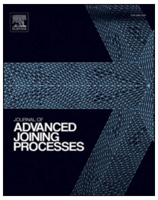
Journal of Advanced Joining Processes 11 (2025) 100318

Contents lists available at ScienceDirect

Journal of Advanced Joining Processes

journal homepage: www.sciencedirect.com/journal/journal-of-advanced-joining-processes



Prediction of weld quality in laser welding of hardmetal and steel using high-speed imaging and machine learning methods

Mohammadhossein Norouzian [a,*], Mahan Khakpour [a], Marko Orosnjak [a], Atal Anil Kumar [b], Slawomir Kedziora [a]

[a] *University of Luxembourg, 6, rue Coudenhove Kalergi, L-1359, Luxembourg*
[b] *Université de Lorraine, Arts et Métiers ParisTech, LCFC, Metz, F-57000 Metz, France*

ARTICLE INFO

Keywords:
Laser welding
Hardmetal
Steel
Weld quality prediction
Machine learning
High-speed imaging

ABSTRACT

Laser welding of steel and hardmetal presents significant challenges due to their differing material properties. Improper laser welding parameters can result in unstable joints, ultimately leading to reduced mechanical strength of the weld. Therefore, defining an optimal process window is critical to ensuring weld quality. In addition, a continuous process monitoring method like High-Speed Imaging (HSI) is essential in real industrial applications to maintain stability and detect potential defects. Understanding plume dynamics helps identify the most important features of weld quality, but it also provides deeper insight into operational parameters that discriminate different weld types. Analysis of individual image plume frames from HSI reveals distinct statistical features that are identified as unique to each welding condition. Performing systematic feature selection using plume morphology, spatter generation and weld quality, we achieved>95 % leveraging Machine Learning (ML) classifiers. Particularly, Gradient Boosting Classifier (GBC), Linear Discriminant Analysis (LDA), Multinomial Logistic Regression (MNL-LR), Support Vector Machine (SVM), and Random Forest (RF), where the RF obtained >99 % classification accuracy of weld quality. The RF was then used in performing Recursive Feature Elimination (RFE), and with the robustness analysis, we managed to reduce the number of features from forty-nine to nine features while maintaining satisfactory performance (Accuracy = 0.981, F1-score = 0.961, AUROC = 0.997). The position of the weld plume, plume eccentricity and plume width are the most essential features that lead to the improvement of node purity and classification accuracy.

Introduction

Hardmetals have gained significant popularity in cutting applications due to their exceptional performance characteristics (Das et al., 2010), such as wear resistance and high hardness (Levin and Ted Hartwig, 2015; Nogami et al., 2021). However, their high cost and limited ductility often make it impractical for tools to be manufactured entirely from these materials (Ren et al., 2018). Instead, hardmetals are the active part of cutting tools and are often joined to a carrier body, typically composed of steel. As a result, there is a growing demand for advanced joining technologies to join these materials effectively (Mir et al., 2021).

Laser technology is recognized as a precise and efficient process for welding thin sheets and dissimilar materials. This is due to its low Heat-Affected Zone (HAZ), high processing speed, and excellent automation flexibility (Amne Elahi et al., 2023), Katayama, 2020). Laser technology can also be utilized for welding hardmetal (WC—Co binder) and steel. However, the primary challenge in welding these materials arises from their significant differences in thermal expansion coefficients and thermal conductivity. These disparities contribute to the formation of defects, such as microcracks, which can compromise the integrity and mechanical performance of the welded joint (Costa et al., 2003), (Mirski et al., 2016), (Norouzian et al., 2023).

For instance, (Costa et al., 2004, and Quintino, 2006; Costa et al., 2003) investigated steel and hardmetal laser welding. They utilized CO_2 and Nd: YAG laser sources and achieved quality welds with reliable results. Subsequent metallurgical analysis revealed the presence of intermetallic compounds and a high concentration of carbides within the weld seam. These factors contributed to increased embrittlement and a reduction in mechanical strength. Research findings indicate that the selection of laser welding methods, such as continuous or pulsed welding, along with the optimization of key parameters, including beam

---

* Corresponding author.
E-mail address: Hossein.norouzian@uni.lu (M. Norouzian).

https://doi.org/10.1016/j.jajp.2025.100318

spot size, power, velocity, and the laser beam's position relative to the material border in butt welding configuration, is critical for achieving high-quality welds (Xu and Li, 2017) (Miranda et al., 2008). The results illustrate that a larger beam spot size, optimized heat energy input, and laser beam offset can improve the joint's quality by reducing the beam power concentration on the hardmetal side. This adjustment helps minimize the tungsten carbide diffusion into the melt pool, ultimately leading to the enhanced mechanical strength of the weld (Schiry and Plapper, 2019). These studies confirm the critical role of selecting an optimized process window. While the importance of these parameters is well recognized, and several reports demonstrate high-quality mechanical joining of hardmetal and steel, it is essential to consider the entire laser welding process monitoring to ensure a quality weld and crack-free joint.

Various methods have been employed to monitor the laser welding process in recent years. Several types of sensors, including high-speed cameras, acoustic sensors, photodiodes, spectrometers, X-ray transmission imaging systems, etc., have been used to enhance process control and ensure weld quality (Kawahito et al., 2009; Yanxi Zhang et al., 2020; Liu et al., 2023). These sensors monitor variations in characteristic quantities throughout the welding process (Deng et al., 2020; Yanxi Zhang et al., 2019) and analyze the relationship between these changing trends and the resulting welding quality. Vision sensors are considered beneficial among the control devices due to their ability to provide high-quality data and information. High-speed imaging (HSI) can be helpful in laser welding, offering process monitoring and precise weld quality evaluation. Imaging can be an efficient controlling method, providing sufficient data for models to be trained and to classify welds based on their appearance and defect. This approach enables automated quality assessment and can significantly improve the precision and reliability of the welding process (Norouzian et al., 2024).

Melt pool monitoring is a widely used method for understanding the behavior of the laser welding process. (Kim et al., 2021) employed an HSI observation technique to observe the weld pool's top and bottom. In another study (Nam and Ki, 2023) used melt pool images to predict absorptance behavior during laser welding to enhance the understanding of energy absorption dynamics and improve the process control and weld quality.

Although keyhole images provide valuable information about the welding process, recording the melt pool presents challenges due to the high light intensity generated by the plume. A powerful laser illumination system and a specific bandpass wavelength filter for the camera are required to enable accurate image capturing. Thus, plume monitoring can serve as a less complex alternative method while providing valuable insights into the welding process. Since the plume is generated by the interaction between the keyhole front and the laser beam, analyzing its characteristics can offer helpful information about weld stability, penetration depth, and potential defects—for instance, (Wang et al., 2013) studied the characteristics of the plume and its correlation with weld speed, which determines the stability of the weld seam. In another study, (Li et al., 2009) investigated the shape of the plasma plume and its relationship with surface impurities and the flow rate of side-assist gas, both of which can significantly influence weld stability. (You et al., 2014) investigated plume features, such as its size, growth, spatter radius, and ejection direction in high-power laser welding, to monitor welding stability. (Xue et al., 2022) studied plume monitoring in high-speed laser welding and concluded that the plume's behavior is significantly influenced by laser power and welding speed. They were able to extract different plume ejection regimes and predict the occurrence of humping defects in the welding process by analyzing various plume features.

Several studies have implemented Machine Learning (ML) algorithms in classifying weld defects and quality (Zhang et al., 2024; Hake et al., 2023; Vasan et al., 2024). Amongst the various ML models for classification, supervised ML algorithms such as logistic regression, discriminant analysis, decision trees, K-nearest neighbor(KNN), support vector machines (SVMs) and random forests have shown promising results. (Malarvel and Singh, 2021) achieved an accuracy between 95 % and 97 % using a multi-class SVM algorithm on X-ray images of weld conditions. (Mu et al., 2013) reported an accuracy of 90.75 % by combining principal component analysis and SVM.(Chianese et al., 2022) reported an accuracy between 73 % and 93 % for different ML algorithms mentioned earlier. (Asif et al., 2022) achieved a classification accuracy of 83 % using logistic regression algorithms to predict the quality of welds.

In sum, existing studies demonstrate that monitoring effectively assesses weld stability and its association with process parameters. However, monitoring laser welding of dissimilar materials, particularly the joining of tungsten carbide with a low binder percentage and steel, requires further research and in-depth analysis of plume dynamics and weld quality control.

To address this challenge, this study investigates the HSI of plume characteristics to address the challenges associated with laser welding of hardmetals and steel, including weld instability and crack propagation. The key features associated with welding quality are identified using supervised ML algorithms to classify the HSI images. As per the author's knowledge, no study currently available in the existing literature leverages HSI for analyzing plume characteristics in the laser welding of hardmetals and steel using supervised ML algorithms for classification.

The rest of the study is structured as follows. The second section provides a systematic and rigorous description of the data workflow procedure, including raw data extraction, data preprocessing for obtaining feature space, description of ML classifiers, and the optimization and sensitivity analysis of the selected RF classifier. The third section provides a step-by-step description of the obtained results, including image processing, welding quality, and ML performance of various classifiers using different performance metrics, considering slightly imbalanced samples for classification. The last section provides concluding remarks, implications and limitations of the study, followed by recommendations and future research directions.

Research design

Data analysis workflow

The data analysis consists of four individual modules (Fig. 1). The first module (1) *Raw Data* consists of the experimental setup, generation of fault labels, and image processing. The second module (2) *Feature Space,* consists of a raw dataset (D_R) and labels (C_k) obtained from image processing, data wrangling step (e.g., filtering, cleaning) and generating a feature subset. The module (3) *ML classifiers* perform data analysis using ML models: Gradient Boosting Classifier (GBC) (Natekin and Knoll, 2013), Linear Discriminant Analysis (LDA) (Tharwat et al., 2017), Support Vector Machine (SVM) (Yongli Zhang, 2012), Logistic Regression (LR) (Bohning, 1992) and Random Forest (RF) (Breiman, 2001). We explain performance metrics, including feature importance analysis. Lastly, after obtaining the results, we focus on the best-fit classifier, i.e., the ML model, and optimize the model by reducing the number of features per Occam's Razor (Blumer et al., 1987) approach. After obtaining the minimum number of features required without reducing the classification metric (e.g., accuracy), we perform a robustness analysis and discuss individual performance metrics while considering each label.

