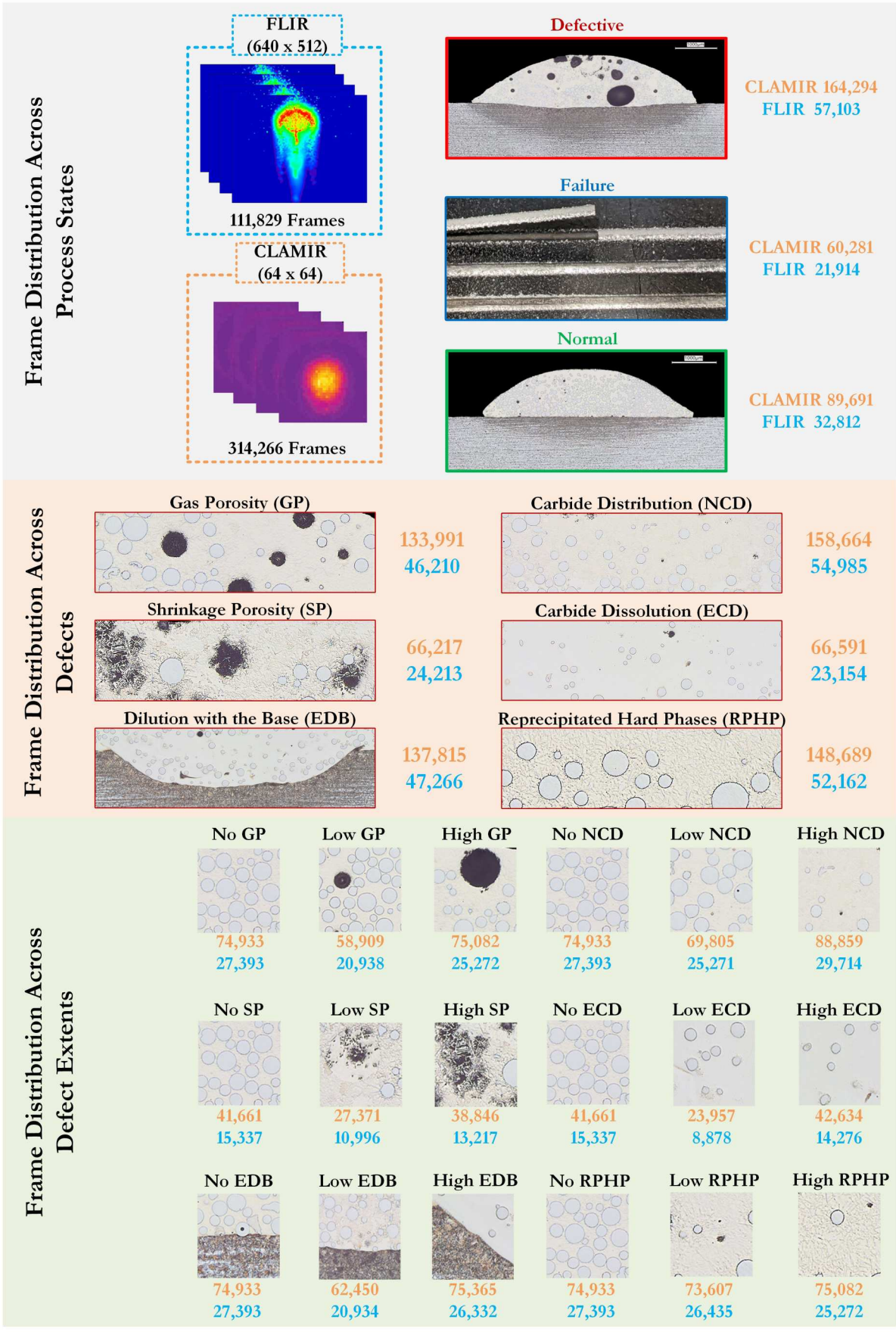

Figure 4. CLAMIR and FLIR frame distributions across different defect labels, their types, and extents from deposited Ni-WC MMC. The microstructural screenshots symbolically show the dominant defect types.

a fixed sized projection. Contrary to the convolutional models that inherit capacity to capture positional relations, the spatial (and temporal) information within and across patches is inserted through embeddings that are added to linear projections of vectorised patch pixels. The key attention module serves as the engine of a transformer architecture. It learns to transform these inputs into query, key, and value vectors and attention scores for each input vector relative to all other input vectors are calculated. The attention scores are later weighted using the value vectors. The output goes through a feed forward network (FFN) following residual connection and layer normalisation. The output of the FFN also goes through another residual connection and layer normalisation, and this pattern of self-attention, addition-normalisation, FFN, and addition-normalisation is repeated several times depending on the depth of a given transformer model. Equations (2–6) represent these operations.

$$z_i = W_p * flatten(P_i) + b_p \quad (2)$$

where z_i represents the projected vector of the i-th patch, W_p is the weight matrix for linear transformation, P_i represents a patch of pixels from the image, and b_p represents bias for linear transformation.

$$x_i = z_i + E_i \quad (3)$$

where x_i is the input vector after adding positional embedding, and E_i is the positional embedding for patch i.

$$Attention(x_i, x_j) = softmax((q_i * k_j^T) / sqrt(d_k)) * v_j \quad (4)$$

where q_i is the query vector derived from xi, k_j is the key vector derived from xj, d_k represents the dimension of the key vectors, and v_j is the value vector derived from xj.

$$FFN(y_i) = ReLU(W_1 * y_i + b_1) * W_2 + b_2 \quad (5)$$

where y_i is the output from the attention mechanism, W_1 represents the weight matrix of the first feedforward layer, b_1 represents the bias of the first feedforward layer, W_2 represents the weight matrix of the second feedforward layer, b_2 represents the bias of the second feedforward layer, and $y_i^{'}$ is the output after first residual connection and normalisation.

$$z_i^{'} = LayerNorm(y_i^{'} + FFN(y_i^{'})) \quad (6)$$

where $z_i^{'}$ is the final output after all layers in the transformer block.

Video vision transformers represent state-of-the-art empirical capabilities to learn AM process variations and relate these variations with resulting defects in the AM parts [46–48]. There are a handful of these architectures available, such as Video Vision Transformer [49], TimeSFormer [50], Video Swin Transformer [51] and VideoMAE [52]. Out of these options, the self-supervised video transformer VideoMAE was selected in the present research to extract features from the melt pool videos and evaluate their potential for fusion. The utilised video masked autoencoder transformer or VideoMAE model has been pre-trained on the Kinetics-400 video dataset [52]. The architecture of the model presented in Figure 5(a) is structured in a typical encoder-decoder fashion. The self-supervised architecture enables the learning of meaningful video representations in the absence of ground truths. Moreover, the architecture deploys a significantly high ratio (≥90%) of masking for patches in the input clip and learns to reconstruct both masked and unmasked patches in a relatively simpler decoder. The masking property may be especially useful for the AM melt pool video clips where the majority of the variations only come from a small region and the remaining regions can be masked randomly. In order to fine-tune the model on the Ni-WC MMC process dataset, and learn features specific to the three process states, a classification variant of the model was used as shown in Figure 5(b) [52]. In the modified version, the decoder was removed, and the classification layer of the model was updated to differentiate three process states. The architectural details for the small variant of VideoMAE classification model are presented in Table 4. Furthermore, the training parameters for fine-tuning the FLIR and CLAMIR camera datasets on VideoMAE-small are presented in Table 5.

2.5. Extent-specific feature selection

Explainable AI methods allow to interpret the inner workings of the ML models and their applications in advanced manufacturing are critical to support the adoption of AI-assisted tools and techniques. Shapley additive explanations (SHAP)-based importance of features is being increasingly used within advanced manufacturing as an explainable AI technique to evaluate the impact of each feature on model output [53]. In this work, SHAP was used to interpret features extracted from fine-tuned vision transformer. SHAP values are based on the cooperative game theory. It offers a method for allocating the 'reward' (i.e. the model's prediction) among a group of 'contributors' (i.e. input features) according to their individual impact. Equation 7 calculates the feature importance $\varphi_j(f)$ using SHAP. The results can provide information about the potential of monitoring systems, the distribution of experimental data, and the significance of model features for defect

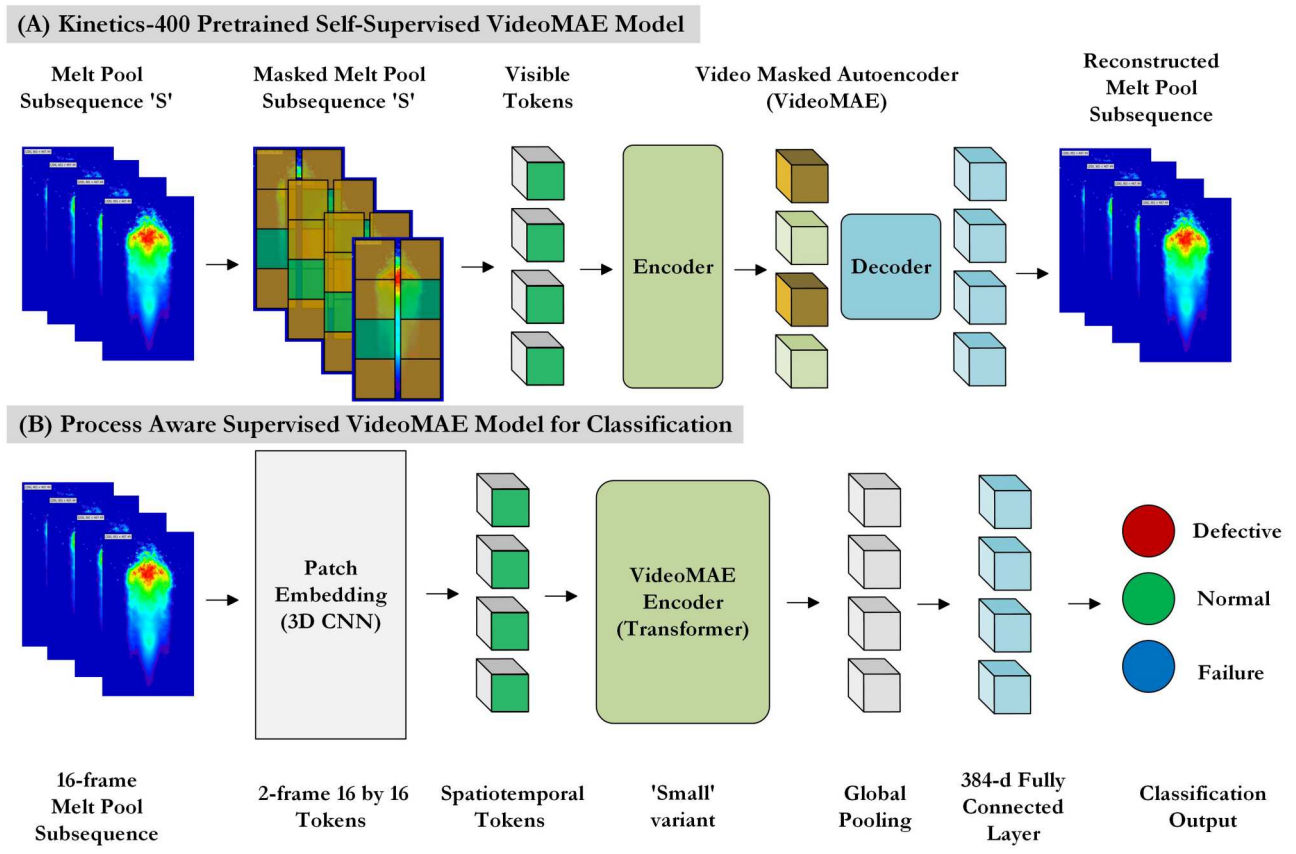

Figure 5. VideoMAE architecture for melt pool sequences. Input clips of 16 frames (224 × 224) were divided into small 3D patches (2-frame × 16 × 16), masked at ≥90%, and passed through the encoder. Spatiotemporal tokens were pooled and processed through a 384-dimensional fully connected layer for classification into process categories (e.g. defective, normal, failure).

detection tasks. SHAP-based rankings of feature importance can also provide insights into the effect of camera fusion and inform the selection of the most important features for detecting specific defects.

$$\varphi_j(f) = \sum_{S \subseteq \{1,\ldots,M\}\setminus\{j\}} \frac{|S|!(M-|S|-1)!}{M!}[f(S\cup\{j\}) - f(S)] \quad (7)$$

where M is the total number of features, S is a subset of features not including j, $f(S)$ is the model's prediction given the features in subset S, $f(S\cup(j))$ is the model's prediction given the features in subset S plus feature j, and $|S|$ is the number of features in subset S.

Table 4. Architectural configuration of VideoMAE model (small) for classification.

Stage	Component	Details
Input		3 Channels, 16 Frames, 224 × 224 per frame
Embeddings	Conv3d Layer	kernel_size = (2, 16, 16), stride = (2, 16, 16), out_channels = 384
Encoder	12 x VideoMAELayer	Each layer includes:
	• Self-Attention	Query, Key, Value: Linear (384, 384)
	• Self-Output	Dense (384, 384)
	• Intermediate	Dense (384, 1536), GELU
	• Output	Dense (1536, 384),
	• Layer Norms (before and after)	LayerNorm ((384,))
Output Normalisation	LayerNorm	LayerNorm ((384,), eps = 1e-05)
Classifier	Linear	Linear (384, 3)

Table 5. Training parameters for fine-tuning VideoMAE-small model on the CLAMIR and FLIR datasets.

Parameter	Values
Architecture variant	Small
Parameters / Weights	22,000,000
Learning rate range	5e-05 – 5e-06
Train and test batch size	8
Optimiser	Adam with betas = (0.9, 0.999) and epsilon = 1e-08
Learning rate scheduler	Linear
Number of fine-tuning steps	2000
Input shape	8 × 16 × 3 × 224 × 224
Output shape	1 × 3 (Normal, Failure or Defective)
Pre-trained (dataset)	Yes (Kinetics-400 weights)
Train/Validation split	80/20

The MWIR CLAMIR- and FLIR-based model checkpoints representing lowest validations losses were used to extract fine-tuned features for subsequent classification of defect extents. The direct use of an extent-level dataset in the video transformer model was found to be incompatible with size requirements of these models, as indicated by previous AM-based studies [47–50]. As a result, the extracted features were used to construct six datasets, each representing one of the anomalies under consideration. Each dataset has three classes representing extent-free (absence), low and high extent of the respective defect. Table 6 presents the distribution of the defect extent datasets derived from fine-tuned VideoMAE models using CLAMIR and FLIR input video clips.

The feature datasets from the six anomalies were subjected to SHAP-based reordering leading to three different rankings, each corresponding to feature importance relative to a single class (absent, low or high). The SHAP importance scores are superior to model-specific feature rankings (e.g. Gini, entropy) as these can provide model-independent insights. Therefore, the model was treated as a black box with inputs and corresponding outputs observed to explain its predictions. The features were iteratively removed from each ranked set, starting from the least important feature to eliminate noise. This approach helped to remove noise while preserving predictive power. To maintain consistency, a baseline tree model (with constant default parameters) was used alongside a stratified GroupKFold ($k = 4$) cross validation approach to evaluate the impact of feature elimination on performance. A total of 36 rounds of feature selection were carried out (3 classes, 6 anomalies, 2 datasets) and the value of $k = 4$ balanced the computational cost with robustness. As tree-based models are relatively robust to irrelevant features, they provided a strong baseline for feature importance evaluation. Stratification ensured similar class distribution in each fold, whereas the group-based segregation limited the data leakage between the training and test folds.

Table 6. Distribution of training samples across defect extent classes (extent-free, low-extent, and high-extent) for each anomaly based on features extracted from fine-tuned CLAMIR and FLIR models. A separate 10% test split from the total dataset was created for both stable and unstable conditions.

		Defect Extent Classes		
Defects	Total Samples	Extent-free Class	Low-extent Class	High-extent Class
GP	4042	1541	1420	1081
SP	2128	863	743	522
EDB	4167	1541	1324	1302
NCD	4042	1541	1361	1140
ECD	2165	863	802	500
RPHP	4351	1541	1420	1390

The elimination process was stopped at a fixed minimum feature set (≥15 features) for each class to prevent over-reduction that could degrade model performance. Ensuring a minimum number of features maintains some level of redundancy, which is beneficial in real-world manufacturing datasets with the potential of environmental noise and added process dynamics. Feature sets corresponding to maximum performance were recovered in each case and were merged into a single feature set, while keeping only the unique features. To eliminate the redundancy among concatenated features, the Spearman correlation coefficient was used to drop highly correlated feature pairs (≥0.85). In contrast to the Pearson correlation, which assumes linearity, the Spearman correlation captures monotonic real-world feature relationships. To accommodate the imbalance in the distribution of samples across the three classes, weighted loss was used during the feature selection process. These steps are symbolically presented for the CLAMIR- and FLIR-based EDB defect extent classes in Figure 6 while using accuracy and F1 score to record performance variations.

2.6. Cross validation and hyperparameter search for extent detectors

Several considerations were taken into account to increase the success of training and to maximise its effectiveness. The challenge to predict the extent of multiple co-existing anomalies is unique to heterogeneous materials such as MMCs. While the anomalies co-exist, their extent varies and can be quantified through optical and scanning electron microscopy. Therefore, defect-specific extent classifiers were developed by careful selection and fusion of features from MWIR cameras. Developing anomaly specific classifiers lowers the data requirements as compared to a single classifier and limits the correlations among the co-existing anomalies.