Experimental setup and procedure

The experiment (Fig. 2) depicts the laser welding configuration. The laser welding procedure uses a 2000 W continuous-wave Nd:YAG laser, Trumpf TrueDisk, where the power represents the average output continuously delivered during welding. The laser source has a 1030 nm wavelength and a 2 mm/mrad beam quality. The laser light is guided through a fiber with a 200 µm diameter to the scanner. Considering the focal distance of 160 mm and the collimator length of 90 mm, this results

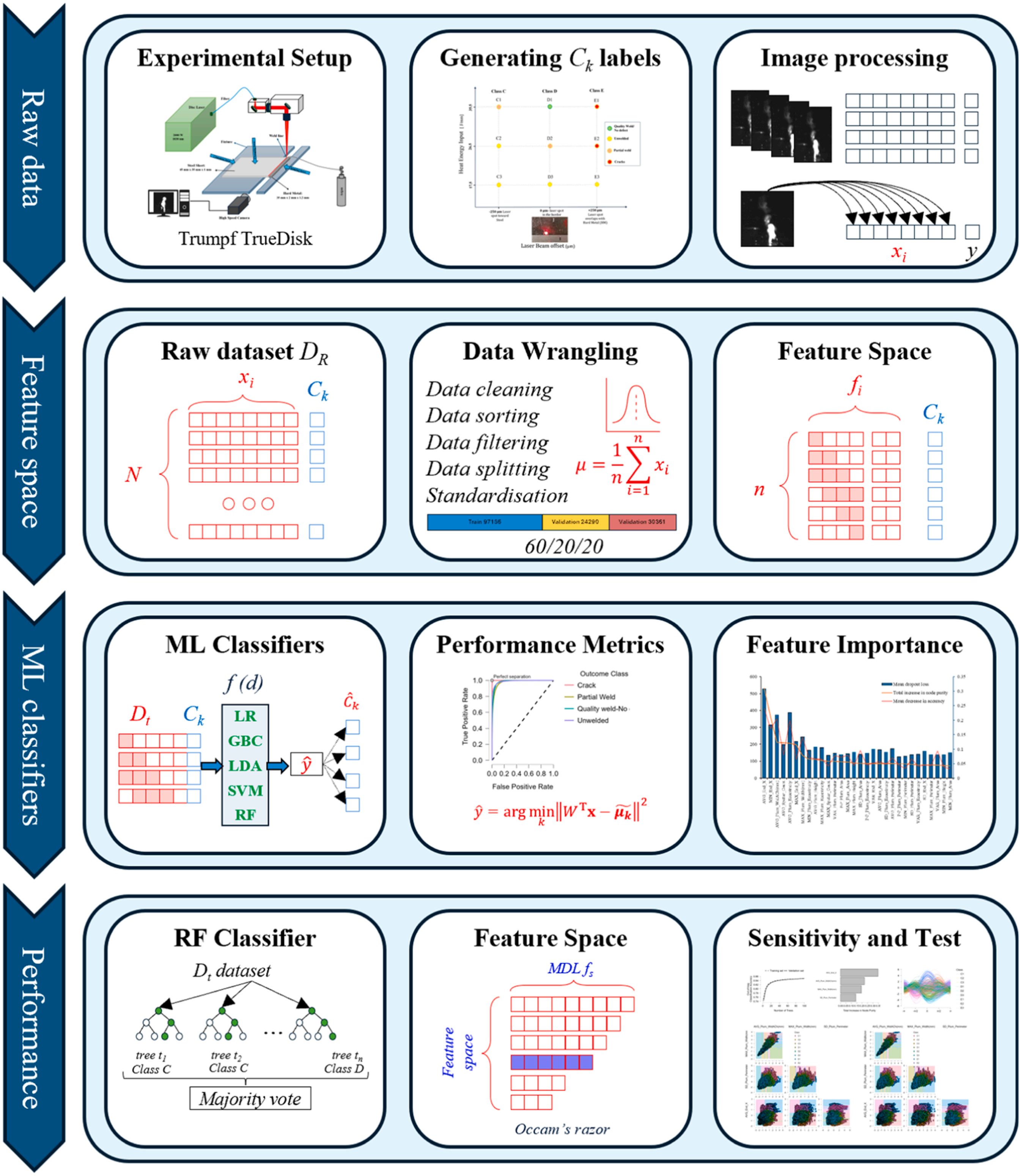


Fig. 1. Data workflow diagram.

in a laser spot diameter of 356 μm. Argon gas is used to flow throughout the welding process to protect the molten zone from oxidation. It is delivered through a gate positioned above the weld pool, angled at 45° for the weld area. A Motion Blitz Cube 4 high-speed camera with a UV filter monitors the welding process. The camera's high frame rate of 11,000 fps and a shutter speed of 1 μs enable the capture of detailed plume images with a resolution of 250 × 448 pixels throughout the procedure, which is crucial for analyzing plume dynamics. Two dissimilar materials, CrMn steel and hardmetal with a 3 % Co binder, are used in the experiment (Schiry and Plapper, 2019). The material properties information is in Table 1.

Both materials are fixed with clamping forces in a butt-welding configuration. The steel sheet has a thickness of 1.0 mm, while the hardmetal piece is 1.2 mm thick. To ensure proper alignment for butt

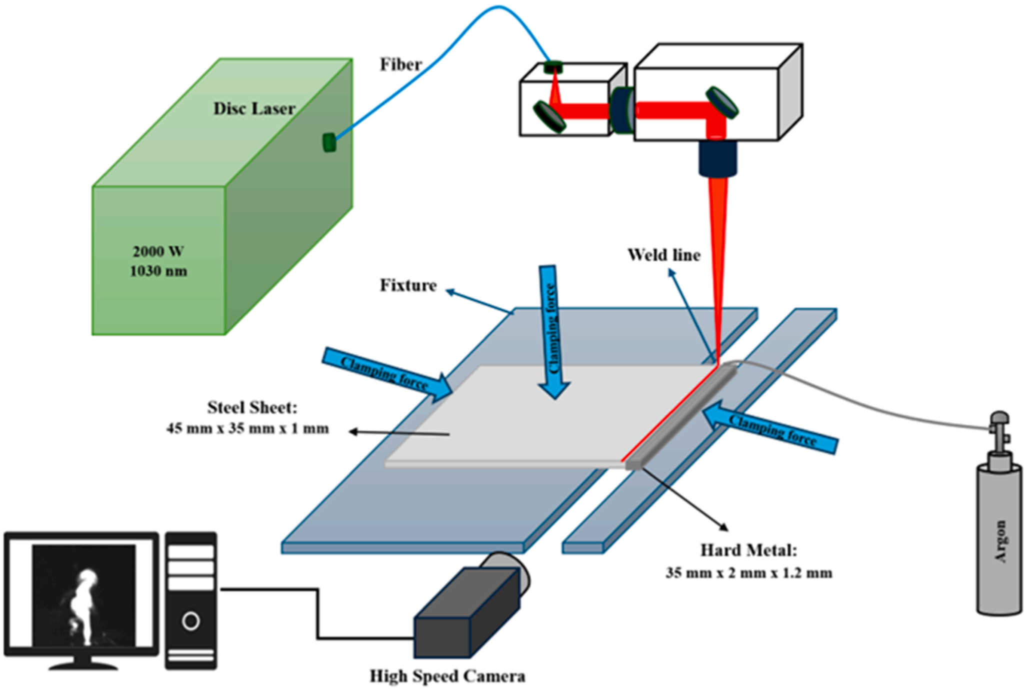


Fig. 2. Laser welding and high-speed camera configuration.

Table 1
Material properties.

Material	Thermal Expansion Coefficient [10^{-6}/K]	Thermal Conductivity [W/m.K]	Hardness [HV10]	Modulus of Elasticity [GPa]
Steel	12	45	255	200
Hardmetal	4.6	95	2020	665

welding, the fixture is designed such that the platform supporting the hardmetal is 0.2 mm lower than that of the steel, allowing the top surfaces of both materials to be level.

The welding procedure uses a linear continuous laser beam with a constant velocity. However, a range of values is considered for other laser parameters, such as laser power and position. Table 2 presents the process parameters and associated class categories. The heat energy input is calculated based on the nominal laser power and welding speed. While the actual energy absorbed by the steel or hardmetal may differ from the calculated values, it is assumed that 100 % absorption efficiency is achieved. This provides a standardized basis for process comparison. The heat energy input is calculated according to the following equation (Tuncel et al., 2023):

$$Q = \frac{P}{V} \tag{1}$$

Table 2
Laser welding process parameters.

Class	Sub-Class	Heat Energy Input [J/mm]	Laser position [μm]
C	C1	35.5	−250
	C2	26.5	
	C3	17.5	
D	D1	35.5	0
	D2	26.5	
	D3	17.5	
E	E1	35.5	+250
	E2	26.5	
	E3	17.5	

Where Q is the heat energy input of laser welding(J/mm), P is the power (W) and V is the welding speed(mm/s).

These process parameters are designed to include potential defects such as micro-cracks, poor joining, lack of fusion in the weld, and good-quality welds. The association between these parameters is defined across three main classes and nine sub-classes. Class D represents the laser position on the steel side, tangent to the border between steel and hardmetal, and is considered the reference laser point (0 μm). Class C corresponds to a welding class in that the laser spot is positioned on the steel side, −250 μm away from the hardmetal border. Finally, Class E indicates welding, where the laser overlaps 250 μm with the hardmetal. Each class is further divided into sub-classes (1, 2, 3), corresponding to different heat energy input levels. Five samples were selected for welding for each sub-class, resulting in 45 welded samples with varying parameters, according to Table 2 for the experiment.

Image preprocessing and feature extraction

The welding zone was isolated by applying Region of Interest (ROI) to remove irrelevant background details. The image processing included five steps. At first, raw images were captured and assessed as valid for preprocessing. Secondly, brightness and contrast were adjusted to discriminate the parts inside the image and reduce the excess light generated by the laser. The images were then converted from RGB to Grayscale using grayscale conversion in OpenCV to reduce computational complexity while maintaining essential intensity details. Lastly, a Gaussian blur filter was used to eliminate unwanted noise and maintain important plume characteristics. Combined with the OpenCV-based processing, the extracted features were used for the classification algorithms.

Five iterations of capturing have been used to ensure consistency and build a robust dataset for each of the nine weld classes. Approximately 9000 images were captured during each iteration, from which an average of 3500 frames were extracted from each section of the sequence. This resulted in a final dataset of approximately 157,000 frames, forming a comprehensive basis for subsequent image processing analysis.

The extraction of features was performed by using plume and spatter characteristics. Plume detection involves isolating high-intensity regions within images to identify the characteristics of plumes. A thresholding technique (Goh et al., 2018) is applied to keep only the brightest regions (intensity range of 250 to 255) for detecting only actual plumes, followed by contour detection to extract geometric properties. A 20-pixel margin was ignored at the bottom of the image to avoid misidentifying reflections as plumes. Spatter detection relied on edge detection methods to identify and isolate individual spatter elements. Additionally, a constraint was set to identify spatters with a size smaller than 50 pixels, ensuring that noise and irrelevant details are minimized. The following flowchart, Fig. 3, represents sequential steps from pre-processing to feature extraction and analysis of each frame.

The following image characteristics are extracted: *Area, Height* and *Width, Perimeter, Eccentricity*, the *X coordinate (End_X)* of the plume, and *Spatter Count*. Specifically, the *Area* is the total pixel count within the detected plume boundary. The *Height* and *Width* are measured using a bounding rectangle around the detected plume boundary. The *Perimeter* was determined by contour tracing to find the perimeter of the plume. *Eccentricity* is calculated using OpenCV's built-in function to derive the plume's second-order moments. Eccentricity values range from 0 for the complete circle to 1 for the highly elongated shape (Gonzalez and Woods, 2018). It computes spatial moments m_{00}, m_{10}, m_{01}, m_{20}, m_{02}, and m_{11} for the detected plume contour. M_{00} represents the total pixel intensity or area of the plume. m_{10}, m_{01} used to calculate the centroid of the object and, m_{20}, m_{02} and m_{11} describe the spread and orientation of the plume.

$$\alpha = \sqrt{\frac{2*\left(m_{20}+m_{02}+\sqrt{\left(m_{20}-m_{02}\right)^2+4m_{11}{}^2}\right)}{m_{00}}} \tag{2}$$

$$\beta = \sqrt{\frac{2*\left(m_{20}+m_{02}-\sqrt{\left(m_{20}-m_{02}\right)^2+4m_{11}{}^2}\right)}{m_{00}}} \tag{3}$$

Using these moments, the major axis (α) and minor axis (β) are derived:

$$Eccentriciy = \sqrt{1-\frac{\beta^2}{\alpha^2}} \tag{4}$$

Next, the *Plume_End_X (Plume X coordinate)* origin identifies the plume's position to track movement on the *X*-axis. The *Spatter Count* demonstrates the number of discrete spatter elements detected using edge detection algorithms. The features, such as area, height, width, and perimeter, are initially calculated based on the number of pixels. These features are then converted from pixel units to millimeters using a scale of 1 pixel = 0.05 mm, determined through the software calibration to make them comprehensible. This calibration utilizes a reference object within the captured images, enabling precise scaling of extracted geometric properties. Fig. 4 depicts the image processing steps and the final extracted raw features.

As a final feature extraction step, we perform feature extraction on the raw data obtained from image preprocessing. The welding plume is highly dynamic and is influenced by the complex interactions between the laser parameters and materials. Due to rapid temporal variations driven by fluctuating process conditions, analyzing individual images in isolation may fail to capture the underlying trends. Thus, incorporating time-domain statistical analysis is crucial to enhancing the interpretability of the extracted features and providing a more comprehensive dataset for ML-based classification. A time-domain statistical feature is extracted by segmenting the image sequence using Time-Sliding Windows (TSW) (Tang et al., 2023; Lv et al., 2023) of 20 consecutive frames to achieve this. The feature selection includes generating a feature subset from raw variables by using a sequence of $n = 20$ TSW to generate a feature subset per Table 3.

Machine learning classification algorithms

Gradient boosting classifier

The GBC algorithm is an ensemble learning algorithm that builds strong classifiers by sequentially adding weak classifiers – typically, decision trees. The model is trained using step-wise optimization. This works by minimizing the loss function with gradient descent. Let D represent the dataset, such that $D = \{(\mathbf{x}_i, y_i)\}$ be a training dataset of N samples, where $\mathbf{x}_i \in \mathbb{R}^d$ is a feature vector of d dimensions, and $y_i \in C_k$, such that $k = \{1\ldots9\}$ classes. The goal of GBC is to learn function $F(\mathbf{x})$:

$$F_k(\mathbf{x}) = \sum_{m=1}^{M} \gamma_{m,k} h_{m,k}(\mathbf{x}), \tag{5}$$

where F_k ($\mathbf{x}$) is the classifier or logit score for class k, $h_{m,k}$ ($\mathbf{x}$) is the weak base learner (typically decision tree), while $\gamma_{m,k}$ is the weight for weak learner determined by the loss function at iteration m. The loss function L for the multi-class classification problem is determined (Hastie et al., 2009):

$$L(y,F) = -\sum_{i=1}^{N}\sum_{k}^{9} \mathrm{I}(y_i = k) log p_k(\mathbf{x}_i), \tag{6}$$

where $\mathrm{I}(y_i = k)$ is an indicator function equals 1 (if true) class and

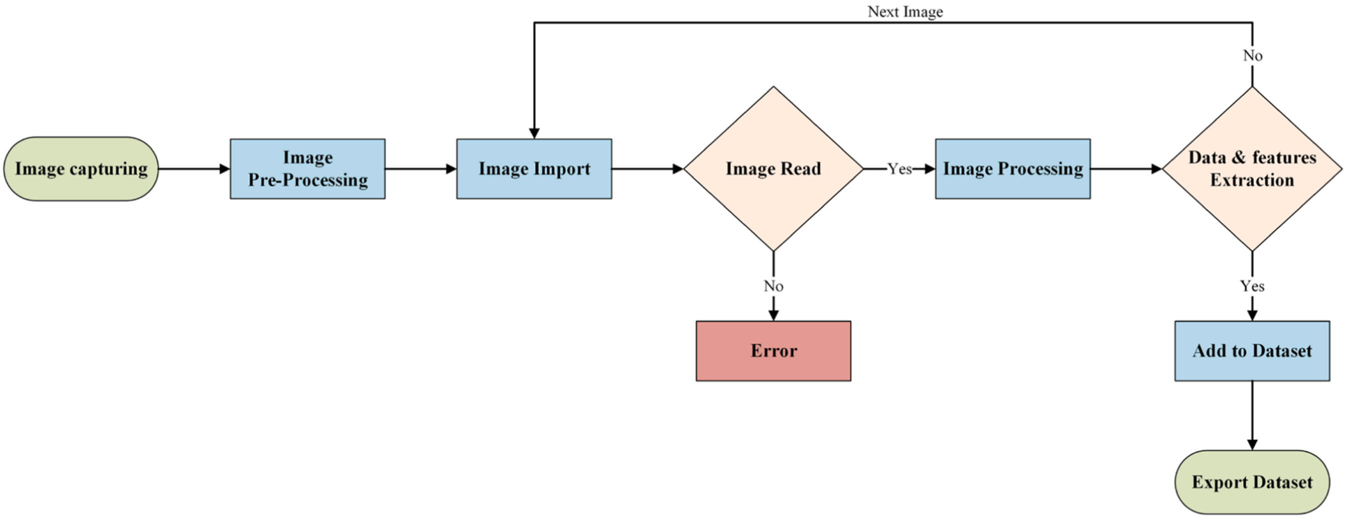


Fig. 3. Image processing flowchart.

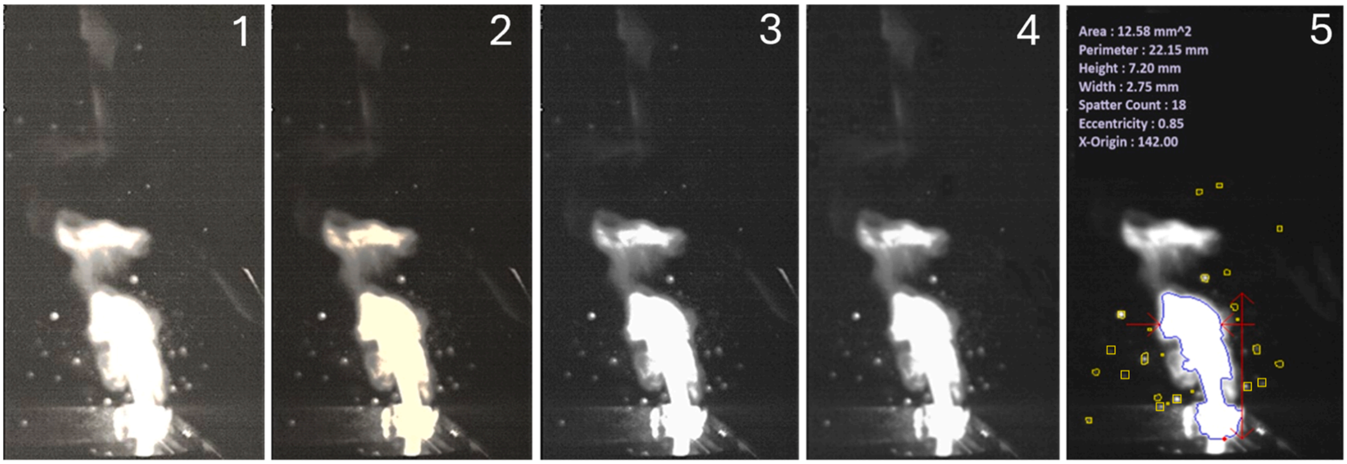


Fig. 4. Filtering and processing images; (1) Raw captured image; (2) Brightness and contrast adjustment; (3) Grayscale conversion; (4) Noise reduction; (5) Raw feature extraction.