For a given set of laser power, powder feed rate and traverse speed, the associated melt pools undergo similar variations that are usually limited. As a result, mixing experiment groups (e.g. bead monitoring data) between training and validation sets can lead to data leakage between the training and testing process [54]. Data leakage can lead to unreasonably high performance that is not representative of the problem complexity. Moreover, when models are evaluated on new data from real-world production environment, the performance can significantly degrade, since training was

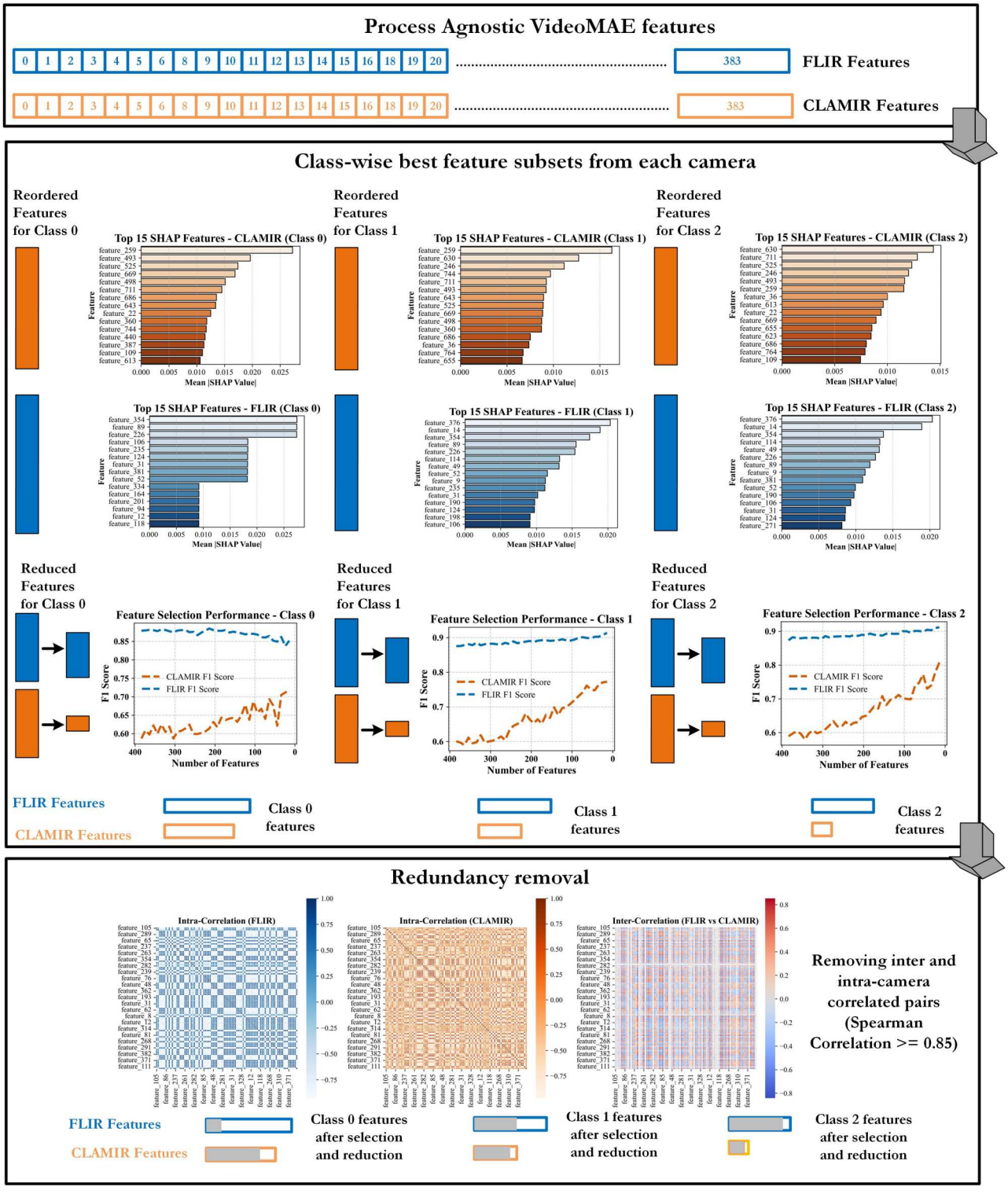

Figure 6. Explainable feature selection using SHAP-based feature importance to identify features relevant to defect severity levels. Class '0', '1', and '2' represent extent-free, low-extent, and high-extent defect levels respectively.

influenced by data leakage. Therefore, it was ensured that the test data comes from beads that did not contribute to training using stratified GroupKFold cross validation [55]. Balanced class weights were used in the loss function to offset the impact of varying class representations in the resulting training and validation folds. Two unseen test sets, each 10% of the original dataset, were used, representing stable and unstable process dynamics. The unstable test set represented variations in process dynamics at the start or end of

deposition. A third test set was generated by applying visual perturbations (rotation, contrast, brightness, and Gaussian noise) to the stable test set. The perturbations were grounded in practical considerations and were reflected by augmentations in the original video clips [19]. The remaining 80% of each dataset were used in cross validation and hyperparameter search to select the best model parameters for each of the six anomaly extent classifiers.

The range of hyperparameter values explored during the grid search approach of the scikit-learn library has been presented in Table 7. A total of four classifiers were used to detect the extent of six anomalies (GP, SP, EDB, NCD, ECD, and RPHP). The algorithms included Random Forest (RF), Support Vector Classifier (SVC), Extreme Gradient Boosting (XGBoost) and Logistic Regression (LR). These four classifiers have been widely used in AM defect detection [13,56]. The total input samples for each defect category are listed in Table 6. The cross validation was performed on each of the stratified folds, and the parameters corresponding to the top-performing fold were recorded. These parameters were used to retrain the model on the complete training dataset. While both accuracy and F1 score were used, weighted accuracy was used to select the top-performing hyperparameter sets. To ensure the representation of all classes in the validation fold, a k value of 4 was used throughout the experiments in GroupKFold.

2.7. Camera fusion for noise and process dynamics

The FLIR and CLAMIR MWIR camera systems have different characteristics (e.g. resolution, frame rate) and both observe the process in different configurations. The CLAMIR camera was co-axially focused on the molten pool, whereas FLIR observed a larger surrounding area in addition to the melt pool region in an off-axial configuration. Due to its off-axial arrangement, the FLIR camera also captured the laser-powder-pool interaction zone. Multi-camera fusion to support data-driven applications for complex builds has been suggested as a potential remedy to overcome situations where one camera system is unable to effectively monitor the critical process phenomena [8]. While the current depositions don't fall in the category of complex build, there could still be value in fusing their outputs for in-process defect detection to mitigate the impact of changing process dynamics. To use the features in the same model, a clip-wise correspondence of frames co-located in the same region was needed, since video models act on a group of several frames (e.g. 10, 16, 32). In this study, low frequency FLIR was taken as the reference and its frames were matched with the high frequency CLAMIR camera, while matching their spatial locality for fusion strategy shown in Figure 7.

The fusion process started by observing the onset of the melt pool in both the FLIR and CLAMIR videos. A time-based matching strategy was used, where time to capture a video clip (e.g. 16 frames for VideoMAE) in FLIR was utilised to estimate and extract corresponding video clips from CLAMIR. Both fair and greedy approaches to fuse the clips originating from two cameras were considered. Fair fusion retrained the same number of features from each camera whereas greedy fusion utilised all available features. A fair-fusion strategy was eventually implemented and each FLIR clip was matched with only one corresponding CLAMIR clip to balance feature representation and avoid redundant information. This choice was reinforced by experiments done in a greedy manner, leading to 2.5-fold overrepresentation for CLAMIR features in the fusion set without enhanced performance.

Table 7. Ranges of parameter values for candidate classifiers during cross-validation and hyperparameter search.

Model Type	Hyperparameters	Values Explored
Random Forest (RF)	n_estimators	50, 100, 200, 300, 500
	max_depth	5, 10, 20, 30
	min_samples_split	2, 5, 10
	min_samples_leaf	1, 2, 4
	max_features	'sqrt', 'log2'
Support Vector Classifier (SVC)	C	0.1, 1, 10, 100
	kernel	'linear', 'rbf', 'poly'
	gamma	'scale', 'auto'
XGBoost (XGB)	n_estimators	50, 100, 200, 300
	max_depth	3, 5, 7, 10
	learning_rate	0.01, 0.1, 0.2
	subsample	0.6, 0.8, 1.0
	colsample_bytree	0.6, 0.8, 1.0
Logistic Regression (LR)	C	0.01, 0.1, 1, 10, 100
	penalty	'l1', 'l2'

3. Results and discussion

This section presents the results from video-vision transformer-based model for feature generation, as well as subsequent feature selection to eliminate noise and redundancy. The resulting features are used to train shallow defect extent classifiers for co-existing anomalies through extensive hyperparameter search and careful cross-validation. The trained models and their performance are presented. The results of camera fusion through the selected features are discussed in detail to support its application to manufacturing datasets having increasing likelihood of external noise and changing process dynamics. SHAP-enabled interpretation of multi-camera features reveal the significance of features from each camera to separate the extent of

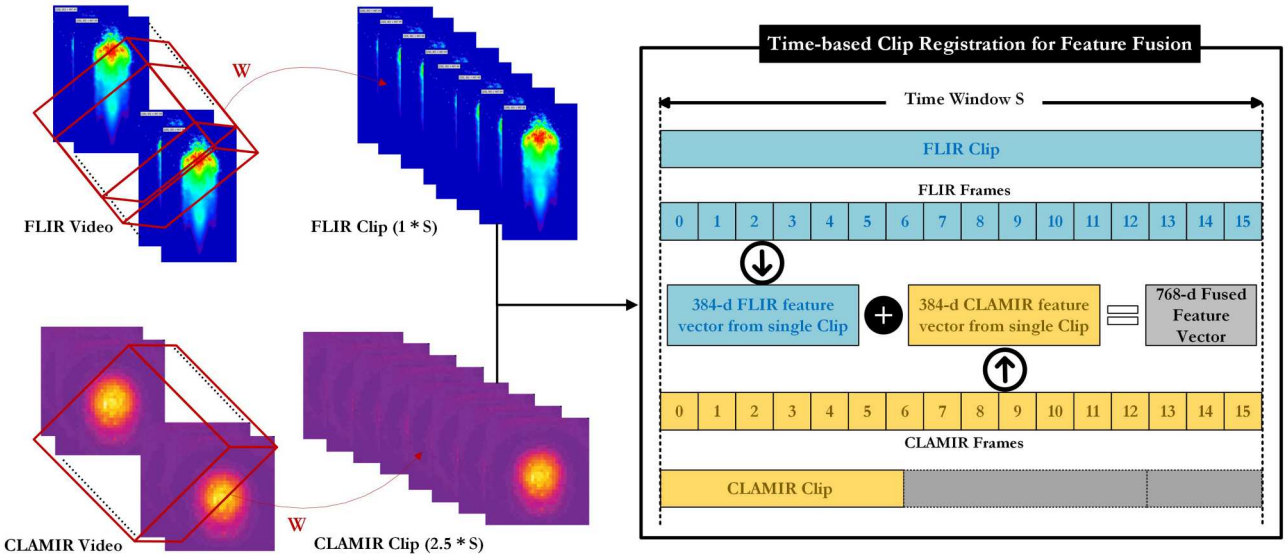

Figure 7. Frame and feature correspondence strategy for camera fusion. The strategies are inspired by the different capture frequencies of co-axial and off-axial MWIR camera systems. 'S' represents the given time window for cross-camera frame correspondence.

these co-existing anomalies unique to Ni-WC MMC. Different fusion strategies are discussed to effectively maximise the complementarity of multi-camera features across changing (stable, unstable, and perturbed) process dynamics.

3.1. Monitoring and modelling comparison for feature separability

The employed co-axial CLAMIR and off-axial FLIR cameras are compared with a co-axial visible light CMOS camera using the same material and process parameters. While the optical and thermal cameras had different resolutions and frame capture rates, the initial comparison employed default settings to conduct sensing potential analysis. Based on the results of this evaluation, the most capable cameras were found to be appropriate to capture the highly transient process. The subsequent analysis employed a time-based registration of clips from the two thermal cameras with different frame capture rates. This helped establish a correspondence between signals being captured from physical phenomena and supported the comparative analysis (e.g. distribution, correlation) for feature fusion and defect extent classification.

It has been shown in the previous research on AI-enabled defect detection in AM that mid-infrared cameras provide better relativity among changing process conditions compared to simple light-based cameras, which only capture variations in geometric or shape-based features [28]. The comparison among the three cameras studied in the present research in terms of their feature separation capacity has been presented in Figure 8. A t-distributed stochastic neighbour embedding (t-SNE) was preferred over uniform manifold approximation and projection (UMAP) for 2D feature visualisation, since it exposes both the local structure within bead features and the global interaction among different beads. The CMOS camera provided poor separation between defective and non-defective features from the pre-trained VideoMAE transformer model. This observation against the visible light camera highlighted the capacity of MWIR cameras to provide high-quality features for learning the differences amongst the wide-ranging thermal conditions. When comparing the potential of the camera systems, the CMOS-based sensor (Figure 8(a)) struggled to differentiate the three processing regimes, as expected. The FLIR and CLAMIR cameras both showed potential to separate defect features. FLIR seemed to outperform CLAMIR across the three processing regimes, but it still contained regions of overlap between the clusters.

CNNs have been the most common vision-based models for defect detection in AM [57]. Figure 9 compares the local and global feature structures between CNN and transformer models using pre-trained and fine-tuned weights corresponding to inputs from six similar processing conditions (Power: 2200 & 2800 W, Speed: 1524 & 2286 mm/min, and Feed Rate: 40 & 80 g/min). Using consistent t-SNE parameters for 2D feature visualisation, it was observed that the self-attention-based transformers provided better starting weights for knowledge transfer and fine-tuning [60]. Fine-tuning was still required to improve local feature

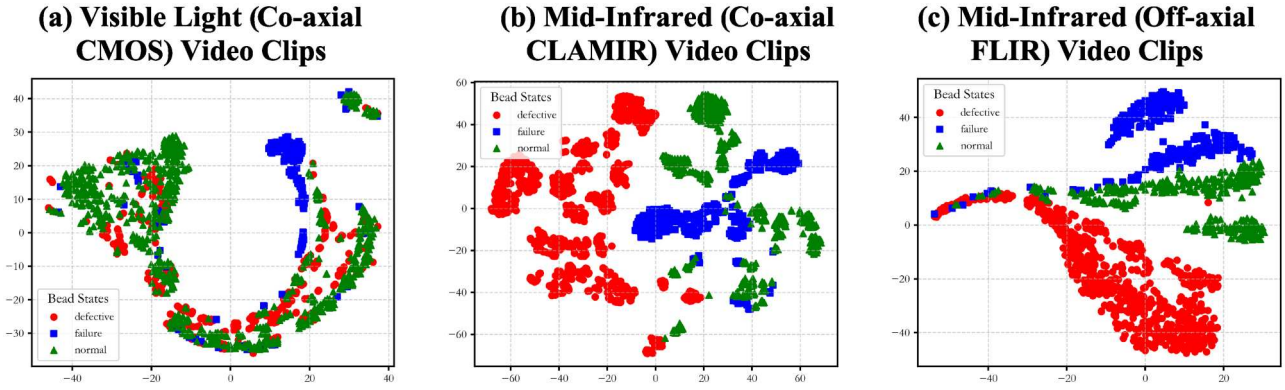

Figure 8. Melt pool data from three monitoring systems with different capture rates, resulting in varying numbers of sampled clips. Each VideoMAE input clip had a shape of (16, 3, 224, 224), where 16 denotes the clip length. (a) Visible light (Co-axial CMOS) video clips, (b) Mid-infrared (co-axial CLAMIR) video clips, (c) Mid-infrared (Off-axial FLIR) video clips.

distributions (e.g. experiment specific features), as well as global feature structure among the three process regimes (e.g. defective, normal and failure). These observations also highlight the differences between co-axial CLAMIR with smaller spatial resolution (higher frame rate) and off-axial FLIR with larger spatial resolution (lower frame rate). The CLAMIR camera provides more precise features which are consistent across parameters, as compared to FLIR that captures relatively generic features. Therefore, CLAMIR is expected to provide strong classification features, whereas FLIR has the potential to support generalisation. Following fine-tuning, both sets of features can separate across process parameters and process regimes.

Figure 10 compares CLAMIR and FLIR video clips under two different VideoMAE model parameter configurations (Kinetics-400 weights versus process-aware weights) for both experiment-wise and process-wise distributions. The experimental distribution is arranged with respect to changing process parameters whereas the process distribution highlights three major states/regimes. Each plot shows a 2D t-SNE embedding, where colour-coded clusters correspond to different beads or process conditions. Notably, the feature distributions change before and after fine-tuning under both scenarios with FLIR features leading to significantly improved separation compared to CLAMIR features. Although both the CLAMIR and FLIR clips benefit from fine-tuning, the higher spatial resolution of FLIR (640×512) allows it to capture more intricate shape and texture details than CLAMIR (64×64). As a result, the FLIR clips exhibit a more visible improvement after fine-tuning, since the model can better learn complex visual patterns. By contrast, CLAMIR's lower resolution restricts the feature richness to primarily pixel intensities, limiting the impact of fine-tuning only to minor feature rearrangement.