Table 3
Time-domain features used for ML classification.

Features generated in TSW	Features*	The equations for generating features
Average (mean) sample value	*AVG_N*	$Sample\ average\ = \frac{1}{n}\sum_{i=1}^{n} N_i$
Sample Standard Deviation	*SD_N*	$Standard\ Deviation\ = \sqrt{\frac{(x_i-\bar{x})^2}{n-1}}$
Minimum value	*Min_N*	$Min(N_i)$
Maximum value	*MAX_N*	$Max(N_i)$
Skewness	*SKEW_N*	$Skewness\ = \frac{n(n+1)}{(n-1)(n-2)}\sum\left(\frac{(x_i-\bar{x})}{\sigma}\right)^3$
Kurtosis	*KURT_N*	$Kurtosis\ = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{(x_i-\bar{x})}{\sigma}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$
Variance sample	*VAR_N*	$Variance\ = \frac{\sum(x_i-\bar{x})^2}{n-1}$

*N = depicts the raw variable expression (e.g., Plume Area, Plume Perimeter, Plume Height, Plume Width, Plume Eccentricity).

0 otherwise, and $p_k(\mathbf{x})$ is the class probability, computed as the function:

$$p_k(\mathbf{x}) = \frac{e^{F_k(\mathbf{x})}}{\sum_{j=1}^{9} e^{F_j(\mathbf{x})}} \quad . \tag{7}$$

Hence, gradient boosting (Friedman, 2002) minimizes the loss function using the gradient of cross-entropy with respect to $F_k(\mathbf{x})$:

$$r_{im,k} = \frac{\partial L(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)}, \tag{8}$$

which is for the softmax multi-classification problem simplified:

$$r_{im,k} = I(y_i = k) - p_k(\mathbf{x}_i), \tag{9}$$

suggesting that the residual for each class k shows the difference between the true class and the predicted probability. The weak learner $h_{m,k}$ $(\mathbf{x})$ is trained to approximate residuals $r_{im,k}$:

$$h_{m,k}(\mathbf{x}) = \operatorname{argmin}_{h} \sum_{i=1}^{N} \left(r_{im,k} - h(\mathbf{x}_i)\right)^2, \tag{10}$$

such that once fitted the optimal weight is determined by minimizing:

$$\gamma_{m,k}(\mathbf{x}) = \operatorname{argmin}_{\gamma} \sum_{i=1}^{N} L(y_i, F_{m-1,k}(\mathbf{x}_i) + \eta\gamma_{m,k} h_{m,k}(\mathbf{x}), \tag{11}$$

where η is the learning rate, controlling the step size in each iteration. Final probability for each class is computed:

$$p_k(\mathbf{x}) = \frac{e^{F_k(\mathbf{x})}}{\sum_{j=1}^{9} e^{F_j(\mathbf{x})}}, \tag{12}$$

where the predicted class label is:

$$\hat{y} = \operatorname{argmax}_{k} p_k(\mathbf{x}) \ . \tag{13}$$

The model tuning includes the following (hyper) parameters: shrinkage = 0.1, interaction depth = 1.0, and minimum observations in node 10. It uses 50 % of training data per tree. The optimization is not fixed and is optimized for a maximum of 100 trees. The loss function is set to Gaussian, and features are scaled with the seed set to 1234. The random seed is set to ensure the reproducibility of the results. The random seed acts as an initialization point for a pseudo-random generator, meaning that whenever the ML algorithm starts with a particular seed, it will lead to the same results. The choice of random seed is arbitrary and can be any integer number.

Linear discriminant analysis

The LDA is a supervised learning method commonly used for classification and dimensionality reduction (Gu et al., 2011). LDA works by projecting data into subspaces to maximize class separability. For estimating within-class separability matrix S_w, the algorithm measures dispersion as:

$$S_W = \sum_{k=1}^{K} \sum_{x \epsilon C_k} (x - \mu_k)(x - \mu_k)^{\mathrm{T}}, \tag{14}$$

where $\boldsymbol{\mu}_k$ is the class mean $\boldsymbol{\mu}_k = 1/n_k\ \Sigma_{x\epsilon Ck}$. The between-class matrix S_B measures the dispersion between class means:

$$S_B = \sum_{k=1}^{K} n_k(\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^{\mathrm{T}}, \tag{15}$$

where the $\boldsymbol{\mu} = 1/N\ \Sigma_k\ n_k\boldsymbol{\mu}_k$ is the global mean, to seek for the optimal projection of matrix $W \epsilon\ \mathbb{R}^{dx(K-1)}$ that maximizes the ratio between-class and within-class probability is to optimize J in the projected subspace as:

$$J(W) = \frac{\det\left(W^{\mathrm{T}} S_B W\right)}{\det\left(W^{\mathrm{T}} S_W W\right)} \quad . \tag{16}$$

The optimal W consists of eigenvectors $S_W^{-1} S_B$ corresponding to the largest eigenvalue. Given rank$(S_B) \leq K\text{-}1$, the LDA reduces dimensionality to $K - 1$ dimensions. Hence, the eigenvalue problem is solved by the generalization:

$$S_B\mathbf{w} = \lambda S_W\mathbf{w}, \tag{17}$$

where eigenvectors $\mathbf{w}_1, \ldots \mathbf{w}_{K\text{-}1}$ form columns of W that are ordered by descending eigenvalues. Next, projecting the data into the LDA subspace is performed:

$$\mathbf{Y} = W^{\mathrm{T}}\mathbf{X}, \tag{18}$$

where $\mathbf{X} \epsilon \mathbb{R}^{d\times N}$ is the data matrix, while projected class means are $\boldsymbol{\mu}_k = W^T \ \boldsymbol{\mu}_k$. Assuming the equal covariance, assigning x to class k is performed:

$$\widehat{y} = \mathrm{arg}\underset{k}{min}\|W^{\mathrm{T}}\mathbf{x} - \widetilde{\boldsymbol{\mu}_k}\|^2 \ . \tag{19}$$

Equivalently, using a discriminant function in the original space, we get:

$$\delta_k(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k^{-1} - \frac{1}{2}\boldsymbol{\mu}_k + \log\pi_k \tag{20}$$

where $\boldsymbol{\Sigma} = S_W/N\text{-}K$ is the pooled covariance estimate, such that $\pi_k = n_k/N$, and $\boldsymbol{\Sigma}^{-1} = (N\text{-}K)S_W^{-1}$. The frequent setbacks in the LDA are the assumptions of normality, i.e., data follows Gaussian (normality) distribution, linear separability and homoscedasticity (i.e., data classes share a common covariance matrix), which leads to misclassification. In setting the algorithm, the estimation method is MLE (Maximum Likelihood Estimate) with scaled features and seed set to 1234.

Multinomial logistic regression

In using LR for classification, there are two primary strategies for generalization of LR to multi-class problems (instead of commonly binary classification) – OvR (One-vs-Rest) and Multinomial (Softmax) LR (Widodo et al., 2023). For more in-depth descriptions, the reader is referred to (Cristea et al., 2024; Gosiewska and Biecek, 2019). However, the OvR is simpler but performs poorer on multi-classes, especially if classes overlap, as in our study. Hence, we focus on the latter. The probability function for each class $k \epsilon \{1, \ldots K\}$, where in our case $K = 9$, estimates the probability that $\mathbf{x}$ belongs to class k:

$$p(y = k|\mathbf{x}, \mathbf{W}) = \frac{e^{\mathbf{w}^{\mathrm{T}}\mathbf{x}}}{\sum_{k=1}^{K} e^{\mathbf{w}_k^{\mathrm{T}}\mathrm{x}}}, \tag{21}$$

where $\boldsymbol{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_k] \epsilon \mathbb{R}^{d\times K}$ represents the weight matrix, while $\mathbf{w}_k \epsilon \mathbb{R}^d$ is the vector for class k. The denominator ensures that the sum of probabilities is 1. The classification boundary in the feature space is determined:

$$\widehat{y} = \mathrm{arg}\underset{k}{max}\mathbf{w}_k^{\mathrm{T}}\mathbf{x} \ . \tag{22}$$

as the largest logit (pre-softmax score). The parameter estimation is determined per log-likelihood function by maximizing the log-likelihood:

$$\mathscr{L}(\mathbf{W}) = \sum_{i=1}^{N}\sum_{k=1}^{K}\mathrm{I}(y_i = k)logP(y_i = k|\mathbf{x}; \mathbf{W}), \tag{23}$$

where $\mathrm{I}(y_i = k)$ is the indicator function (1 if $y_i = k$, 0 otherwise). Equivalently, minimizing the negative log-likelihood (cross-entropy loss) is performed as:

$$J(\mathbf{W}) = -\sum_{i=1}^{N}\sum_{k=1}^{K}\mathrm{I}(y_i = k)\log\left(\frac{e^{\mathbf{W}_k^{\mathrm{T}}\mathbf{x}_i}}{\sum_{j=1}^{K} e^{\mathbf{W}_j^{\mathrm{T}}\mathbf{x}_i}}\right) \ . \tag{24}$$

The gradient loss concerning $\mathbf{w}_k$ is determined:

$$\nabla_{\mathbf{w}_k}J(\mathbf{W}) = -\sum_{i=1}^{N}\mathbf{x}_i[\mathrm{I}(y_i = k) - P(y_i = k|\mathbf{x}_i; \mathbf{W})] \tag{25}$$

which typically minimizes the loss using an iterative optimization method (e.g., gradient descent). Lastly, regularization (L1 or L2) should be used to penalize large weights to prevent overfitting. Nevertheless, as in previous cases, it is essential to note that classification struggles when classes overlap, in addition to the linearity (log-odds) assumption, and class probabilities may be affected. Consistent with previous algorithms, the features are scaled, and the seed is set to 1234.

Support vector machine

Support Vector Machine (SVM) is a supervised learning algorithm widely used for classification tasks, particularly in high-dimensional spaces where it excels at identifying decision boundaries with maximum margin (Zhang et al., 2017). The core principle of SVM is to find a hyperplane that best separates the classes while maximizing the margin, which is the distance between the closest points (support vectors) of different classes. Given a training dataset $D = \{(\mathbf{x}_i, y_i)\}$, such that $i = 1, \ldots N$ where $\mathbf{x}_i \epsilon \mathbb{R}^d$ is a D-dimensional feature vector and $y_i \epsilon \{-1,1\}$ represents the class labels, the goal of SVM is to learn a decision function of the form:

$$f(x) = w^T x + b \tag{26}$$

Where, w is the weight vector, and b is the bias term. The optimization problem for the linear SVM is formulated as:

$$\underset{w,b}{min}\frac{1}{2}||w||^2 \tag{27}$$

subject to constraints:

$$y_i\left(w^T x_i + b\right) \geq 1, \forall i = 1, \ldots, N \tag{28}$$

where the margin is maximized by minimizing $||w||^2$. For non-linearly separable data, SVM employs the kernel trick to project data into a higher-dimensional space, where a linear separator can be found. The most commonly used kernel functions include linear kernel, polynomial kernel, radial bias function kernel, sigmoid kernel where the choice of kernel depends on the data distribution. This study chooses the RBF kernel for its flexibility in capturing complex decision boundaries. The dataset is standardized, and the random seed is set to 1234 to ensure reproducibility. While SVM is robust to overfitting in high-dimensional spaces, it struggles with large datasets due to computational complexity. However, it remains a powerful classifier, especially in settings where decision boundaries are well-defined and class overlap is minimal.

Random forest

To define a problem setup for RF, let's again consider a dataset $D = \{(\mathbf{x}_i, y_i)\}$, such that $i = 1, \ldots N$ where $\mathbf{x}_i \epsilon \mathbb{R}^d$ is a feature vector, and $y_i \epsilon \{1, \ldots K\}$ denotes the class membership. The RF, like the GBC, is an ensemble method that performs bagging (bootstrap aggregation) of multiple DTs to improve generalization and robustness (Breiman, 2001). To perform ensemble construction, we first delve into (bootstrapped) sampling for each tree $t \epsilon \{1, \ldots T\}$ by drawing a bootstrapped sample D_t by randomly selecting N number of samples from D dataset with replacement. An important note here is that there is always a remaining sample not contained in D_t, forming a sample called an OOB (Out-of-Bag) set. The OOB sample is also used here to validate the RF model's validity by measuring classification performance metrics (e.g., accuracy). At each node split during tree construction, a random subset of $m \leq d$ features is considered (m = sqrt(d) is typical for classification). During each split, the goal is to maximize purity, i.e., Gini impurity:

$$G = \sum_{j=1}^{C_k} p_k(1 - p_k), \tag{29}$$

where p_k is the proportion of samples belonging to class k, and Ck is the number of classes (in our case, $k = 9$). After constructing the ensemble T

trees, the final classification is determined through majority voting:

$$\widehat{Y} = \arg \max_{j \in \{1 \ldots k\}} \sum_{t=1}^{T} l(h_t(\mathbf{x}) = j), \tag{30}$$

where $h_t(\mathbf{x})$ represents the prediction of t^{th} tree given X, l is the indicator function if the prediction class is j, and 0 otherwise. To determine feature importance, we rely on Mean Dropout Loss (MDL), Mean Decrease in Accuracy (MDA), and Total Increase in Node Impurity (TINP). The MDL measures the cross-entropy loss with features averaged over 50 permutations. The MDA evaluates the drop in model accuracy with randomly shuffled features, while TINP computes the reduction in Gini across all nodes when a particular feature m is used. Unlike previous methods, the RF has no formal assumptions for feature independence, but correlated features may dilute importance. The RF is robust to noise because of its resilience to irrelevant features due to feature randomization. Lastly, compared to LDA, LR, and SVM, the RF captures complex interactions and depends on the linearity assumption. However, it is less interpretable than LDA, LR, and SVM, but we provide important features in the analysis. The parameters and hyperparameters include training data used per tree at 50 %, features auto split, and maximum optimization trees set to 100. The features are scaled, and the seed is set to 1234. For a detailed description of RFs, the reader is referred to (Louppe, 2014; Breiman, 2001; Kaur and Malhotra, 2008).

Performance metrics

Performance metrics for classification algorithms help evaluate how well a model distinguishes between different classes (Jordan et al., 2006). Precision (P) measures the number of correct instances retrieved, divided by all retrieved instances. Recall (R) measures the number of correct instances retrieved divided by all correct instances. Accuracy is another measurement defined as the proportion of true instances retrieved, both positive and negative, among all instances retrieved. Accuracy is a weighted arithmetic mean of precision and inverse precision.

$$Precision(P) = \frac{TP}{TP + FP} \tag{31}$$

$$Recall(R) = \frac{TP}{TP + FN} \tag{32}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{33}$$

Here, TP is true positive (positive sample correctly classified), FP is false positive (positive outcomes that the model predicted incorrectly), FN is false negative (negative outcome that the model predicted incorrectly) and TN is true negative (negative sample correctly classified).

The F1-score gives the harmonic mean between precision and recall.

$$F_1 - score = 2 \cdot \frac{P \cdot R}{P + R} \tag{34}$$

An F1-score of 1 indicates a perfect balance between precision and recall, meaning the classifier makes no false positives or negatives. Conversely, an F1-score of 0 signifies that the model completely fails to identify the positive class, either by misclassifying all positives or predicting none at all.

The Receiver Operating Characteristic (ROC) curve represents the trade-off between the true and false positive rates. The area under this curve, also known as the Area Under the Receiver Operating Characteristic (AUROC) curve, is generated by adjusting the classification threshold, which influences the balance between true positives and false positives. A score of 1 for AUROC indicates perfect differentiation between positive and negative outcomes, while a score of 0 signifies the complete absence of positive classifications.

Results and discussion

Laser welding

Fig. 5 presents the welding images of all classes, illustrating the effect of laser positioning and heat energy input. Analysis of each weld type confirmed that C3, D3, and E3 sub-classes lack fusion due to insufficient power (17.5 J/mm), making an unachievable joint under these conditions. However, when the heat energy input increases to 26.5 J/mm and 35.5 J/mm, sufficient energy is available to form a keyhole, improving the laser penetration. The laser beam offset also plays a critical role in determining weld quality. The initial results confirm that the precise positioning of the laser beam is a critical parameter in achieving high-quality welds. When the laser beam is shifted toward the hardmetal, cracks appear (classified as E1 and E2). The formation of microcracks occurs when the laser beam interacts with the hardmetal, which has a lower thermal expansion coefficient than steel. Consequently, this disparity in expansion induces stress within the weld zone that leads to crack formation and its propagation to the hardmetal.

Another factor contributing to crack formation is the diffusion of carbon and tungsten from the hardmetal into the melt pool, forming brittle intermetallic compounds. At the steel's liquidus temperature, an interaction occurs between the eutectic WC—Co phase and the infiltrating molten steel, resulting in the decomposition of WC—Co. This process facilitates the diffusion of tungsten and carbon into the weld zone. During solidification, W, Fe, and C form intermetallic compounds, predominantly along the austenite grain boundaries, creating a network-like structure (Fig. 6). Additionally, within the austenite grains—which later transform into martensite—a minor fraction of WC and mixed carbides precipitate (Barbatti et al., 2007). This phenomenon increases the likelihood of cracking when the laser beam interacts predominantly with the hardmetal.

Other sub-classes, like D2 and C1, exhibit unstable welds and inconsistent weld integrity. Full bonding is achieved in certain areas, while in others, the bonding fails. In sub-class C1, the heat energy input is sufficient to generate a hole; however, due to the distance of the beam from the metal, the formation of a high-quality weld is not feasible. Conversely, in sub-class D2, the laser position is well-balanced and tangent to the hard metal border; however, the insufficient heat energy input results in a joint with multiple points of incomplete bonding. Only sub-class D1 produced high-quality welds with optimal fusion and defect-free joints among all tested conditions. With its precise laser positioning and heat energy input of 35.5 J/mm, this sub-class represents a quality weld.

After examining and analyzing the weld quality, the weld types within this process window are classified into four main categories: "Weld with Cracks", "Partial Weld", "Unwelded", and "Quality Weld". Specifically, sub-classes C3, D3, E3, and C2 fall under "Unwelded," while sub-classes C1 and D2 belong to "Partial Weld." Additionally, Classes E1 and E2 are categorized as "Weld with Cracks." Finally, sub-class D1 is identified as a "Quality Weld". Fig. 7 depicts the weld types in their correlation with process parameters. This process window provides a comprehensive understanding of the key challenges of welding steel and hardmetal. Process parameters' influence is crucial, and this window offers valuable insights into optimizing these parameters for improved weld quality and stability.