Cluster quality metrics (Silhouette Coefficient or SC [58], Calinski–Harabasz Index or CHI [59], Davies-Bouldin Index or DBI [60]) were utilised to measure feature separation before and after fine-tuning. Equations 8–10 show the basic formulae for these clustering metrics. The SC metric measures the cohesion and separation of clusters by calculating the average Silhouette score for each sample. The Silhouette scores range from −1 to 1, with higher values signifying better clustering quality. The CHI metric evaluates the ratio of between-cluster dispersion to within-cluster dispersion. A higher CHI score indicates better-defined clusters, with the optimal clustering having the maximum CHI value. The DBI metric computes the ratio of within-cluster distances to between-cluster distances. A lower DBI score indicates better clustering quality, with the optimal clustering having the minimum DBI value.

$$\mathrm{SC} = \frac{(b_i - a_i)}{\max(a_i, b_i)} \tag{8}$$

where a_i is the average distance from the i-th point to all other points in the same cluster and b_i is the minimum average distance from the i-th point to all points in the other clusters.

$$\mathrm{CHI} = \frac{Tr(B_k)}{Tr(W_k)}\left(\frac{(k-1)}{(n-k)}\right) \tag{9}$$

where $Tr(B_k)$ is the trace of the between-cluster dispersion matrix, $Tr(W_k)$ is the trace of the within-cluster dispersion matrix, k represents the number of clusters, and n is the total number of data points.

$$\mathrm{DBI} = \frac{1}{k}\sum_{i=1}^{k} max_{\{j \neq i\}}\left(\frac{s_i + s_j}{d_{ij}}\right) \tag{10}$$

where s_i is the average distance between each point in cluster i and the centroid of cluster i, d_{ij} is the distance

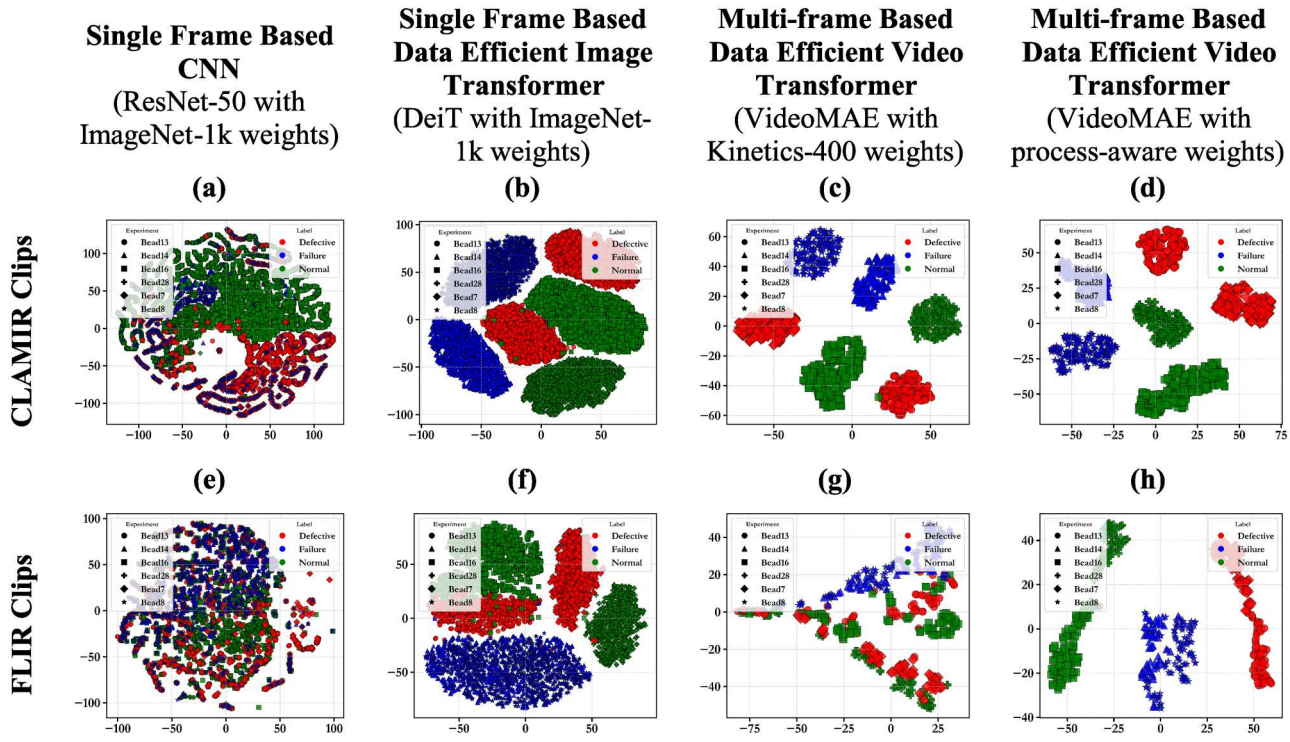

Figure 9. t-SNE comparison of CNN (a, e) and transformer-based (b, c, d, f, g, h) models across CLAMIR and FLIR clips. Transformers provided superior pre-trained features, but fine-tuning was key for learning both local and global structures. Fine-tuning on CLAMIR enabled clear separation of defect categories, while FLIR revealed structure across process states, particularly under changing conditions (blue clusters).

between the centroids of clusters *i* and *j*, and *k* represents the number of clusters.

Table 8 presents cluster quality metrics (SC, DBI, and CHI) for the CLAMIR and FLIR features extracted using VideoMAE models with Kinetics-400 pretraining and process-aware fine-tuning. Across both camera models, fine-tuning improves the SC values, indicating better-defined cluster separations. For instance, the SC for defective CLAMIR features increases from 0.0389 (Kinetics-400) to 0.2087 (process-aware), and for FLIR, from 0.1794 to 0.3542, demonstrating improved cluster cohesion. Similarly, the CHI values significantly increase after fine-tuning, with FLIR's normal features rising from 183.28 to 3580.65, reinforcing the superior cluster separability in FLIR, due to its higher spatial resolution. Meanwhile, the DBI values decrease for all cases after fine-tuning, particularly for CLAMIR's defective features (8.54 to 1.81) and FLIR's normal features (3.98 to 0.90), indicating reduced within-cluster variance and stronger feature consistency. Overall, the results highlight that fine-tuning with process-aware weights substantially enhances feature separability, with FLIR benefiting the most due to its higher spatial resolution, while CLAMIR sees moderate improvements primarily in cluster cohesion.

The plots in Figure 11 show training and validation loss curves during the fine-tuning process of the VideoMAE model. CLAMIR, with its lower spatial resolution and simpler features (melt pool zone without extended tail and pool-powder interactions), converges faster and reaches its minimum training loss earlier (step #1080) compared to FLIR (step #1170). The corresponding validation errors (0.000021 versus 0.000076) also indicate that CLAMIR stabilises more quickly. Nonetheless, both models achieve lower validation loss as compared to training loss around the 1000th step of the fine-tuning process. The plots display the first 1500 steps out of a total of 2000 fine-tuning steps to emphasise critical regions of variation. The training parameters are listed in Table 5. The two VideoMAE-small models were fine-tuned in a GPU-accelerated (NVIDIA RTX-4090 Ti) environment using PyTorch and the Hugging Face Trainer API for efficient training and evaluation. The training process was conducted on video data, where sequences of frames were extracted and processed using the built-in VideoMAEImageProcessor function to ensure consistency in input representation. Model evaluation and logging were integrated with Weights & Biases (WandB) to track performance metrics, and checkpoints were systematically saved locally and uploaded to the Hugging Face Hub for

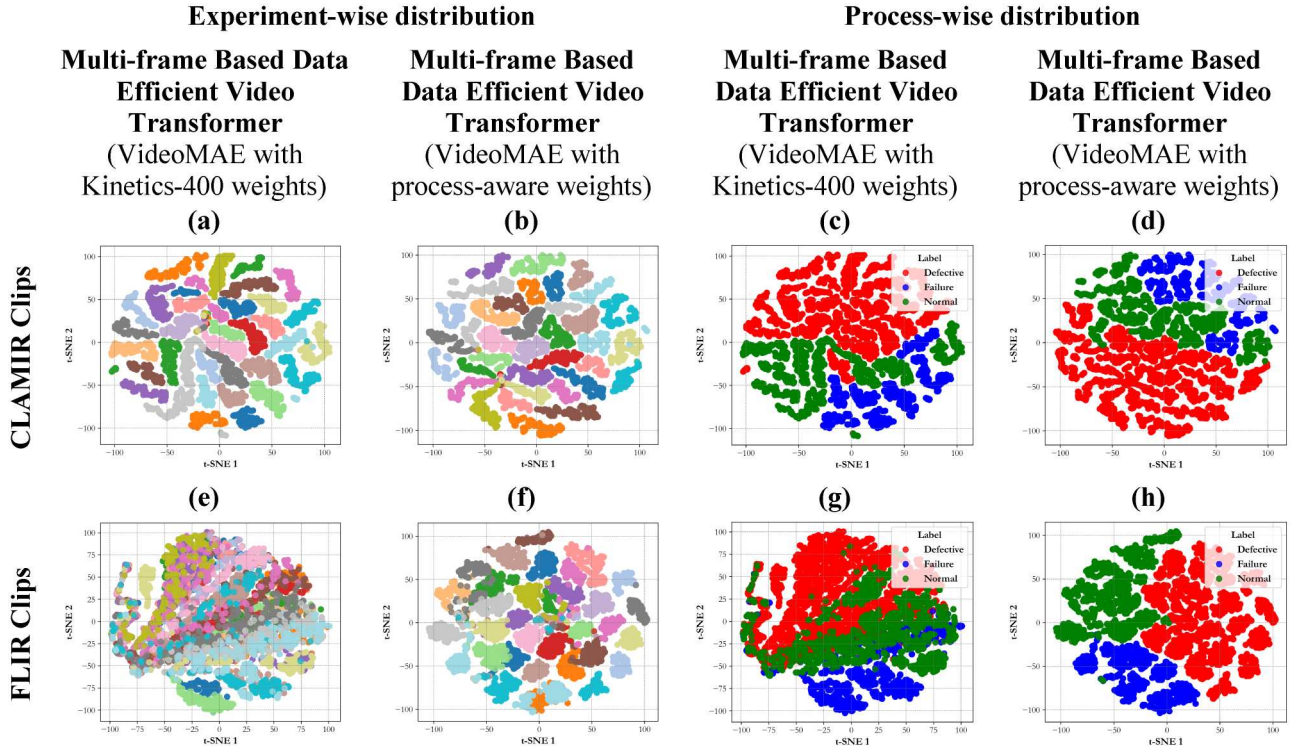

Figure 10. t-SNE plots of CLAMIR (64 × 64, a–d) and FLIR (640 × 512, e–h) clips using VideoMAE with Kinetics-400 vs. process-aware weights under experiment- and process-wise distributions. Each colour-coded cluster represents feature embeddings. FLIR benefited more from fine-tuning due to its higher spatial resolution, while CLAMIR showed limited feature rearrangement post fine-tuning.

reproducibility. The fine-tuned models captured spatio-temporal representations of process regimes to classify sequences into normal, defective, and failure states.

3.2. Exploration of multi-camera deep learning features

The relative distributions of pre-trained and fine-tuned features by comparing most important feature pairs from two cameras were explored, as presented in Figure 12. Only the most important features (top 35) from each dataset for classifying defective, normal, and failure processes are visualised to enhance clarity and facilitate interpretation in the plots. The pre-trained CLAMIR and FLIR features are mostly similar in their standardised distributions. StandardScalar from scikit-learn library was used for standardisation. Some CLAMIR features (CLAMIR_209, CLAMIR_4, CLAMIR_322) have narrower densities, whereas some FLIR features have relatively wider distributions (FLIR_377, FLIR_149, FLIR_310). This is expected due to CLAMIR's spatial resolution being much smaller, leading to high intensity regions with increased similarity and correlations. In comparison, FLIR features are spread out, making probability density peaks of lower amplitude in comparison to CLAMIR.

Following fine-tuning, both the order of feature importance and their distributions were updated, highlighting the impact of process-aware (e.g. defective, failure, normal) parameter updates. Most CLAMIR (29 out of 35) and FLIR features (30 out of 35) are replaced with newer and different features. The repeated

Table 8. Comparison of cluster quality metrics SC, DBI, and CHI for separating defective, failure and normal features from each camera model before and after fine-tuning.

		VideoMAE-kinetics features			VideoMAE-process features		
	Cluster Metrics	Defective	Failure	Normal	Defective	Failure	Normal
CLAMIR	SC	0.0389	0.1460	−0.0105	0.2087	0.2210	0.1368
	DBI	8.5437	10.3283	14.2143	1.8162	1.9607	1.9155
	CHI	70.0835	38.3025	18.2771	1513.6204	857.9514	978.7210
FLIR	SC	0.1794	0.2325	−0.0113	0.3542	0.2914	0.3916
	DBI	1.9569	1.4983	3.9770	1.3836	1.1638	0.8957
	CHI	1103.1317	1099.3968	183.2822	2376.2586	1805.9869	3580.6540

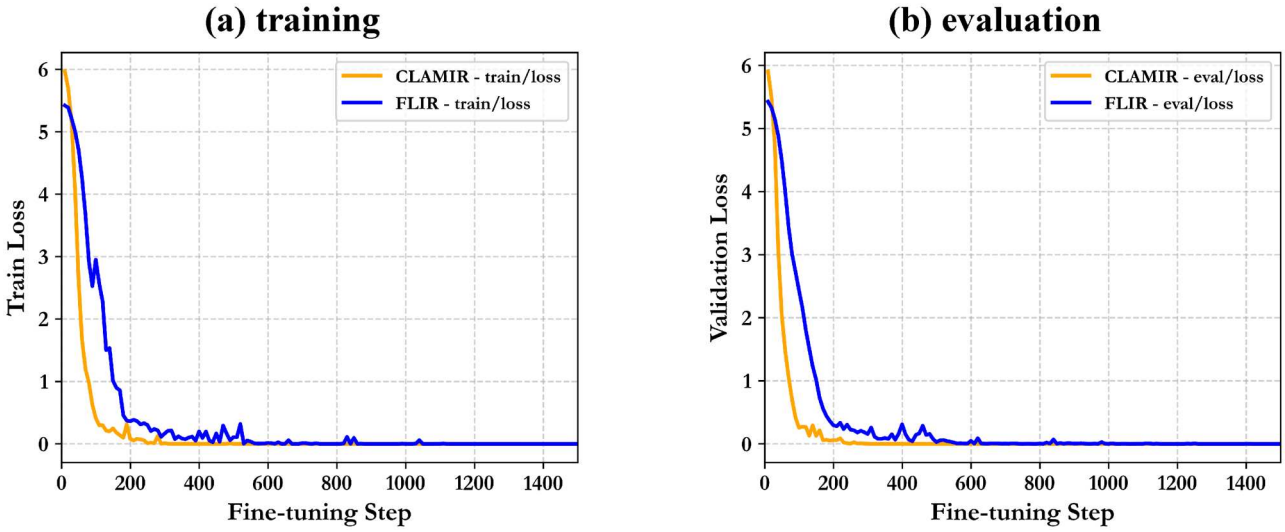

Figure 11. Training (a) and validation (b) loss curves for CLAMIR and FLIR during VideoMAE fine-tuning process. The plot shows 1500 steps out of the total 2000 fine-tuning steps to focus on regions of variations.

feature for CLAMIR (100, 140, 209, 229, 285, 302) and FLIR (12, 16, 49, 179, 230) highlight their maintained significance for separating the process states. Similarly, the fine-tuning also leads to a change in the feature distributions producing non-overlapping density curves for important CLAMIR and FLIR features. The fine-tuned features from each camera also exhibit bimodal and multimodal distributions as opposed to the unimodal pre-trained distribution, indicating class specific distributions for each of three regimes. Interestingly, a few pairs of CLAMIR and FLIR features maintain matching or similar distributions even after fine-tuning (FLIR_16 vs. CLAMIR_369, FLIR_292 vs. CLAMIR_367, FLIR_49 vs. CLAMIR_236). Upon investigation, distinct peaks were found to correspond to specific process regimes, demonstrating the effectiveness of fine-tuning in learning process-aware features across both datasets, with more unique FLIR peaks compared to CLAMIR.

The non-linear and linear correlations between the 35 most important CLAMIR and FLIR features are presented in Figure 13 and in Appendix Figure A1, respectively. The maps are discussed in four parts. The top-left region highlights the FLIR-FLIR feature correlations. It is noticeable that the intra-FLIR correlations are not significantly strong, highlighting that the FLIR features are capturing different aspects of the process. The top-right and bottom-left regions show mirrored representation of FLIR-CLAMIR feature correlations. Inter-camera feature correlations are much weaker than intra-camera feature correlations for the most important features. This indicates that the FLIR and CLAMIR features are uncorrelated, a key aspect that is a strong incentive for combining all features to enhance in-process microstructure defect detection and classification. The bottom right region highlights intra-CLAMIR feature correlations. There are several visible clusters of strong positive or negative correlations. This highlights the relevance of feature selection or dimensionality reduction for CLAMIR, since its features may be capturing similar information. It was also observed that moving from the most important features to features with less importance increased the intra-camera feature correlations significantly, whereas the inter-camera correlations remained in a similar range. The correlations' map highlights two aspects: the potential for inter-camera feature fusions and the relevance of intra-camera feature reductions. For both types of correlations, more highly correlated pairs are found within each individual dataset (linear intra-FLIR: 448 pairs, non-linear intra-FLIR: 114 pairs, linear intra- CLAMIR: 706 pairs and non-linear CLAMIR: 612 pairs) compared to across datasets (linear FLIR-CLAMIR: 5 pairs and non-linear FLIR-CLAMIR: 1 pair). The cross-camera linear correlations are much weaker than non-linear correlations. In both cases, correlations between CLAMIR features highlight greater redundancy than FLIR. Overall, these observations warrant careful selection and reduction of features to effectively enhance information for defect extent classification.