Although the process window has been established and the key parameters influencing quality welds and defect formation have been identified, process monitoring remains critical to ensure the final product's quality without relying on visual inspection or random quality control methods. This underscores the importance of using a high-speed camera to capture images during welding, enabling a detailed analysis of the welding process and defect detection throughout the process. The results presented in Fig. 7 serve as a reference, derived from the experimental studies, for the next step of image processing and ML. The results establish a foundation for understanding the association between

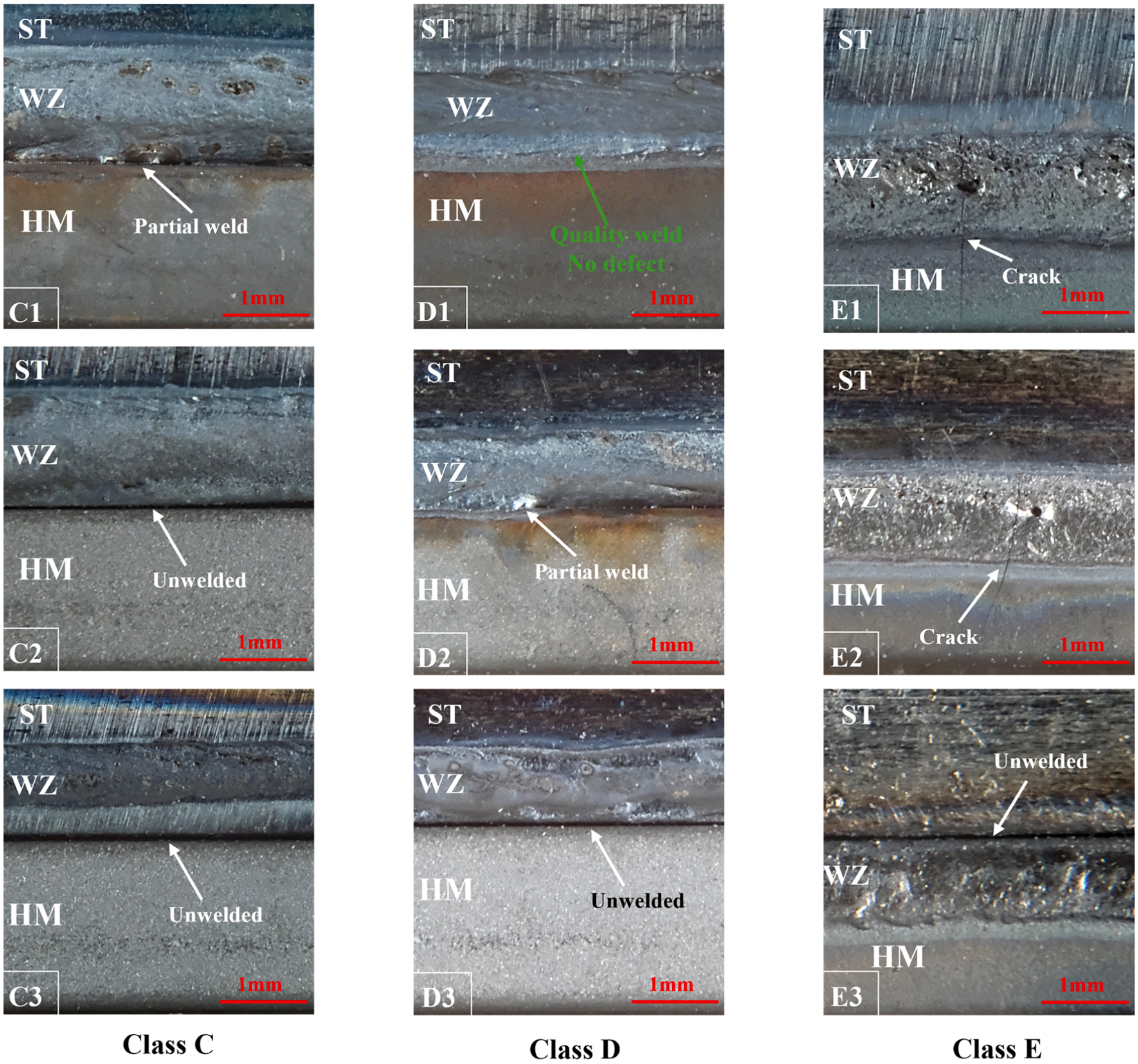


Fig. 5. Class C (left): No joining between Hardmetal (HM) and Steel (ST) in C2 and C3. Partial weld in C1; Class D (Middle): D3 has no joining, D2 welded partially and D1 is welded completely without any defects in the weld zone (WZ); Class E (Right): E3 is unwelded, E2 and E1 have cracks.

welding parameters, the extracted plume and spatter image features during the process, and the final joint quality.

Plume and spatter high-speed imaging

The high-speed camera captured plume images throughout the welding process. Fig. 8 presents the plume images for each class. The initial analysis of HSI and plume feature extraction from sub-class E1 and E2 welds—where heat energy input is elevated and the laser beam overlaps with the hardmetal—reveals a significant presence of spatter. The increased vaporization and keyhole instability lead to molten metal ejection, resulting in more spatter. The more substantial vapor-induced recoil pressure within the keyhole further disrupts the molten pool. In these particular sub-classes, where the laser is shifted toward the hardmetal, due to the lower melting and vaporizing temperature point of cobalt as the binder than tungsten carbide, the explosive vaporization of this element generates a spatter, which likely contains tungsten carbide particles.

In contrast, this behavior is not observed in other sub-classes, such as D1, C1, where the same heat energy input levels are applied. This observation suggests that increased spatter generation may serve as an early indicator of welds with microcracks. Although many spatters can indicate crack propagation and "Weld with Cracks" category, it cannot be the only extracted feature for identifying other weld types. This suggests that additional features from plume must be extracted for different classes to characterize the weld type accurately.

Analyzing plume images across different sub-classes reveals distinct features associated with varying heat energy inputs. A long, nearly straight plume is observed at a low heat energy input of 17.5 J/mm (sub-classes C3, D3, and E3), indicating a controlled evaporation process. In contrast, as the laser energy input increases (sub-classes C1, D1, E1), the plume size becomes shorter and more turbulent, likely due to higher vaporization rates and increased material ejection. This plume behavior serves as another distinguishing factor among different weld types. Insufficient energy input prevents keyhole formation, leading to an "Unwelded" classification, where no bonding occurs. In this scenario, plume height is a reliable indicator for identifying a lack of fusion, as these cases exhibit significantly longer plume heights than other weld types.

Since the welding process demands high precision and laser beam positioning can significantly impact the outcome, making the accurate detection of quality welds challenging. For instance, sub-classes C1 (partial weld) and D1 (quality weld) exhibit similar plume images, complicating their differentiation. In this context, among the extracted image processing features, the X-coordinate position of the plume, "*Plume_End_X*" feature, has proven particularly effective in classifying D1

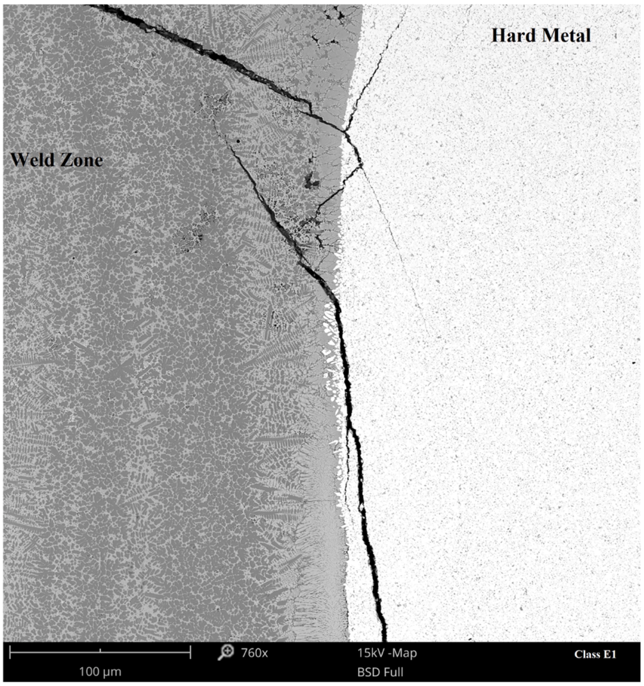


Fig. 6. SEM image of cross-section of the sub class E1 weld joint with cracks – Weld Zone exhibiting a network-like structure at the Hardmetal interface.

as a quality weld and distinguishing it from C1. The differences in plume behavior across each class provide valuable insights into the quality of the resulting welds. However, a comprehensive analysis of all frames throughout the welding process is required to fully understand the plume dynamics for each sub-class and extract the relevant features. To achieve this, image processing techniques are employed to capture the distinctive features of each plume and correlate them with the corresponding weld types.

After capturing images for each sub-class, the image processing phase extracts relevant plume and spatter features. This process results in a large dataset comprising 150,000 images, yielding approximately 1050,000 numerical feature values. These features and datasets enable the application of ML classification models capable of predicting the final weld type and identifying defects like cracks based on plume and spatter images, forming a foundation for future real-time monitoring systems. The following section discusses various ML models trained on these datasets in detail.

Machine learning classification results

The classification results (Fig. 9) suggest the following. The overall classification accuracy shows that RF (0.98), followed by SVM (0.89) and GBC (0.86), are the best-performing classifiers. However, given the slightly imbalanced dataset where most of the dataset contains information about "Unwelded" (44.5 %) labels, followed by "Quality weld" (11 %), "Weld with Crack" (22.4 %) and "Partial Weld" (22.1 %), we relied on ROC (Receiver Operating Characteristic) and Area Under ROC as it adds more validity in discussing the classification results. Hence, the results of AUROC (Fig. 9A) shows that in all cases, the classification of "Partial weld" had the lowest performance as it overlaps with "Quality weld" and "Unwelded" labels. Based on the analysis, it can be seen that LR (i.e., Multinomial Regression – MNR) performs the worst, primarily due to the assumption of normality and linear dependency among predictors, which can also be said for LDA and SVM. However, MNR relies on probability estimates, unlike LDA and SVM, which perform classification on linear discrimination and hyperplane separation, respectively.

This also stands for "Quality weld", as it mostly overlaps with "Partial weld", and the classification performance is consistent with other classifiers. The "Unwelded" class seems to be second to best classification, although it can since several samples may have contributed to better separation from different courses. The "Weld with Crack" performs best across all models with little to no overlap with other categories. The analysis is confirmed visually in terms of the consistency of the class separability via ROC curves (Fig. 9A), which again proves the high classification accuracy of "Weld with Crack", while showing poor separability and overlap of "Partial weld" and "Quality weld". Lastly, the RF shows perfect separability of "Weld with Crack" (100 %) and almost perfect separability of the "Unwelded" class (99 %) while showing a bit

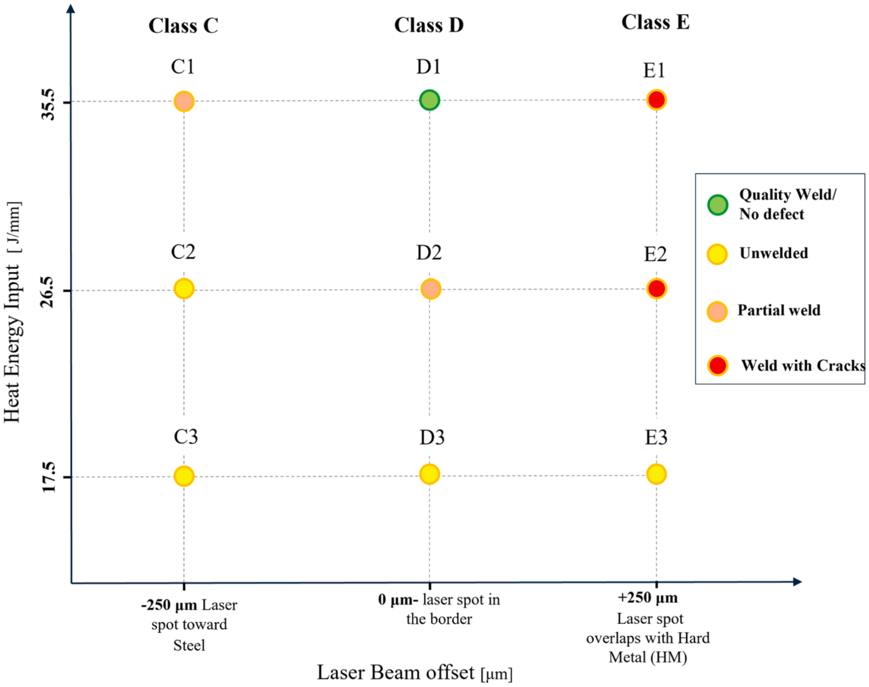


Fig. 7. The association between weld types and process parameters.

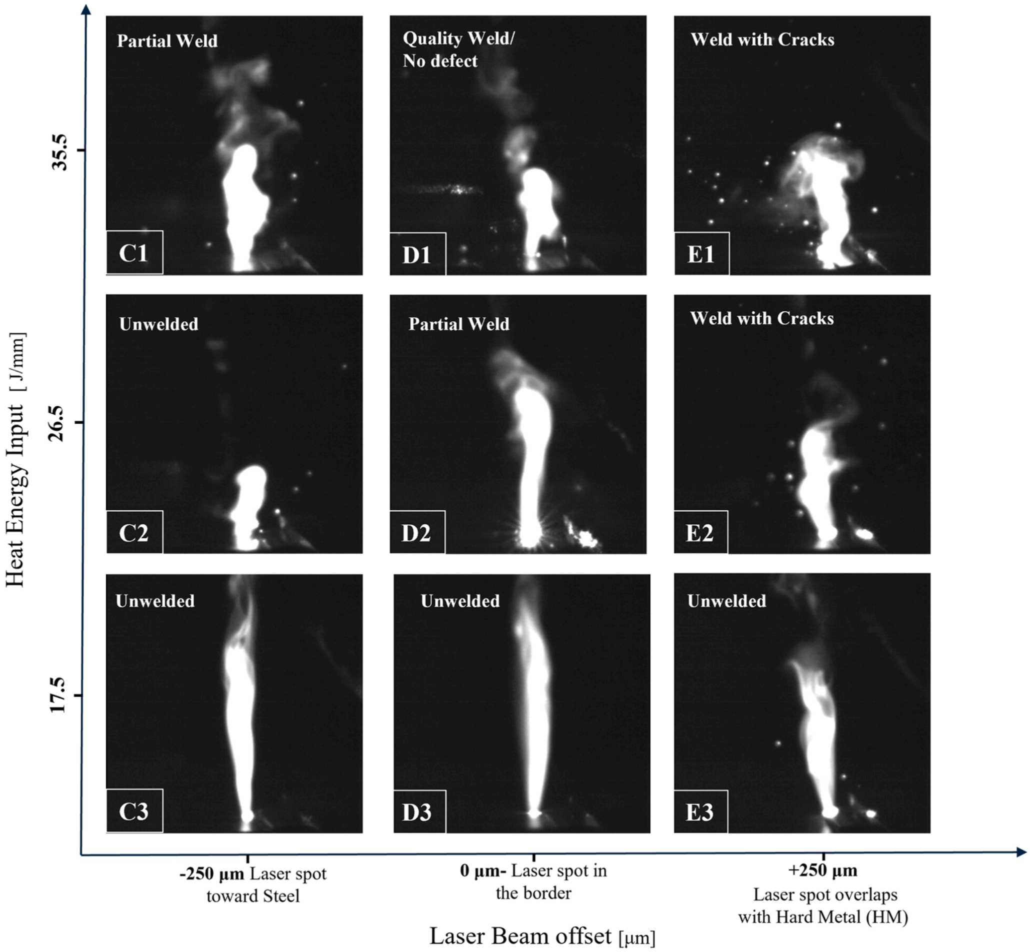


Fig. 8. Images of plumes for different sub-classes.

of overlap in "Partial Weld" and "Quality weld" classes, obtaining 96.8 % and 98.0 % prediction accuracy, respectively.

Finally, given that RF performs best, we focus on analyzing the underlying mechanisms leading to a high model classification. To do so, we first investigate the feature importance analysis using MDL, MDA and TINP results (Fig. 9C). The results show that AVG_End_X has the highest contribution (MDL = 280.31) to the classification accuracy. In contrast, MIN_End_X has the highest contribution in terms of a decrease in the accuracy of prediction (MDA = 0.289), suggesting that the position of the laser beam significantly affects the detection of "Weld with Crack" and separation of classes as it determines the outcome of the welding process. In addition, AVG_Plum_Width (MDL = 223.705) and AVG_Plum_Eccentricity (MDL = 263.092) also significantly contribute to the classification accuracy. Lastly, given that the idea is to follow Occam's Razor approach, i.e., to maintain the least number of predictors while maintaining the classification accuracy, we perform optimization of features by Recursive Feature Elimination (RFE) by iteratively removing features without losing too much information and performance of the model.

The feature elimination process included eliminating features with the least MDL contribution to the model performance. This included removing features until we optimised the solution to maintain the 99 % AUROC metric. The results (Table) show the process of eliminating features starting from all included features (*RF_All*) until we manage to come to the ten features (*RF_10*). The reason for stopping at 10 features is that the model has significantly reduced the prediction accuracy and AUROC. However, we also added the RF_9 model (with nine features). Still, the process of obtaining the nine most essential features is explained in the following subsection since it did not include the process of removing the features based on the feature importance metrics (i.e., MDL) but rather removing redundant features based on the correlation score, thus avoiding collinearity of the features.

Table 4.

In the following, we explain the reduction of feature space based on the collinearity of features used for ML models. The analysis leveraged the correlation matrix, mainly relying on Pearson's *r* correlation. In addition, we provide additional insights into changes in feature behavior by reporting TINP and MDA. Besides, we report on out-of-bag results, which contain the subsample left out from the bagging, i.e., bootstrapping of samples to show the rigour and validity in predicting unseen samples. Alongside this, we report ROC results, including all performance metrics that could interest the reader.

Optimization and robustness

The Pearson's correlation matrix (Fig. 10A) reports several potential collinearities among the features used. Firstly, we start with the correlation between AVG_Plum_Perimeter and AVG_Plum_Height ($r = 0.947$, $p < 0.001$). The underlying reason for such a correlation is that the plum perimeter is obtained from the measures of height and width, which again is closely related to AVG_Plum_Area, as they closely explain the geometric characteristics of the plume in the image processing. However, given that the results are obtained from a data-driven approach, ignoring the underlying dependencies, it becomes reasonable why different features contribute to the classification accuracy. Nevertheless, after removing the feature of AVG_Plum_Perimeter – selected based on the lower MDL score – the model still preserves 99 % prediction accuracy based on AUROC = 0.997 and some lower Accuracy = 98.1. In fact,