3.3. Feature selection and reduction

Following the observations on prevalent noise and redundancy in our dataset, feature selection and reduction was performed. The original feature sets for all six co-existing anomalies were reduced using SHAP-

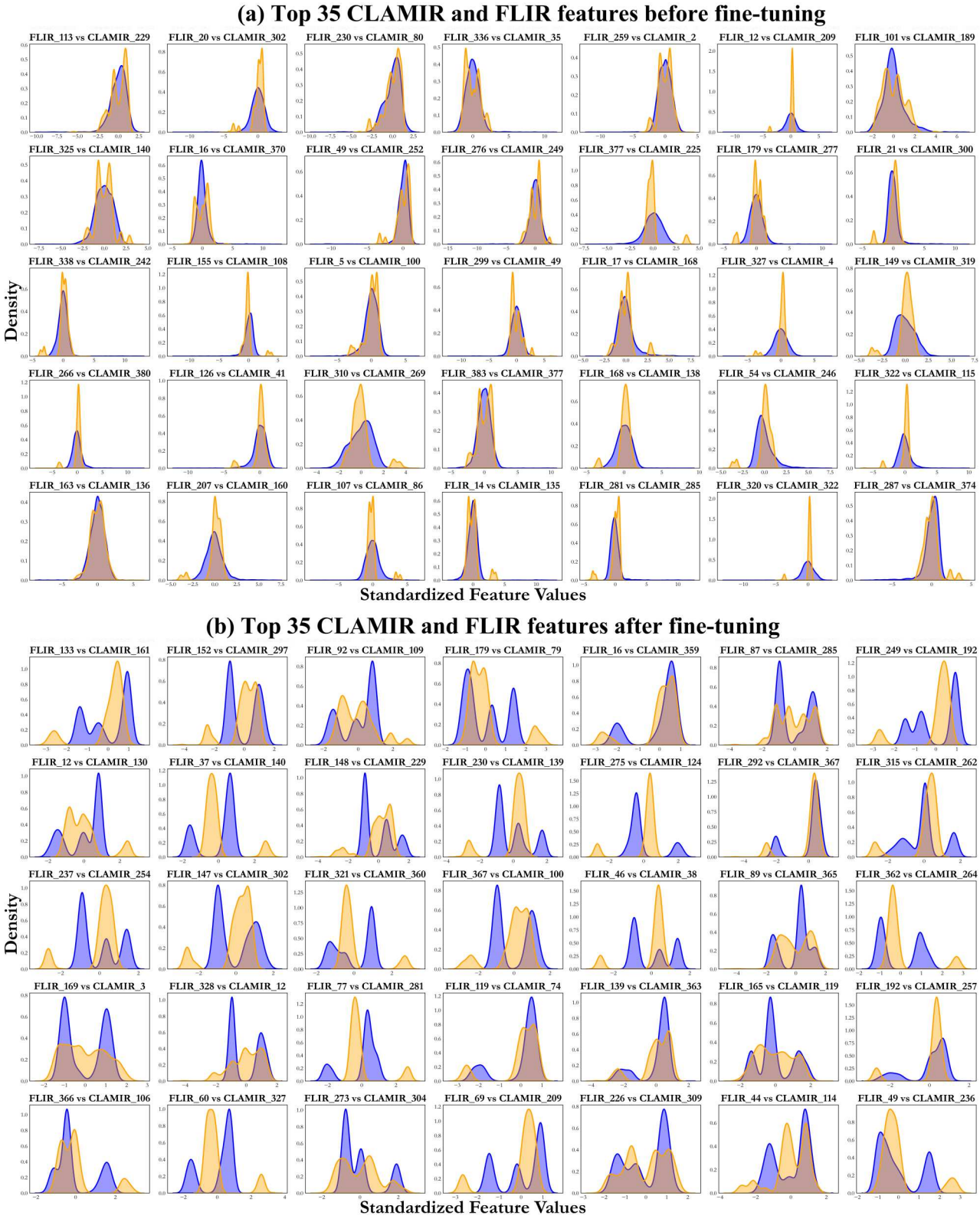

Figure 12. Comparison of 35 most important pre-trained (a) and fine-tuned (b) features from CLAMIR and FLIR in standardised form using kernel density estimation for continuous feature values.

based feature importance in a tree-based classifier. Tables 9 and 10 show the reduction of CLAMIR and FLIR features, respectively. The classification layer of the VideoMAE model outputs 384 features for each input video clip of shape $16 \times 3 \times 224 \times 224$ (e.g. one layer before the task layer). The extracted features

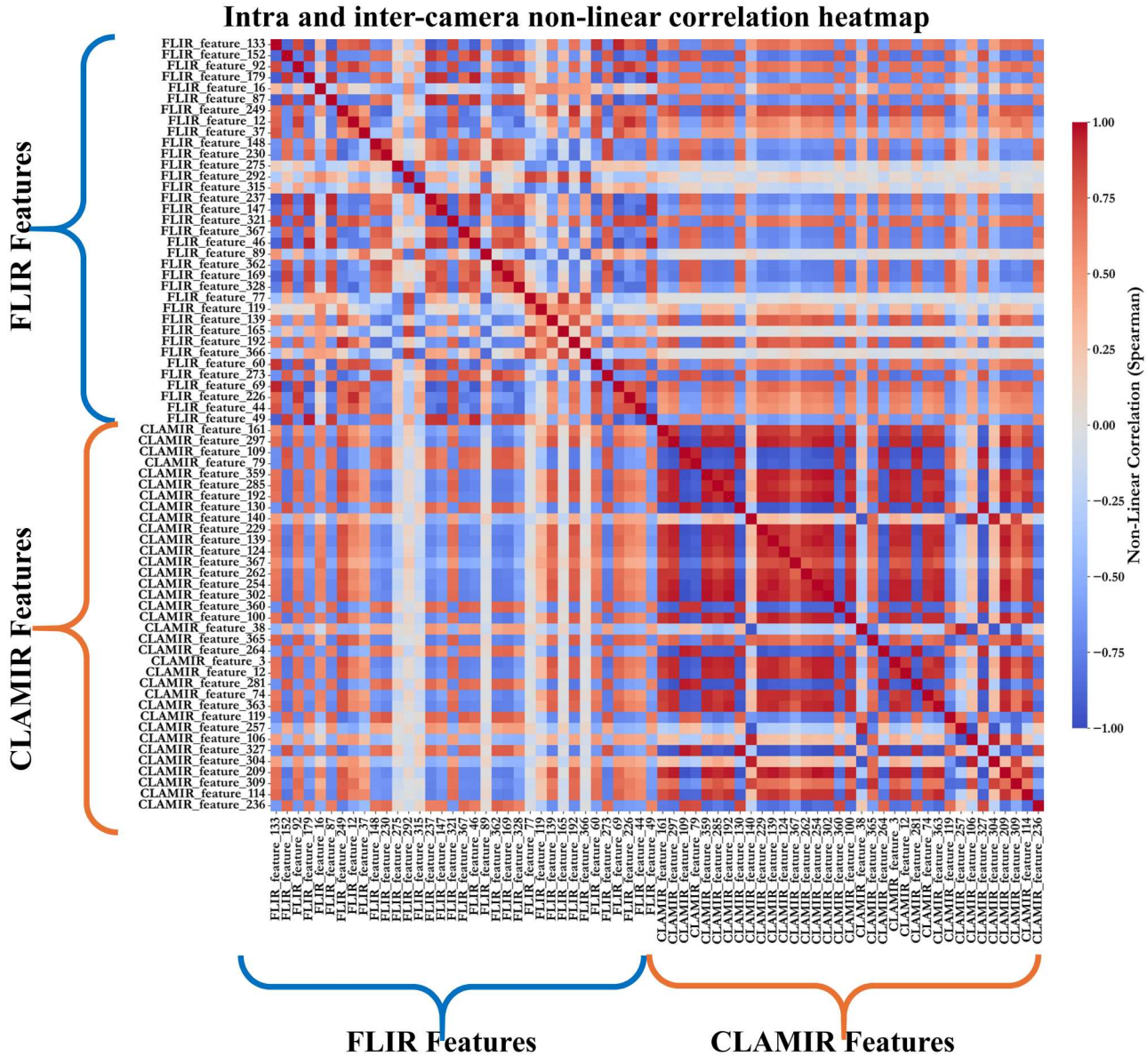

Figure 13. Intra and inter-camera non-linear feature correlations for the most important (1–35) fine-tuned VideoMAE features.

were arranged into six datasets, one for each defect category. As shown in Figure 4, each dataset contained three ground truth labels to separate the extent of defect into absent, low, and high classes. The SHAP values enabled reorganisation of the extracted VideoMAE features for each extent class based on their importance. Recursive elimination of least significant features to most significant features allowed the removal of noisy features in each class. Feature elimination was stopped with at least 15 features for each class. The class-wise selected features were combined, and unique features were maintained. Finally, the highly correlated features across the three classes were removed using a Spearman correlation value of 0.85, leading to a final reduced set with minimum noise and redundancy. The feature selection process led to the most significant features for each defect category. The CLAMIR features saw higher reduction compared to FLIR, owing to extreme redundancy as discussed in the last Section. Moreover, both camera datasets contained shared features across all classes, indicating their overall importance for defect extent separation task. Notably, the CLAMIR features were reduced to as low as 12 for both ECD as well as RPHP extent separation with most features retained for SP at 40. On the other hand, the FLIR features were ranked more uniformly, leading to a higher number of features at the end, such as 299 for ECD and 290 for GP with lowest features retained for SP at 49.

Table 9. Explainable selection of CLAMIR features for defect-extent classes.

Defect	Original features	Class-wise selected features	Combined feature (unique)	Reduced features
GP	384	Absent: 34 Low: 14 High: 24	42	22
SP	384	Absent: 74 Low: 14 High: 14	76	40
EDB	384	Absent: 14 Low: 14 High: 14	22	13
NCD	384	Absent: 34 Low: 24 High: 44	53	28
ECD	384	Absent: 14 Low: 14 High: 14	23	12
RPHP	384	Absent: 14 Low: 14 High: 14	19	12

3.4. Best extent classifiers

Reduced features were used for RF, SVC, XGB, and LR model selection. The models were trained through extensive hyperparameter search and stratified GroupK-Fold cross-validation [55]. The selected hyperparameters for top-performing models are shown in Tables 11 and 12 for CLAMIR and FLIR respectively. Both the CLAMIR and FLIR datasets yielded distinct parameters for each defect category within each algorithm type. The results highlight model specific tuning required for optimal performance across the defect categories.

Among CLAMIR-based defect extent classifiers, RF estimators varied significantly across the categories, with higher values (e.g. 300 for GP and 200 for EDB) indicating the need for more estimators. The maximum depth values remained relatively low, except for ECD and RPHP, which required deeper trees (e.g. 30). The minimum samples split, and minimum samples leaf values were generally small, promoting more splits, and the maximum features varied between 'log2' and 'sqrt'. The SVC models highlighted changing complexity with C values ranging from 0.1 to 100 and kernel types switching between 'rbf' and 'poly'. The XGB models showed a preference for lower learning rates (0.01–0.2) and moderate tree depths (mostly 3–5), with subsample values consistently at 0.6, except for ECD (1.0). In the case of the LR models, the C parameter fluctuated from 0.01 to 100, indicating different levels of regularisation, and the solver was fixed as 'liblinear', with penalty switching between 'l1' and 'l2'.

Table 10. Explainable selection of FLIR features for defect-extent classes.

Defect	Original features	Class-wise selected features	Combined feature	Reduced features
GP	384	Absent: 334 Low: 14 High: 14	336	290
SP	384	Absent: 34 Low: 54 High: 44	60	49
EDB	384	Absent: 214 Low: 14 High: 14	219	188
NCD	384	Absent: 134 Low: 14 High: 14	140	126
ECD	384	Absent: 374 Low: 14 High: 14	374	299
RPHP	384	Absent: 24 Low: 74 High: 14	74	51

Among the FLIR-based defect extent classifiers, the RF estimators also varied significantly across the categories, with highest value of 500 for RPHP, indicating the need for more estimators. The maximum depth values were consistently low (except for RPHP), suggesting a preference for shallower trees. The minimum samples split, and minimum samples leaf values varied slightly with more smaller values for deeper splits, and maximum features varied between 'log2' (GP, SP, and EDB) and 'sqrt' (NCD, EDC, and RPHP). The SVC models highlighted changing complexity with C values ranging from 0.1 (GP, SP, EDB, RPHP) to 100 (NCD) and kernel types switching between 'rbf' (EDB, ECD) and 'poly' (GP, SP, NCD, RPHP). The XGB models showed a range of estimators (e.g. 50 to 300) with maximum depth values (e.g. 3, 7, and 10) reflecting the varying models' complexity. The learning rates and subsamples were optimised for generalisation. The C parameter in the LR models ranged from 0.01 to 100 with a mix of 'l1' and 'l2' penalties and the 'liblinear' solver.

Table 13 lists the performance in weighted accuracy and weighted F1 score for the best performing models during the hyperparameter search (RandomizedSearchCV). The reported performances come from the top-performing fold during stratified GroupKFold cross validation instead of a random validation set to avoid data leakage from training to validation sets leading to unrealistically high performances. Overall, the CLAMIR-based models achieve lower performance compared to the FLIR models across the six defect categories. Five out of six CLAMIR-based classifiers were selected from the RF parameter grid (0.87 GP, 0.83 SP, 0.88 EDB, 0.86 NCD, 0.94 ECD) with the exception of the RPHP classifier, which was selected from the SVC parameter grid (0.87 RPHP). The performance of the best classifiers for the FLIR dataset is distributed among the different models with the SVC (0.94 GP, 0.95 SP, 0.97 NCD), LR (EDB 0.93, ECD 0.97) and RF (RPHP 0.97) models. Notably,

Table 11. Best hyperparameters for CLAMIR-based models.

Classifiers	Parameter	GP	SP	EDB	NCD	ECD	RPHP
Random Forest (RF) Parameter Grid	n_estimators	300	50	200	100	100	50
	max_depth	10	5	5	5	30	30
	min_samples_split	10	2	2	2	2	10
	min_samples_leaf	4	2	2	4	2	1
	max_features	'log2'	'log2'	'sqrt'	'sqrt'	'sqrt'	'log2'
	class_weight	'balanced'					
Support Vector Classifier (SVC) Parameter Grid	C	0.1	0.1	10	0.1	10	100
	kernel	'rbf'	'poly'	'poly'	'rbf'	'poly'	'poly'
	gamma	'scale'	'scale'	'scale'	'scale'	'scale'	'scale'
	class_weight	'balanced'					
XGBoost (XGB) Parameter Grid	n_estimators	100	300	50	50	50	200
	max_depth	3	10	5	5	5	3
	learning_rate	0.01	0.01	0.1	0.2	0.01	0.2
	subsample	0.6	0.6	0.6	0.6	1.0	0.8
	colsample_bytree	0.8	1.0	0.6	0.6	0.8	0.6
Logistic Regression (LR) Parameter Grid	C	0.1	100	0.1	0.01	100	100
	penalty	'l1'	'l1'	'l2'	'l2'	'l2'	'l2'
	solver	'liblinear'					

the inclusion of LR amongst the top classifiers indicates that the FLIR-based defect extent features are more linearly separable than CLAMIR. However, the performance of the LR FLIR models is similar to the performance of the three other classifiers for EDB and ECD. The top-performing models for each defect category are emboldened in Table 13.

3.5. Impact of noise and process dynamics

The models corresponding to the best set of hyperparameters in each case were trained on data from all folds prior to saving. The top-performing models were then tested using three unseen test sets representing stable, unstable, and perturbed process dynamics. Each test set accounted for 10% of the overall datasets, whereas the perturbed test set was generated by applying visual augmentations to the original video clips. The augmented features were then subjected to multiplicative Gaussian noise with increasingly perturbed distribution. The unstable test set reflected scenarios where the visual appearance of the captured frames varied significantly due to changes in the deposition conditions (e.g. process start, process end, intricate feature deposition). In the current work, the unstable test set consisted of clips corresponding to the start of the process, as they exhibited significant changes even at the same process parameters. Both noise and changing dynamics are grounded in realistic scenarios leading to variations in the input data. Therefore, the evaluation of selected models under these conditions provides useful insights for their industrial deployment [64]. Figure 14 highlights the start and end dynamics relative to the stable region. Differences in the Structural Similarity Index Measure (SSIM) are visible for both the CLAMIR and FLIR inputs; however, the FLIR similarity (that starts from ∼0.88 to ∼0.94 and fluctuates towards the end ranging ∼0.90 to ∼0.94) changes significantly as compared to CLAMIR (ranging ∼0.9980 to ∼0.9999 at the start and ∼0.9965 to ∼0.9985 at the end) when

Table 12. Best hyperparameters for FLIR-based models.

	Hyperparameter	GP	SP	EDB	NCD	ECD	RPHP
Random Forest (RF) Parameter Grid	n_estimators	50	50	100	100	100	500
	max_depth	5	5	5	5	5	10
	min_samples_split	2	10	2	2	10	2
	min_samples_leaf	2	2	4	4	1	1
	max_features	log2	log2	log2	'sqrt'	'sqrt'	'sqrt'
	class_weight	'balanced'					
Support Vector Classifier (SVC) Parameter Grid	C	0.1	0.1	0.1	100	1	0.1
	kernel	'poly'	'poly'	'rbf'	'poly'	'rbf'	'poly'
	gamma	'scale'	'scale'	'scale'	'scale'	'scale'	'scale'
	class_weight	'balanced'					
XGBoost (XGB) Parameter Grid	n_estimators	300	50	50	100	300	100
	max_depth	10	3	7	3	7	3
	learning_rate	0.01	0.01	0.1	0.01	0.1	0.1
	subsample	0.6	0.8	0.6	0.6	0.6	0.8
	colsample_bytree	1.0	1.0	1.0	0.8	0.8	0.8
Logistic Regression (LR) Parameter Grid	C	1	100	0.01	0.02	1	0.1
	penalty	'l2'	'l1'	'l2'	'l2'	'l2'	'l1'
	solver	'liblinear'					

Table 13. Best performing models for CLAMIR and FLIR from hyperparameter search.

	CLAMIR		FLIR	
Models	Weighted Accuracy	Weighted F1 Score	Weighted Accuracy	Weighted F1 Score
GP RF	**0.8717**	**0.7886**	0.9330	0.8917
GP SVC	0.8679	0.7898	**0.9385**	**0.9008**
GP XGB	0.8626	0.7738	0.9376	0.8986
GP LR	0.7738	0.6441	0.9275	0.8825
SP RF	**0.8359**	**0.7064**	0.9122	0.8481
SP SVC	0.8207	0.6867	**0.9466**	**0.9109**
SP XGB	0.8251	0.7036	0.9284	0.8798
SP LR	0.8113	0.6741	0.9300	0.8825
EDB RF	**0.8830**	**0.8129**	0.9288	0.8870
EDB SVC	0.8826	0.8131	0.9268	0.8838
EDB XGB	0.8577	0.7721	0.9283	0.8862
EDB LR	0.8096	0.6890	**0.9305**	**0.8897**
NCD RF	**0.8677**	**0.7891**	0.9575	0.9316
NCD SVC	0.8611	0.7792	**0.9705**	**0.9525**
NCD XGB	0.8053	0.6828	0.962	0.9387
NCD LR	0.8186	0.7055	0.9584	0.933
ECD RF	**0.94**	**0.8974**	0.9648	0.9396
ECD SVC	0.922	0.8641	0.9658	0.9413
ECD XGB	0.8488	0.6777	0.9645	0.9393
ECD LR	0.9095	0.8293	**0.9686**	**0.9462**
RPHP RF	0.8082	0.7014	**0.9696**	**0.9529**
RPHP SVC	**0.873**	**0.7957**	0.9694	0.9526
RPHP XGB	0.8201	0.7173	0.9694	0.9526
RPHP LR	0.7928	0.6824	0.9613	0.9401

The results of the highest performing model in each defect category are emboldened.

subjected to process instability. This behaviour results from FLIR capturing more shape-based features compared to CLAMIR, which only focusses on a subset of high intensity melt pool pixels.

The degrading performances of the CLAMIR and FLIR models across the six defect categories, as these are subjected to features from unstable process dynamics (by replacing the stable test set), are presented in Figure 15(a–f). For noise and instability comparison, we focused on the RF classifiers, as it was the top-performing model type in most CLAMIR-based defect extent classifiers. The CLAMIR models were found to provide better generalisability in performance compared to the FLIR models, especially in the case of GP (Figure 15(a)), SP (Figure 15(b)), NCD (Figure 15(d)), and RPHP (Figure 15(f)). The instability in process dynamics had a similar impact on the ECD (Figure 15(a)) model. This suggests that the CLAMIR dataset provides more robust feature representations that maintain predictive performance under unstable process conditions. One possible explanation for this behaviour is CLAMIR's reliance on intensity features over shape features, which are less affected under instability. The disparity in the performance degradation highlights the varying sensitivity of the two datasets to process-induced variations and emphasises the importance of selecting appropriate data sources for defect extent classification in dynamic manufacturing environments. One notable exception was the EDB model where similar performance degradation occurred for both the CLAMIR and FLIR models suggesting that both models prioritised features (e.g. intensity) with similar dependence on process instability.

Similarly, to evaluate the impact of noise on the predictive performance of the models, visual augmentations grounded in realistic scenarios were applied to the CLAMIR and FLIR video clips. These included frame rotations, brightness, contrast and Gaussian noise. The augmentations for FLIR were kept 5 times higher in magnitude than CLAMIR due to its frame resolution being 8–10 times greater, providing higher capacity against lower perturbation magnitudes (rotation: random ±30°, contrast: ±20% with constant brightness, brightness: ±20% with constant contrast, noise: randomly selected variance between 10 and 50). All augmentations were applied using the Albumentations library to manage frame-wise consistency [61].

The impact of visual noise on the feature distributions from FLIR and CLAMIR cameras is compared in Figure 16. The FLIR density plots (Figure 16(a)) show how different augmentations alter the distribution of the FLIR features compared to the original data (blue curves). While rotation minimally affects the feature distributions, contrast and brightness adjustments shift the peaks, and Gaussian noise broadens the distributions, introducing greater variability in the feature values. In the case of CLAMIR (Figure 16(b)), the original data remains sharply concentrated around specific values. Augmentations such as contrast, brightness, and noise introduce broader, more dispersed distributions for CLAMIR compared to FLIR, with noise having the most evident effect by spreading the feature values across a wider range. The stability of CLAMIR is affected more significantly by noise owing to its reliance on a small number of intensity pixels across a wide range of processing conditions (e.g. laser power from 1000 W to 5000 W) whereas the stability of FLIR is disturbed by changing process dynamics as it affects high fidelity shape features captured by the thermal camera.

The performance of RF classifiers based on CLAMIR and FLIR models when input data is subjected to noise is presented in Figure 17(a–f). Notably. and as opposed to instability, noise had more adverse effect on CLAMIR performance compared to FLIR, which remained around at ~80% overall performance even under extremely noisy data. As a result, the reliance of the CLAMIR-

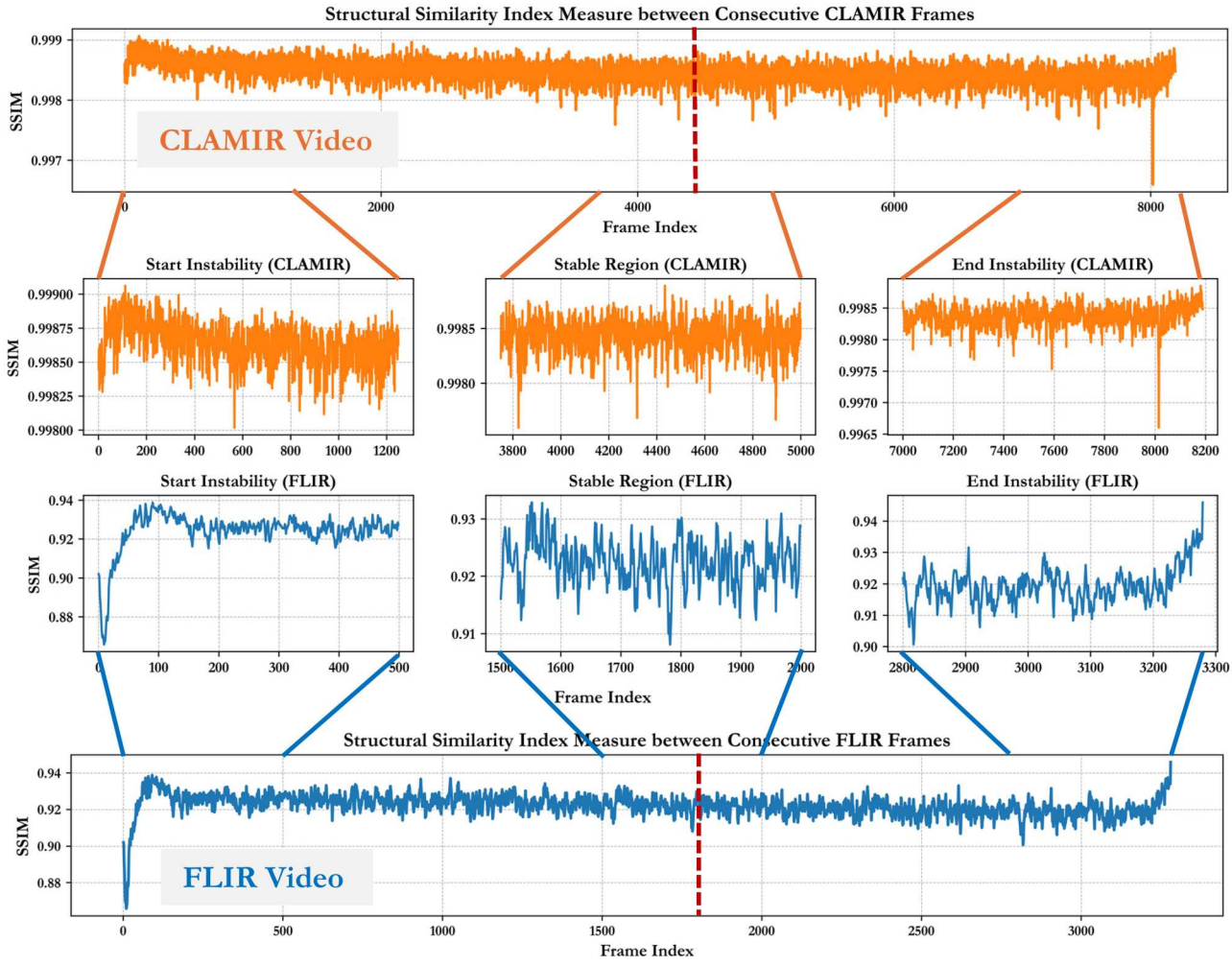

Figure 14. CLAMIR and FLIR signals across a single track with highlighted stable and unstable regions based on SSIM and extracted microstructural images from optical metallography.

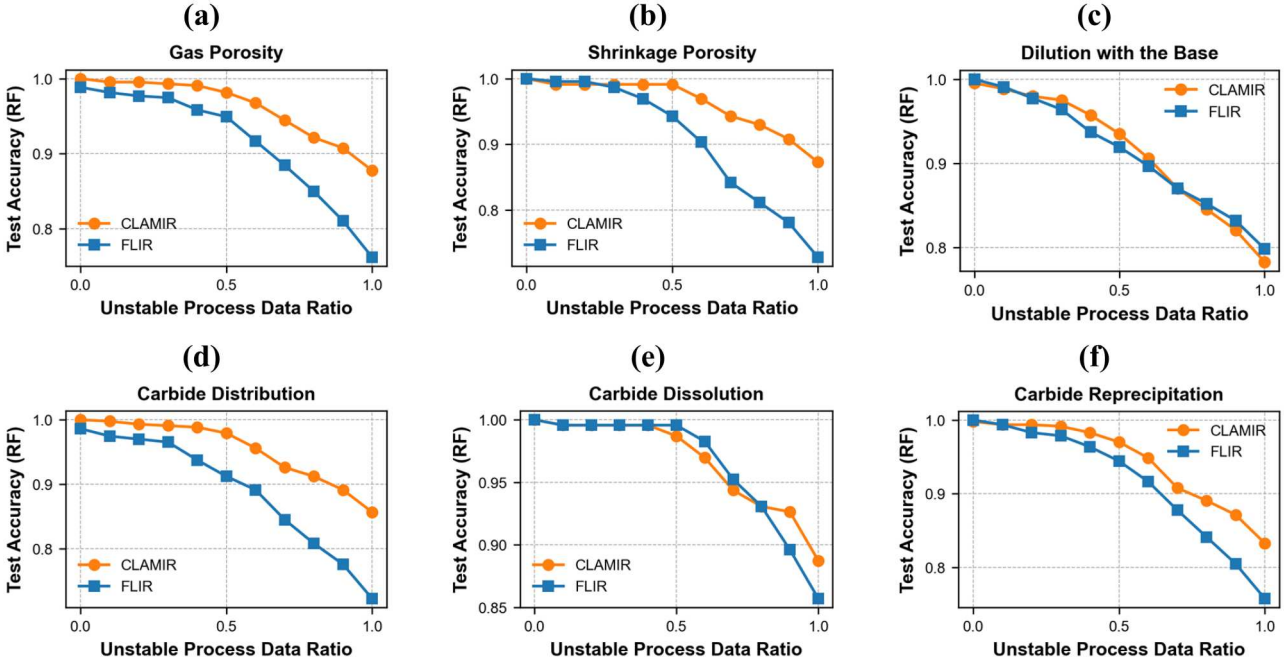

Figure 15. Impact of the unstable process dynamics on the overall performance of the six defect extent classifiers (RF model).

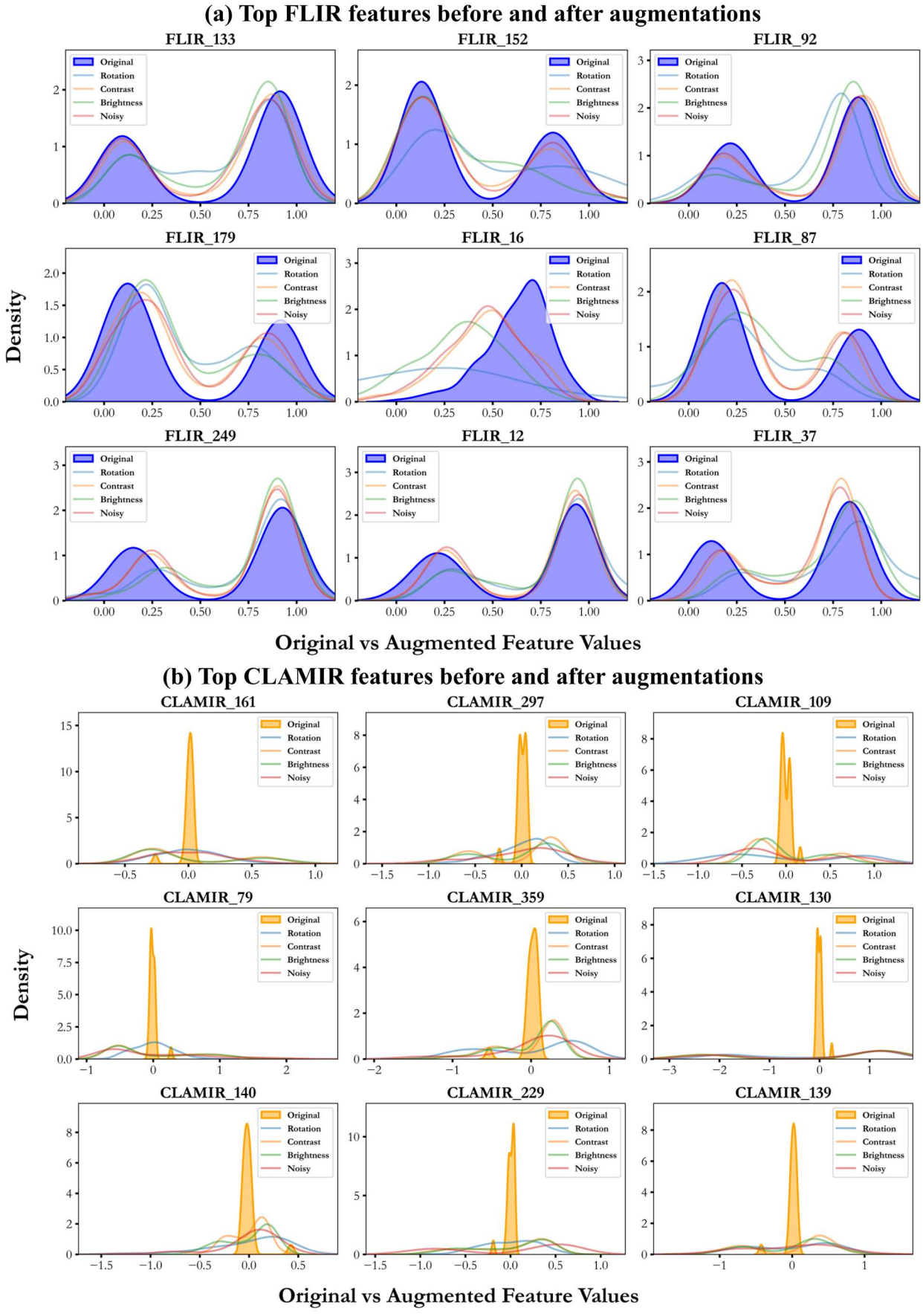

Figure 16. Original and augmented values for standardised FLIR and CLAMIR feature distributions (top 9). CLAMIR values are Zero mean shifted to support comparison between original and augmented distributions. (a) Top FLIR features before and after augmentations, (b) Top CLAMIR features before and after augmentations.

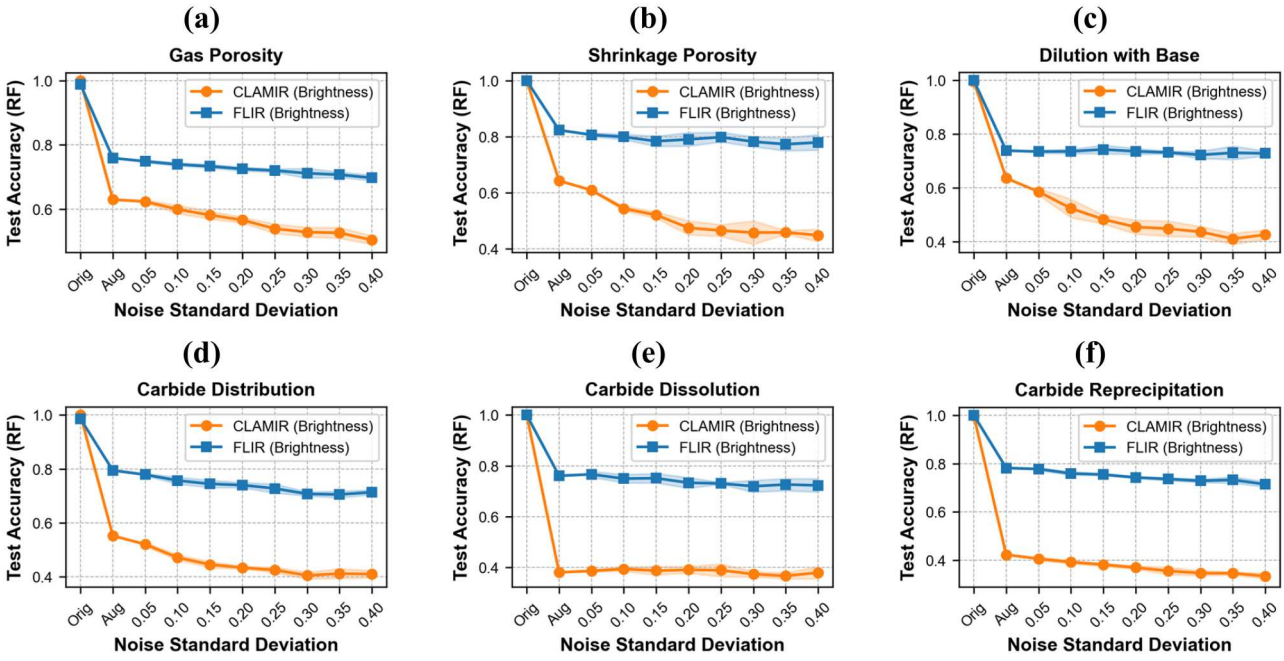

Figure 17. Impact of noise on the overall performance of the six defect extent classifiers, with the example of brightness-based augmentations followed by multiplicative Gaussian noise applied to augmented features (RF model).

based models on a few important features improved its precision but led to poor generalisability. Noise had different impact on specific classifiers. FLIR-based SP (Figure 17(b)), NCD (Figure 17(d)), and RPHP (Figure 17(f)) maintained ≥80% performance when subjected to noise, whereas the performance for FLIR based GP, EDB, and ECD was slightly lower. Gaussian noise was then applied multiplicatively to each augmented feature, scaling its value by a factor of $1 + N\ (0, \sigma)$, where σ represents the noise standard deviation. The noise level was progressively increased from 0.05 to 0.40 in increments of 0.05 to simulate varying levels of perturbation for assessing model robustness. This lowered the computational cost as compared to the visual augmentations and provided an assessment of feature stability for both CLAMIR and FLIR.

Fusion of the features was explored to improve the overall performance, as well as increase the generalisability of the models to changing data distributions from noisy and unstable conditions. This was done by repeating the hyperparameter search for the best models on joint CLAMIR-FLIR feature datasets. In the case of changing process dynamics, the lower performance of the FLIR models was related to the lack of confidence in separating the higher extent of defects from the lower extent. The FLIR-based confusion matrices in Figure 18 highlight significant false negatives for the high-extent class among all defect categories (e.g. GP, SP, EDB, NCD, and RPHP), while maintaining superior or comparable performance on absent and low-extent classes. The CLAMIR-based models see a drop in predicting low extent classes in the case of SP and EDB, as indicated by the false negatives in the respective confusion matrices. The direct fusion of features was explored to exploit the complementary performance of the datasets across defect extent classes (with overall weighted accuracies of GP SVC 0.9380, SP RF 0.8841, EDB LR 0.9478, NCD RF 0.9609, ECD RF 0.9714, RPHP RF 0.9789). However, as shown in Figure 19, direct feature fusion didn't improve the performance on high-extent class for GP, SP, NCD and RPHP (46.7% (Figure 18(c)), 40.5% (Figure 18(f)), 71.7% (Figure 18(l)) and 44.1% (Figure 18(r)) high-extent false negatives). These results are interpreted in the next Section to explain the performance of feature fusion, as well as explore better fusion strategies.

3.6. Fusion of extent separating features

The impact of fusion on extent separating features from each camera was investigated to explain the lower performance of the feature fusion-based models on the GP, SP, NCD, and RPHP defect categories. As shown in Figure 19, the direct fusion of features leads to random selection of the top CLAMIR and FLIR features for high extent class, often being biased to FLIR features over CLAMIR features, which leads to lower performance for the instability scenario. In the case of GP (feature_27 in Figure 19(c)) and NCD (feature_185 in Figure 19(o)), only a single CLAMIR feature ends up in the top 15

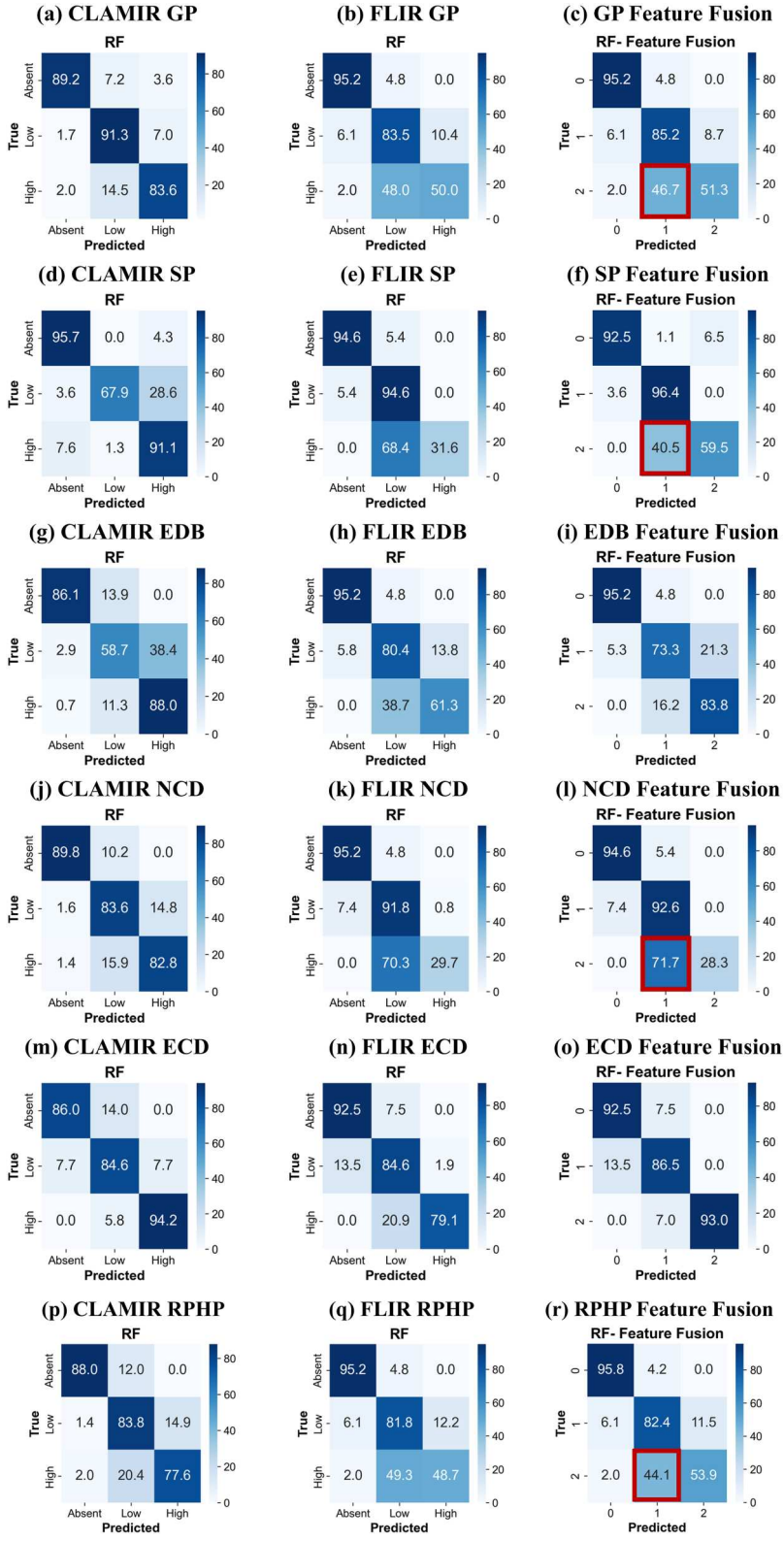

Figure 18. Comparison of CLAMIR and FLIR model performances againt direct feature fusion using unstable test set. The low performance on separating high-extent class after fusion is highliged in red using false negatives for GP (c), SP (f), NCD (l) and RPHP (r). (a) CLAMIR GP, (b) FLIR GP, (c) GP Feature Fusion, (d) CLAMIR SP, (e) FLIR SP, (f) SP Feature Fusion, (g) CLAMIR EDB, (h) FLIR EDB, (i) EDB Feature Fusion, (j) CLAMIR NCD, (k) FLIR NCD, (l) NCD Feature Fusion, (m) CLAMIR ECD, (n) FLIR ECD, (o) ECD Feature Fusion, (p) CLAMIR RPHP, (q) FLIR RPHP, (r) RPHP Feature Fusion.

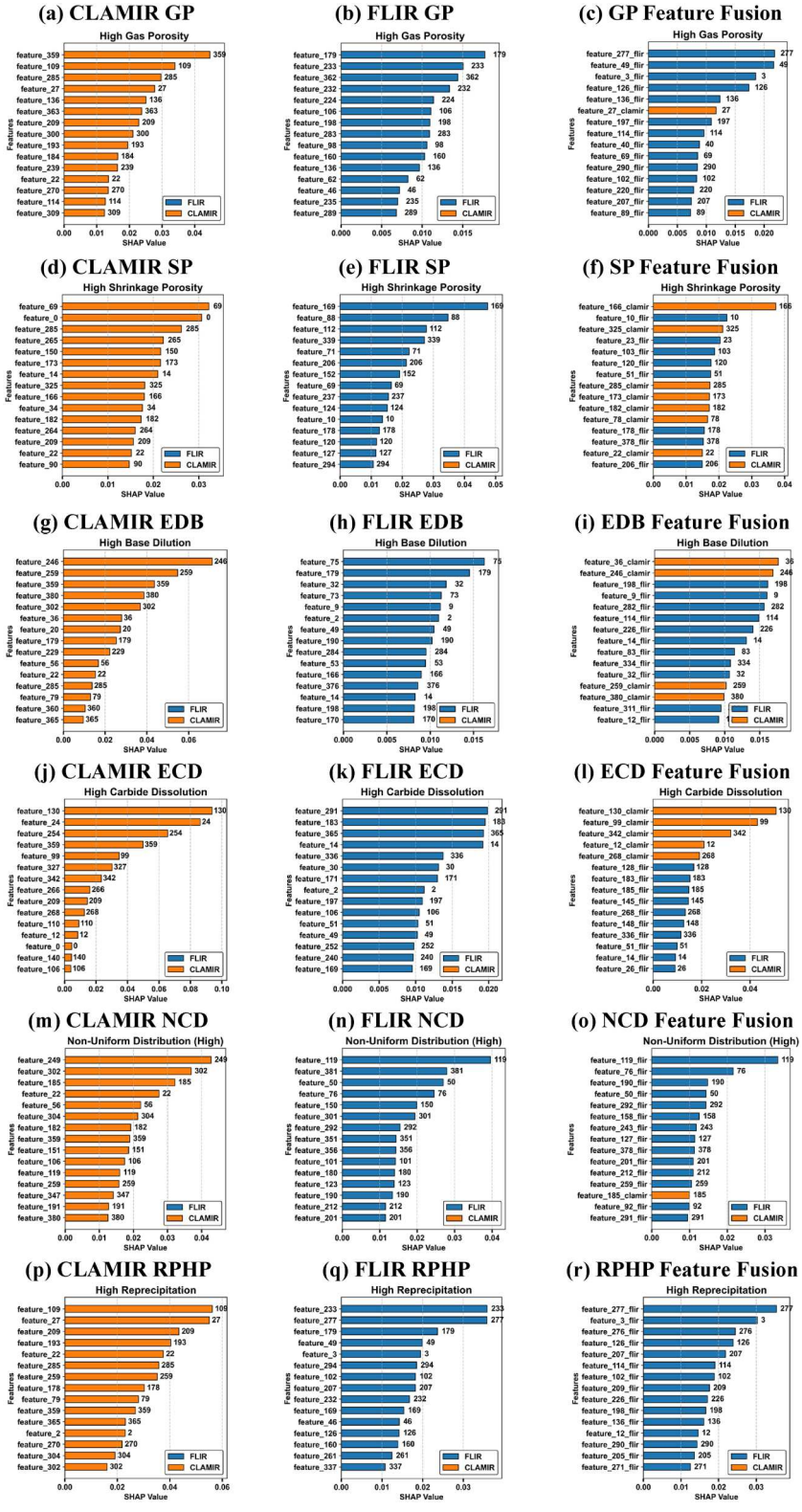

Figure 19. Impact of fusion on the most significant features for detecting high-extent of anomalies (RF model). (a) CLAMIR GP, (b) FLIR GP, (c) GP Feature Fusion, (d) CLAMIR SP, (e) FLIR SP, (f) SP Feature Fusion, (g) CLAMIR EDB, (h) FLIR EDB, (i) EDB Feature Fusion, (j) CLAMIR ECD, (k) FLIR ECD, (l) ECD Feature Fusion, (m) CLAMIR NCD, (n) FLIR NCD, (o) NCD Feature Fusion, (p) CLAMIR RPHP, (q) FLIR RPHP, (r) RPHP Feature Fusion.

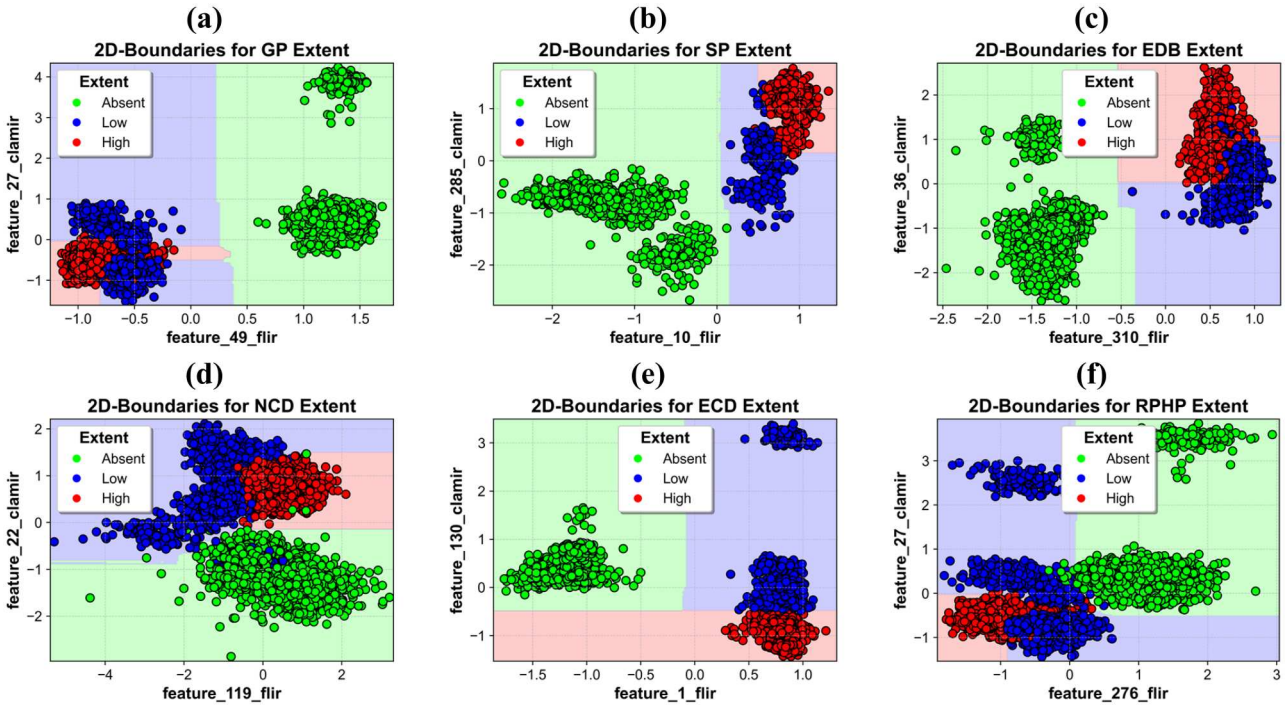

Figure 20. Decision boundaries using one CLAMIR and one FLIR feature from the overall ranked feature set across six anomalies.

features following fusion, whereas no CLAMIR feature was found among the top fusion features in the case of RPHP (in Figure 19(r)). Although the fusion models for SP leveraged more CLAMIR features, these weren't the most effective ones to separate the extents of low and high defects. By contrast, the EDB (Figure 19(i)) and ECD (Figure 19(l)) models that performed relatively better on the fusion dataset retained top CLAMIR features (feature_246, feature_130) among the overall feature rankings, enabling better separation of the high-extent from the other classes.

The distribution of unstable test data across the important features in terms of values and impact was evaluated using SHAP, as highlighted in Appendix Figure A2. The SHAP feature importance values on x-axis are based on high extent predictions. This way, the features are interpreted in terms of their impact on separating the high-extent of defect categories from the low-extent or extent-free cases. Across the four defect categories (GP, SP, NCD, RPHP), where the FLIR models struggle with the classification of high-extent defects, CLAMIR provides stronger features with a greater variance in the SHAP values, enabling interpretation. On the other hand, the FLIR features are more balanced with a lower variance in SHAP values. This characteristic impacts their performance when detecting high-extent defects. Therefore, the CLAMIR features (e.g. feature_359 of GP, feature_69 of SP, feature_249 of NCD, and feature_109 of RPHP) are more dominant in separating the extents than FLIR, owing to the significant spread of the SHAP values, as well as stronger positive and negative shifts to support model predictions (e.g. higher positive values influencing likely model prediction for the high-extent class and vice versa). Upon fusion, the model depends more on the FLIR features than CLAMIR, lowering the confidence in predicting high-extent anomalies.

To aid visualisation, 2D separation boundaries for the six defects were plotted using one CLAMIR and one FLIR feature from the overall ranked set in a simplified RF model. Figure 20(a–f) displays the decision criteria for separating the difficult low-/high-extent boundary using the FLIR and CLAMIR training set features. It is important to note that these are not the actual decision boundaries, as the models depend on more than two features to perform the classification task. However, this visualisation provides insights into the challenges associated with maintaining model performance under noise and instability. Since both low- and high-extent represent the presence of a defect, it is more difficult to separate them from each other, as opposed to separating non-defective features. This observation is consistent across the six defect categories. Moreover, added variations in the test set through noise and instability can be expected to further increase the complexity of the extent separation task.

In the case of noisy data, the CLAMIR features failed to add value to the fusion process across all classes, leading

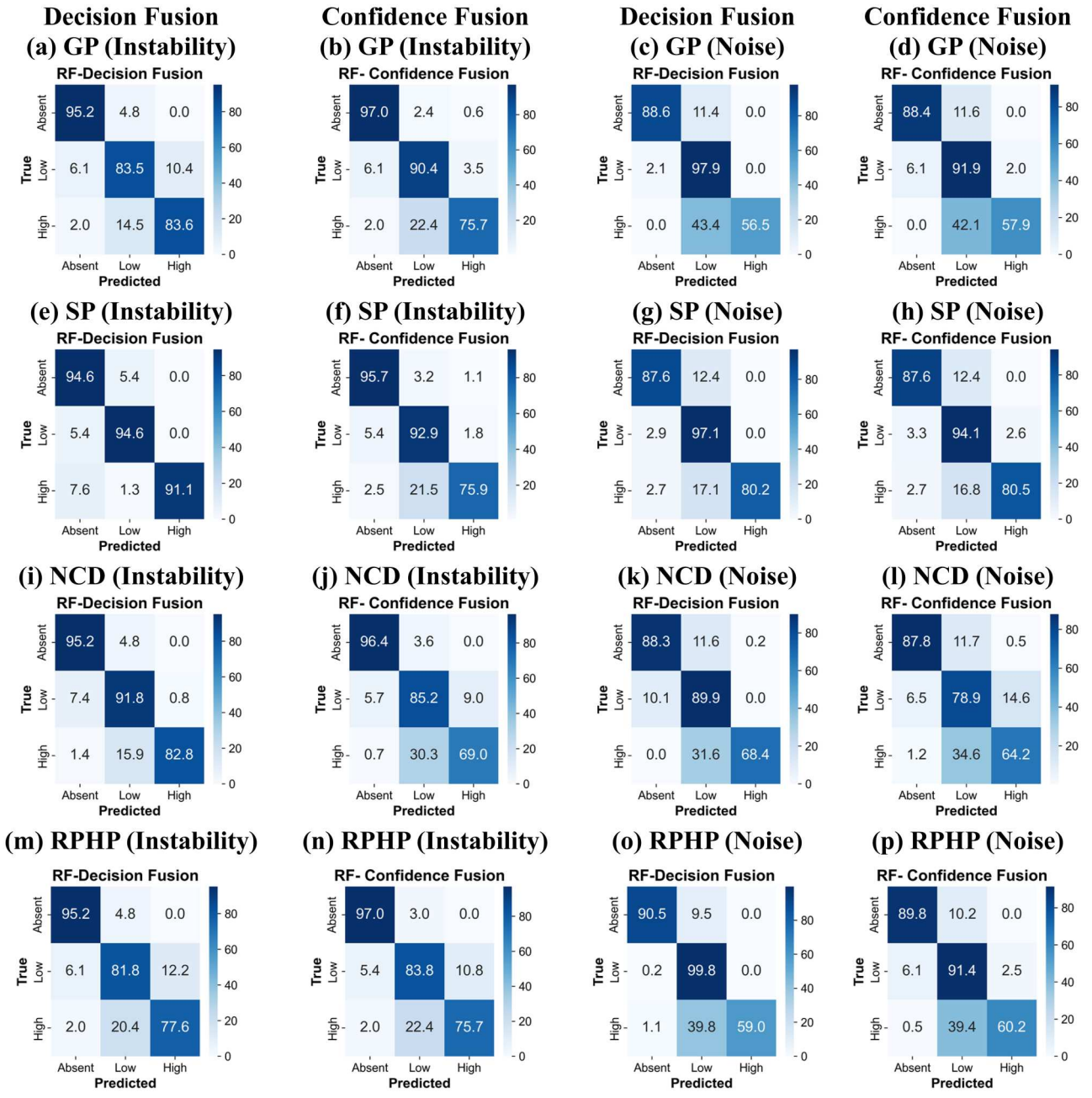

Figure 21. Comaprison of decision and confidence based fusions on defect-extent classification under instability and noise (RF model). (a) GP (Instability), (b) GP (Instability), (c) GP (Noise), (d) GP (Noise), (e) SP (Instability), (f) SP (Instability), (g) SP (Noise), (h) SP (Noise), (i) NCD (Instability), (j) NCD (Instability), (k) NCD (Noise), (l) NCD (Noise), (m) RPHP (Instability), (n) RPHP (Instability), (o) RPHP (Noise), (p) RPHP (Noise).

to lower performance with fusion compared to the single FLIR models. As a result, a decision level or confidence-based fusion with priority for FLIR predictions was explored. Based on the observations from the direct fusion of features and its impact across camera features, as well as defect classes, confidence-based and decision level fusions more effectively leverage the GP, SP, NCD and RPHP features from two cameras. Confidence based fusion is grounded on predicted probabilities for all three classes (e.g. absent, low, high). If the FLIR model is more confident for the first two classes (prob $\geq$ 0.6), the FLIR predicted labels were assigned, otherwise the CLAMIR predictions were retained. Similarly, for high-extent of defects (class 2), if CLAMIR is more confident (prob $\geq$ 0.6), its predictions were used to classify class 2 across all defect categories. If both models are unable to meet the criteria, then the model with the highest confidence is used to make the prediction. This

approach provided a better way to leverage the complementarity between the two cameras; however, it is still prone to false negatives for low and high extent classes due to the sensitivity for confidence threshold. The decision-based fusion utilised the top performance on each extent class from the CLAMIR and FLIR features for noise (0: FLIR, 1: CLAMIR, 2: FLIR) and instability (0: FLIR, 1: FLIR, 2: CLAMIR). While the rule-based decision level fusion maintains the highest performing results across the three classes, fusing the predictions based on the confidence of each model ensures better generalisability for unseen test sets.

The improved fusion results under instability and noise, as presented in Figure 21, provide important insights. Overall, the analysis highlights the effectiveness of careful fusion to retain top-performing features from each dataset. In the case of process instability, confidence-based fusion preformed better on class 0 (absent) across the four categories of interest. Its performance on class 1 (low-extent) was comparable with the decision fusion results. However, in the case of high-extent separation, decision fusion outperformed confidence-based fusion, due to its reliance on stable CLAMIR feature results. In the case of process noise, both strategies yielded similar results for performance on class 0 and decision level fusion performed better on class 1. Notably, the source of retained performance on class 2, in cases with lowest performance (57.9% GP in Figure 21(d), 64.2% NCD in in Figure 21(l) and 60.2% RPHP in Figure 21(p)) were FLIR features, indicating stronger resistance to severe noise. Nonetheless, it is important to highlight that noise had adverse effect on the high-extent predictions from both the CLAMIR and FLIR datasets, which limited the performance improvement through fusion, irrespective of the strategy.

4. Conclusions

The DED AM process, owing to its material deposition freedom, is well suited to repair and enhance complex geometries of functional parts/tools. In the present research, powder-fed laser DED-based deposition was advanced to address the challenge related to process development with a Ni-WC MMC that is applied to modify the surface condition and enhance the wear resistance of a ferrous alloy used in mining machinery. Inspired by the existing lack of understanding for the complex relationship between the DED process parameters and the resulting Ni-WC MMC microstructure, this research investigated the potential of using two MWIR camera systems (FLIR and CLAMIR) to monitor the melt pool and surrounding regions for ML-based microstructure defect detection. Co-existing micro-scale anomalies of the matrix, reinforcement and their interactions are unique to AMed Ni-WC MMCs and present a significant detection challenge for industrial applications. The main contribution of this work was to explain the selection and fusion of MWIR features from fine-tuned vision transformer for enhanced detection of Ni-WC MMC anomalies and their extent. Moreover, data-ablation studies to evaluate the impact of noise and process dynamics revealed practical considerations for multi-camera feature selection and fusion when deployed for real-world production.

Comprehensive experiments at industrial scale were undertaken both in a production environment, as well as in an R&D setup. The datasets from these experiments were used to investigate the potential of different camera systems (CMOS, CLAMIR, & FLIR) to capture features relevant to the process states and microstructure defects. State-of-the-art vision transformer model VideoMAE was fine-tuned to learn DL features from monitored video clips of major process states. These features were found to provide better separability as compared to image-based models. The fine-tuning process further improved the ability of features for defect detection. SHAP, an explainable AI technique, was used to select the most relevant features for each defect category (GP, SP, EDB, NCD, ECD, RPHP) to maximise model performance, leading to significant feature reduction for smaller MWIR camera. The extracted features were employed in four shallow classifiers (RF, SVC, XGBoost, LR) to discover the best hyperparameters through cross-validation and grid search. FLIR, owing to its significantly large frame resolution (8–10 times), outperformed the CLAMIR features when detecting the extent of six co-existing anomalies (weighted accuracies: GP 0.938, SP 0.946, EDB 0.930, NCD 0.970, ECD 0.968, RPHP 0.969). The best performing models from the CLAMIR and FLIR datasets were subjected to changing environment conditions and process dynamics. CLAMIR was found to outperform FLIR consistently on unstable test set (by an average of 8.5 percentage points at 100% unstable test set), whereas FLIR provided stronger generalisation under noise as compared to CLAMIR (by an average of 30.67 percentage points at augmented test set). These observations highlighted the complementarity of the MWIR cameras for practical conditions and their dominant dependence on precise intensity (CLAMIR) and generalisable shape (FLIR) features. Subsequently, fusion strategies (e.g. feature, decision, and confidence levels) were evaluated to maximise the performance of the most relevant CLAMIR and FLIR features under instability and noise.

The key findings of the proposed detection pipeline demonstrated the suitability of a dual MWIR camera setup for effectively overcoming data perturbations in

industrial repair and overlay applications through careful selection and fusion of DL features. By identifying the most relevant camera features for detecting the extent of specific co-existing anomalies in Ni-WC MMCs, the approach enhanced trust and interpretability of the proposed defect detection framework. Several parameter sets were identified as defect-free (e.g. 2800 W, 1524 mm/min, and 80 g/min), providing a reference point for future process development and potential extension to overlapping track depositions [62]. The substantial reduction in feature dimensionality through selection, along with the successful application of shallow classifiers, can support efficient and scalable deployment on industrial machines. Most notably, the demonstrated ability to relate in-situ monitoring features to defect the extent of co-existing anomalies offered a compelling foundation for future extension of the framework towards regression-based predictions to support finer and localised microstructural control over the deposited overlays.

One limitation of the current work is the use of a single-track monitoring dataset as opposed to overlapping tracks, which is usually the expected deposition level for Ni-WC overlays to enhance part surfaces. Since the detected anomalies are primarily caused due to higher thermal conditions, it is expected that the MWIR cameras can maintain the same relativity between the features of defective and non-defective overlapping tracks, enabling the adaptation of the developed single-track models from the current work. Moreover, a multi-camera setup can effectively support complex 2D and 3D depositions by overcoming hidden or occluded regions from one camera's viewpoint. Another limitation of the current work is the use of classical shallow classifiers for defect extent detection that may be limited in their modelling capacity. This can be overcome by enhancing the dataset size at the defect extent level in the future. Therefore, the most relevant future works include experiments with overlapping tracks to both increase the size of the current defect extent dataset as well as evaluate the applicability of the proposed framework for the higher dimensions of process complexity. In the future, more DED-specific defects such as cracks in the overlapping tracks can also be considered as the framework is expanded from single tracks to layer-level depositions. The models can be deployed on production machines and used in inference or evaluation mode with a fast detection time (e.g. ms) to support in situ defect detection.

Author contributions

All listed authors meet the criteria for authorship and agree to be accountable for all aspects of this work. **Mutahar Safdar**: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Gentry Wood**: Resources, Data curation, Writing – original draft. **Max Zimmermann**: Resources, Data curation, Writing – original draft. **Guy Lamouche**: Conceptualisation, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Priti Wanjara**: Conceptualisation, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Yaoyao Fiona Zhao**: Conceptualisation, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Research Council Canada [grant number NRC INT-015-1].

Data availability statement

Raw data were generated at Apollo-Clad Laser Cladding, a division of Apollo Machine and Welding Ltd. Derived feature-level data that support the findings of this study are openly available in Zenodo at https://zenodo.org/records/15120833.

ORCID

Yaoyao Fiona Zhao http://orcid.org/0000-0003-4927-0514

References

[1] Pei E, Bernard A, Gu D, et al. *Springer handbook of additive manufacturing*. Cham: Springer Nature; 2023.
[2] Mehrpouya M, Dehghanghadikolaei A, Fotovvati B, et al. The potential of additive manufacturing in the smart factory industrial 4.0: a review. Appl Sci. 2019;9(18):3865. doi:10.3390/app9183865
[3] ASTM. Additive manufacturing – general principles – fundamentals and vocabulary. American Society of Testing Materials; 2022.
[4] Guo W, Wang L, Su L, et al. Microstructure evolution and mechanical properties of Q690D steel repaired by wire based laser directed energy deposition. Virtual Phys Prototyp. 2024;19(1):e2375108. doi:10.1080/17452759.2024.2375108
[5] Murr LE. A metallographic review of 3D printing/additive manufacturing of metal and alloy products and components. Metallogr Microstruct Anal. 2018;7:103–132. doi:10.1007/s13632-018-0433-6
[6] Wood G, Mendez P. Disaggregated metal and carbide catchment efficiencies in laser cladding of nickel-tungsten carbide. Weld J. 2015;94(11):343–350.
[7] Sing S, Kuo CN, Shih CT, et al. Perspectives of using machine learning in laser powder bed fusion for metal additive manufacturing. Virtual Phys Prototyp. 2021;16(3):372–386. doi:10.1080/17452759.2021.1944229

[8] Chen L, Bi G, Yao X, et al. In-situ process monitoring and adaptive quality enhancement in laser additive manufacturing: a critical review. J Manuf Syst. 2024;74:527–574. doi:10.1016/j.jmsy.2024.04.013
[9] Nguyen NV, Hum AJW, Do T, et al. Semi-supervised machine learning of optical in-situ monitoring data for anomaly detection in laser powder bed fusion. Virtual Phys Prototyp. 2023;18(1):e2129396. doi:10.1080/17452759.2022.2129396
[10] Srisungsitthisunti P, Kaewprachum B, Yang Z, et al. Real-time quality monitoring of laser cladding process on rail steel by an infrared camera. Metals. 2022;12(5):825. doi:10.3390/met12050825
[11] Binega E, Yang L, Sohn H, et al. Online geometry monitoring during directed energy deposition additive manufacturing using laser line scanning. Prec Eng. 2022;73:104–114. doi:10.1016/j.precisioneng.2021.09.005
[12] Razvi SS, Feng S, Narayanan A, et al. A review of machine learning applications in additive manufacturing. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference; Anaheim, CA. American Society of Mechanical Engineers; 2019.
[13] Chen L, Bi G, Yao X, et al. Multisensor fusion-based digital twin for localized quality prediction in robotic laser-directed energy deposition. Robot Comput Integr Manuf. 2023;84:102581. doi:10.1016/j.rcim.2023.102581
[14] Choi JY, Xue T, Liao S, et al. Accelerating phase-field simulation of three-dimensional microstructure evolution in laser powder bed fusion with composable machine learning predictions. Addit Manuf. 2024;79:103938. doi:10.1016/j.addma.2023.103938
[15] Chung Baek AM, Kim T, Seong M, et al. Multimodal deep learning for enhanced temperature prediction with uncertainty quantification in directed energy deposition (DED) process. Virtual Phys Prototyp. 2025;20(1):e2474532. doi:10.1080/17452759.2025.2474532
[16] Safdar M, Lamouche G, Paul PP, et al. *Engineering of additive manufacturing features for data-driven solutions: sources, techniques, pipelines, and applications*. Cham: Springer Nature; 2023.
[17] Qin J, Hu F, Liu Y, et al. Research and application of machine learning for additive manufacturing. Addit Manuf. 2022;52:102691. doi:10.1016/j.addma.2022.102691
[18] Zhang Y, Safdar M, Xie J, et al. A systematic review on data of additive manufacturing for machine learning applications: the data quality, type, preprocessing, and management. J Int Manuf. 2023;34(8):3305–3340.
[19] Yang H, Zhang S, Lu Y, et al. Spatiotemporal monitoring of melt-pool variations in metal-based additive manufacturing. IEEE Robot Autom Lett. 2022;7(3):8249–8256. doi:10.1109/LRA.2022.3187540
[20] Cherif L, Safdar M, Lamouche G, et al. Evaluation of key spatiotemporal learners for print track anomaly classification using melt pool image streams. IFAC-PapersOnLine. 2023;56(2):4733–4739. doi:10.1016/j.ifacol.2023.10.1235
[21] Safdar M, Li YF, El Haddad R, et al. Accelerated semantic segmentation of additively manufactured metal matrix composites: generating datasets, evaluating convolutional and transformer models, and developing the MicroSegQ+ tool. Expert Syst Appl. 2024;251:123974. doi:10.1016/j.eswa.2024.123974
[22] Mendez PF, Barnes N, Bell K, et al. Welding processes for wear resistant overlays. J Manuf Process. 2014;16(1):4–25. doi:10.1016/j.jmapro.2013.06.011
[23] McGowan E, Gawade V, Guo WG. A physics-informed convolutional neural network with custom loss functions for porosity prediction in laser metal deposition. Sensors. 2022;22(2):494. doi:10.3390/s22020494
[24] Tian Q, Guo S, Melder E, et al. Deep learning-based data fusion method for in situ porosity detection in laser-based additive manufacturing. J Manuf Sci Eng. 2021;143(4):041011. doi:10.1115/1.4048957
[25] Tian Q, Guo S, Guo Y. A physics-driven deep learning model for process-porosity causal relationship and porosity prediction with interpretability in laser metal deposition. CIRP Ann. 2020;69(1):205–208. doi:10.1016/j.cirp.2020.04.049
[26] Yin M, Zhuo S, Xie L, et al. Online monitoring of local defects in robotic laser additive manufacturing process based on a dynamic mapping strategy and multibranch fusion convolutional neural network. J Manuf Syst. 2023;71:494–503. doi:10.1016/j.jmsy.2023.10.005
[27] Kim S, Jeon I, Sohn H. Infrared thermographic imaging based real-time layer height estimation during directed energy deposition. Opt Lasers Eng. 2023;168:107661. doi:10.1016/j.optlaseng.2023.107661
[28] Xie X, Bennett J, Saha S, et al. Mechanistic data-driven prediction of as-built mechanical properties in metal additive manufacturing. npj Comput Mater. 2021;7(1):86. doi:10.1038/s41524-021-00555-z
[29] Knaak C, von Eßen J, Kröger M, et al. A spatio-temporal ensemble deep learning architecture for real-time defect detection during laser welding on low power embedded computing boards. Sensors. 2021;21(12):4205. doi:10.3390/s21124205
[30] Gonçalves DA, Stemmer MR, Pereira M. A convolutional neural network approach on bead geometry estimation for a laser cladding system. Int J Adv Manuf Technol. 2020;106(5):1811–1821. doi:10.1007/s00170-019-04669-z
[31] Zhang B, Liu S, Shin YC. In-process monitoring of porosity during laser additive manufacturing process. Addit Manuf. 2019;28:497–505. doi:10.1016/j.addma.2019.05.030
[32] Khanzadeh M, Chowdhury S, Marufuzzaman M, et al. Porosity prediction: supervised-learning of thermal history for direct laser deposition. J Manuf Syst. 2018;47:69–82. doi:10.1016/j.jmsy.2018.04.001
[33] Jamnikar ND, Liu S, Brice C, et al. In situ microstructure property prediction by modeling molten pool-quality relations for wire-feed laser additive manufacturing. J Manuf Process. 2022;79:803–814. doi:10.1016/j.jmapro.2022.05.013
[34] Jamnikar ND, Liu S, Brice C, et al. In-process comprehensive prediction of bead geometry for laser wire-feed DED system using molten pool sensing data and multimodality CNN. Int J Adv Manuf Technol. 2022;121(1):903–917. doi:10.1007/s00170-022-09248-3
[35] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; Las Vegas, NV. 2016.
[36] Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention. PMLR. 2021;139:10347–10357.
[37] Ali N, Tomesani L, Ascari A, et al. Fabrication of thin walls with and without close loop control as a function of scan

strategy via direct energy deposition. Lasers Manuf Mater Process. 2022;9(1):81–101. doi:10.1007/s40516-022-00164-8
[38] Shah U, Liu X, Benatar A, et al. Computational analysis of the ultrasonic effects on resistance spot welding process. J Manuf Process. 2022;81:191–201. doi:10.1016/j.jmapro.2022.06.050
[39] Malmivirta T, Hamberg J, Lagerspetz E, et al. Hot or not? Robust and accurate continuous thermal imaging on flir cameras. 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom); Kyoto, Japan. IEEE; 2019.
[40] DebRoy T, Wei HL, Zuback JS, et al. Additive manufacturing of metallic components–process, structure and properties. Prog Mater Sci. 2018;92:112–224. doi:10.1016/j.pmatsci.2017.10.001
[41] Sayood K. *Introduction to data compression*. Cambridge: Morgan Kaufmann; 2017.
[42] Zhang Y, Soon HG, Ye D, et al. Powder-bed fusion process monitoring by machine vision with hybrid convolutional neural networks. IEEE Trans Ind Inf. 2020;16(9):5769–5779. doi:10.1109/TII.2019.2956078
[43] Zhao X, Imandoust A, Khanzadeh M, et al. Automated anomaly detection of laser-based additive manufacturing using melt pool sparse representation and unsupervised learning. 2021 International Solid Freeform Fabrication Symposium. Austin, TX: University of Texas at Austin; 2021.
[44] Ho S, Zhang W, Young W, et al. DLAM: deep learning based real-time porosity prediction for additive manufacturing using thermal images of the melt pool. IEEE Access. 2021;9:115100–115114. doi:10.1109/ACCESS.2021.3105362
[45] Guo S, Guo WG, Bain L. Hierarchical spatial-temporal modeling and monitoring of melt pool evolution in laser-based additive manufacturing. IISE Trans. 2020;52(9):977–997. doi:10.1080/24725854.2019.1704465
[46] Selva J, Johansen AS, Escalera S, et al. Video transformers: a survey. IEEE Trans Pattern Anal Mach Intell. 2023;45(11):12922–12943. doi:10.1109/TPAMI.2023.3243465
[47] Lee SK, Ko H. AMTransformer: a Koopman theory-based transformer for learning additive manufacturing dynamics in laser processes. Int J AI Mater Des. 2024;1(2):76–91. doi:10.36922/ijamd.3919
[48] Li Q, Huang T, Liu J, et al. Time-series vision transformer based on cross space-time attention for fault diagnosis in fused deposition modelling with reconstruction of layer-wise data. J Manuf Process. 2024;115:240–255. doi:10.1016/j.jmapro.2024.01.082
[49] Zhang W, Wang J, Tang M, et al. 2-D transformer-based approach for process monitoring of metal 3-D printing via coaxial high-speed imaging. IEEE Trans Ind Inf. 2024;20(3):3767–3777. doi:10.1109/TII.2023.3314071
[50] Safdar M, Lamouche G, Paul PP, et al. Feature engineering in additive manufacturing. In: Engineering of additive manufacturing features for data-driven solutions: sources, techniques, pipelines, and applications. Cham: Springer Nature Switzerland; 2023. p. 17–43.
[51] Arnab A, Dehghani M, Heigold G, et al. Vivit: a video vision transformer. Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC. 2021.
[52] Bertasius G, Wang H, Torresani L. Is space-time attention all you need for video understanding? ICML; 2021.
[53] Liu Z, Ning J, Cao Y, et al. Video swin transformer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; New Orleans, LA. 2022.
[54] Tong Z, Song Y, Wang J, et al. Videomae: masked autoencoders are data-efficient learners for self-supervised video pre-training. Adv Neural Inf Process Syst. 2022;35:10078–10093.
[55] Abhilash PM, Luo X, Liu Q, et al. Towards next-gen smart manufacturing systems: the explainability revolution. npj Adv Manuf. 2024;1(1):8. doi:10.1038/s44334-024-00006-9
[56] Kaufman S, Rosset S, Perlich C, et al. Leakage in data mining: formulation, detection, and avoidance. ACM Trans Knowl Discov Data. 2012;6(4):1–21. doi:10.1145/2382577.2382579
[57] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–2830.
[58] Pandiyan V, Baganis A, Axel Richter R, et al. Qualify-as-you-go: sensor fusion of optical and acoustic signatures with contrastive deep learning for multi-material composition monitoring in laser powder bed fusion process. Virtual Phys Prototyp. 2024;19(1):e2356080. doi:10.1080/17452759.2024.2356080
[59] Valizadeh M, Wolff SJ. Convolutional neural network applications in additive manufacturing: a review. Adv Ind Manuf Eng. 2022;4:100072. doi:10.1016/j.aime.2022.100072
[60] Safdar M, Xie J, Ko H, et al. Transferability analysis of data-driven additive manufacturing knowledge: a case study between powder bed fusion and directed energy deposition. ASME. J. Comput. Inf. Sci. Eng. May 2024;24(5):051010. doi:10.1115/1.4065090
[61] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65. doi:10.1016/0377-0427(87)90125-7
[62] Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat Theory Methods. 1974;3(1):1–27. doi:10.1080/03610927408827101
[63] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;PAMI-1(2):224–227. doi:10.1109/TPAMI.1979.4766909
[64] Safdar M, Paul PP, Lamouche G, et al. Fundamental requirements of a machine learning operations platform for industrial metal additive manufacturing. Com Ind. 2024;154: 104037.
[65] Buslaev A, Iglovikov VI, Khvedchenya E, et al. Albumentations: fast and flexible image augmentations. Information. 2020;11(2):125. doi:10.3390/info11020125
[66] Cai Y, Wang Y, Chen H, et al. Searching optimal process parameters for desired layer geometry in wire-laser directed energy deposition based on machine learning. Virtual Phys Prototyp. 2024;19(1):e2352066. doi:10.1080/17452759.2024.2352066

Appendix

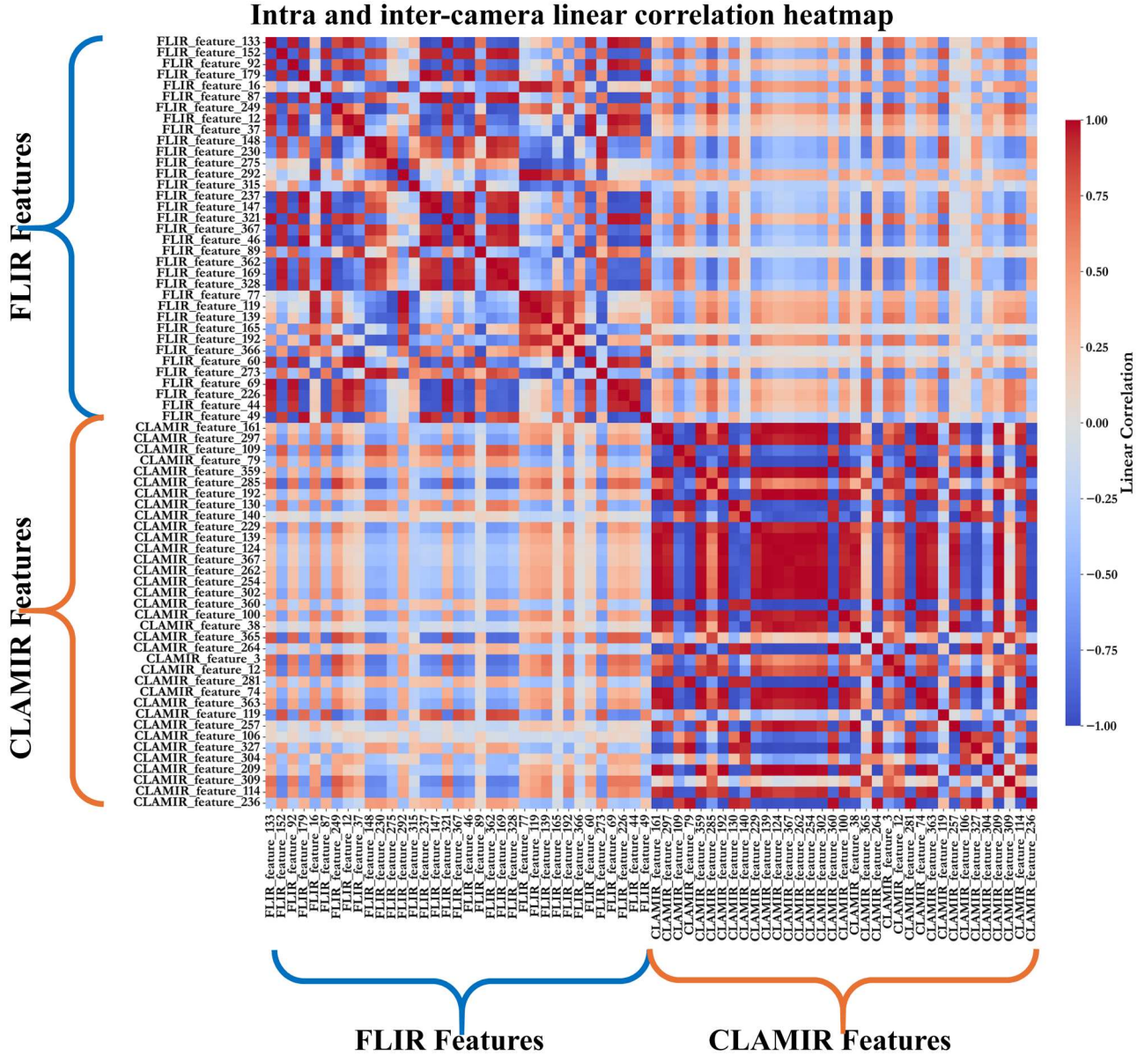

Figure A1. Intra and inter-camera linear feature correlations for the most important (1–35) fine-tuned VideoMAE features.

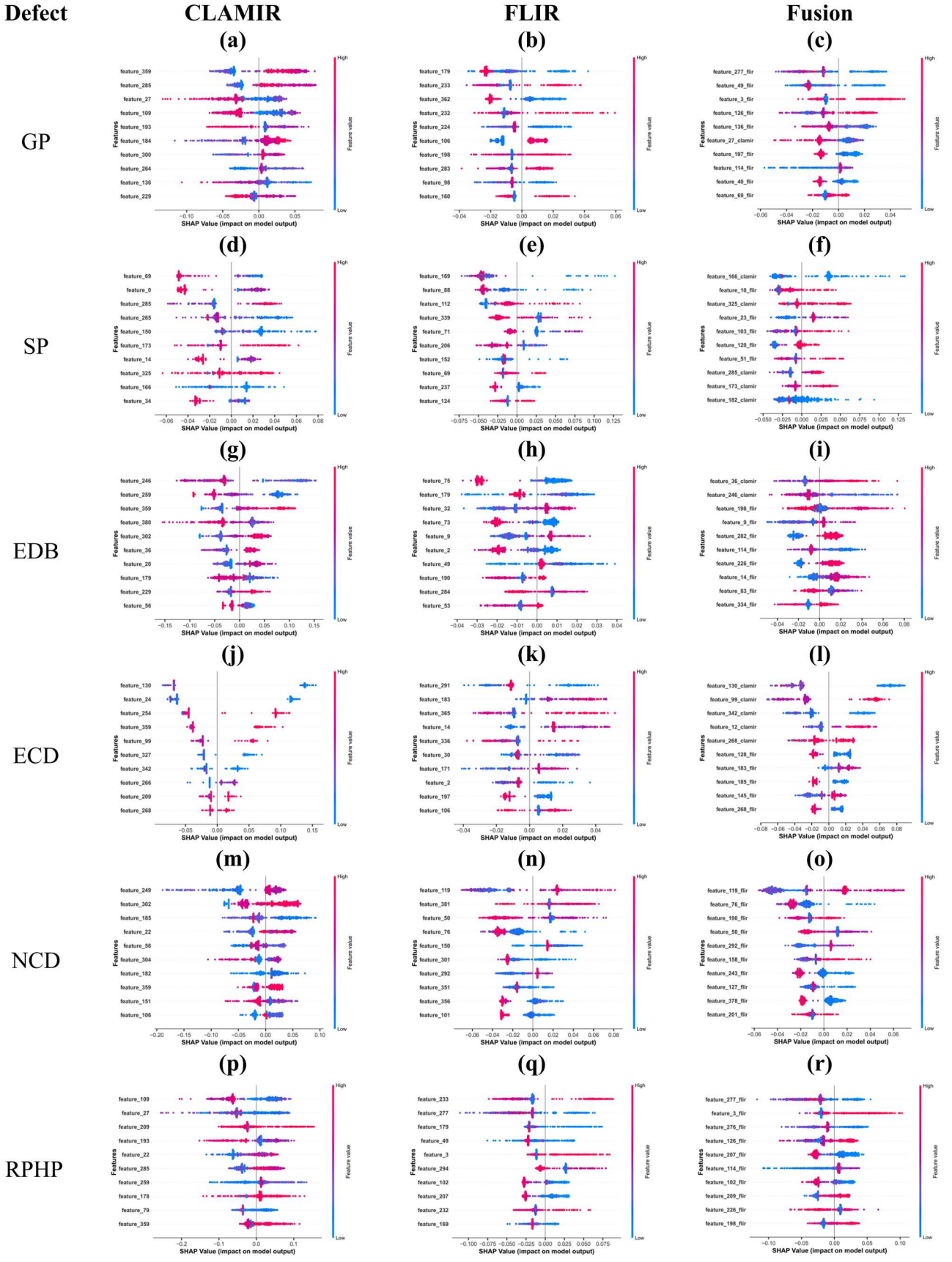

Figure A2. Distributions of feature SHAP values for predicting high extent class across the six classifiers (RF model). Each dot represents a data instance, with its position on the *x*-axis showing the SHAP value, indicating whether the feature increased (positive value) or decreased (negative value) the model's prediction. The colour gradient from blue to pink reflects the magnitude of the feature value, where blue represents lower values and pink represents higher values.