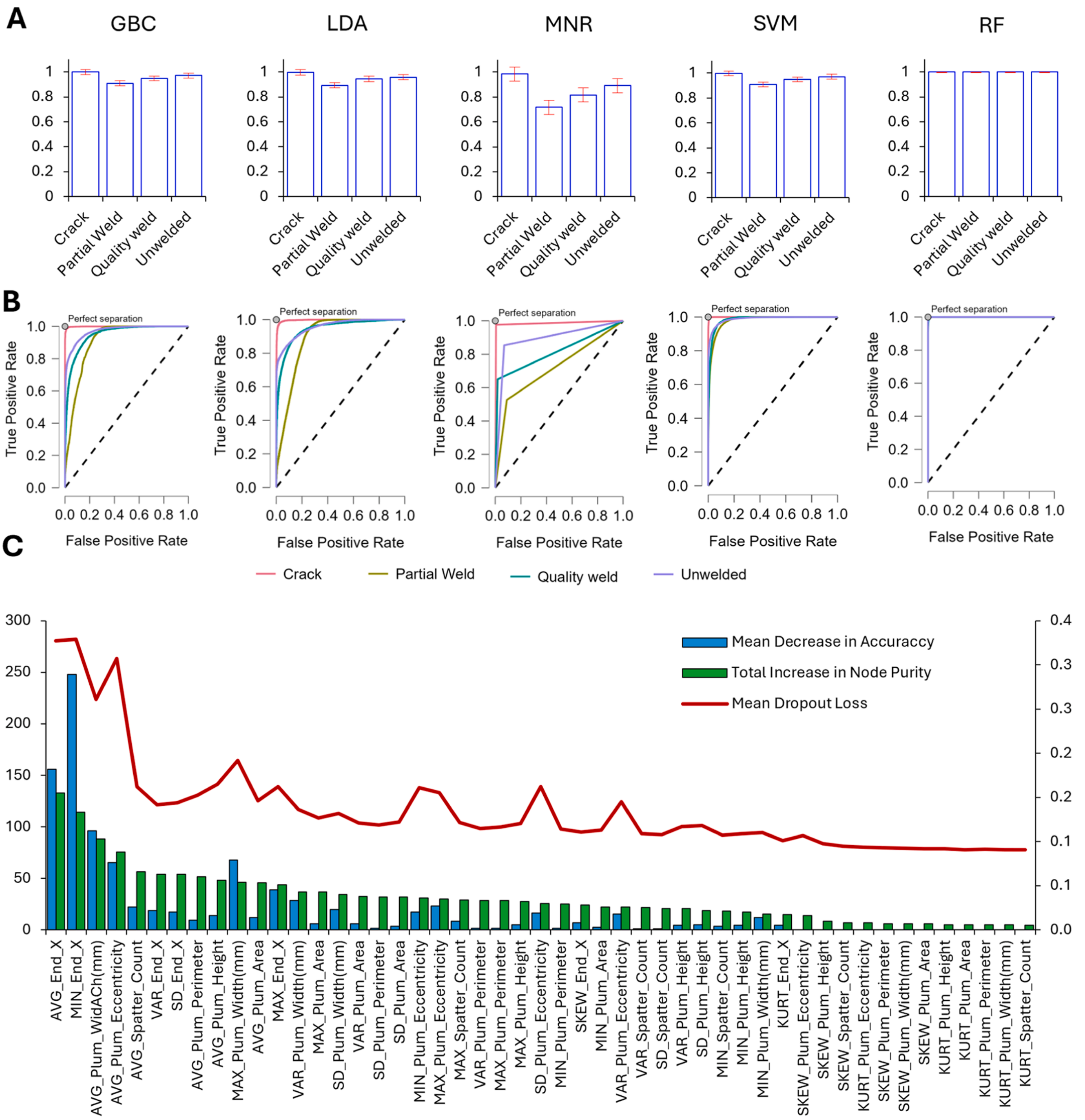


Fig. 9. Performance of ML Classification considering (A) Training AUROC characteristics with confidence intervals. (B) Receiver Operating Characteristic Curve of ML classifiers. (C) The plot of feature importance from RF is the best-performing model, where MDL is represented on the y_1-axis, and MDA and TINP on the y_2-axis are concerned with features (*X*-axis).

during the training process (Fig. 10F), the AUROC resulted in 1.00, i.e., >99.9 % AUROC and 98 % prediction accuracy across label categories, suggesting that it led to the improvement of the model with fewer features. However, after perturbations and manually performing the analysis of different scenarios by removing highly correlated features, the prediction results led to a high drop in prediction accuracy (in some cases <80 %), and the final decision was made to preserve the model of nine features (Fig. 10B and C), containing: AVG_End_X, Min_End_X, AVG_Plum_Eccentricity, AVG_Plum_Width(mm), AVG_Plum_Height, SD_Plum_Width(mm), VAR_End_X, AVG_Plum_Area and AVG_Scatter_Count.

Conclusions

Concluding remarks

Laser beam welding of steel to hardmetal with a low binder content enables defect-free joining. The study results indicate that improper optimization of process parameters (e.g., heat energy input and laser beam position) can lead to defects such as unstable welds or crack formation. This underscores the presence of a narrow process window for the welding process, emphasizing the importance of non-destructive condition monitoring in ensuring weld quality. HSI revealed that

Table 4
Performance metrics after the recursive feature elimination process.

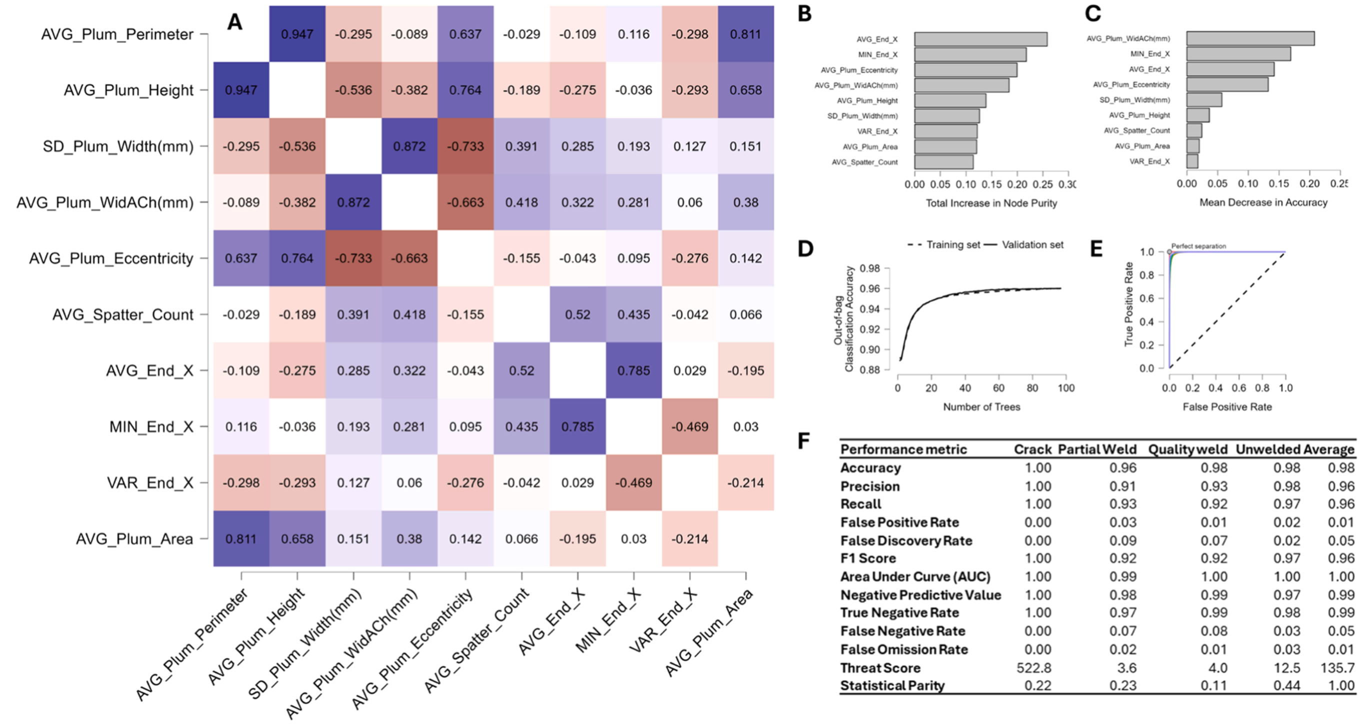
	Performance Metrics	Weld with Crack	Partial Weld	Quality weld	Unwelded	Average
RF_All	Accuracy	1.000	0.989	0.994	0.993	0.994
	Precision (Positive Predictive Value)	1.000	0.968	0.980	0.994	0.988
	Recall (True Positive Rate)	1.000	0.982	0.965	0.991	0.988
	F1 Score	1.000	0.975	0.972	0.993	0.988
	Area Under Curve (AUROC)	1.000	0.999	0.999	1.000	1.000
RF_37	Accuracy	1.000	0.990	0.994	0.994	0.995
	Precision (Positive Predictive Value)	1.000	0.973	0.979	0.995	0.989
	Recall (True Positive Rate)	1.000	0.983	0.969	0.992	0.989
	F1 Score	1.000	0.978	0.974	0.994	0.989
	Area Under Curve (AUROC)	1.000	0.999	0.999	1.000	1.000
RF_30	Accuracy	1.000	0.990	0.995	0.994	0.995
	Precision (Positive Predictive Value)	1.000	0.973	0.982	0.994	0.990
	Recall (True Positive Rate)	1.000	0.983	0.972	0.992	0.990
	F1 Score	1.000	0.978	0.977	0.993	0.990
	Area Under Curve (AUROC)	1.000	0.999	0.999	1.000	1.000
RF_25	Accuracy	1.000	0.990	0.995	0.994	0.995
	Precision (Positive Predictive Value)	1.000	0.974	0.979	0.994	0.989
	Recall (True Positive Rate)	1.000	0.982	0.973	0.992	0.989
	F1 Score	1.000	0.978	0.976	0.993	0.989
	Area Under Curve (AUROC)	1.000	0.999	0.999	1.000	1.000
RF_20	Accuracy	1.000	0.985	0.992	0.991	0.992
	Precision (Positive Predictive Value)	1.000	0.960	0.970	0.992	0.984
	Recall (True Positive Rate)	1.000	0.975	0.958	0.987	0.984
	F1 Score	1.000	0.967	0.964	0.990	0.984
	Area Under Curve (AUROC)	1.000	0.999	0.999	0.999	0.999
RF_15	Accuracy	1.000	0.980	0.990	0.987	0.989
	Precision (Positive Predictive Value)	1.000	0.945	0.968	0.988	0.979
	Recall (True Positive Rate)	1.000	0.967	0.944	0.982	0.979
	F1 Score	1.000	0.956	0.956	0.985	0.979
	Area Under Curve (AUROC)	1.000	0.998	0.998	0.999	0.999
RF_12	Accuracy	1.000	0.973	0.986	0.982	0.985
	Precision (Positive Predictive Value)	0.999	0.928	0.945	0.984	0.971
	Recall (True Positive Rate)	1.000	0.950	0.928	0.976	0.971
	F1 Score	1.000	0.939	0.937	0.980	0.971
	Area Under Curve (AUROC)	1.000	0.996	0.997	0.998	0.998
RF_10	Accuracy	1.000	0.963	0.982	0.975	0.980
	Precision (Positive Predictive Value)	0.999	0.907	0.922	0.976	0.960
	Recall (True Positive Rate)	1.000	0.927	0.910	0.968	0.960
	F1 Score	0.999	0.917	0.916	0.972	0.960
	Area Under Curve (AUROC)	1.000	0.993	0.996	0.997	0.996
RF_9	Accuracy	1.000	0.964	0.983	0.976	0.981
	Precision (Positive Predictive Value)	0.999	0.909	0.927	0.977	0.961
	Recall (True Positive Rate)	1.000	0.931	0.917	0.968	0.961
	F1 Score	0.999	0.920	0.922	0.972	0.961
	Area Under Curve (AUROC)	1.000	0.993	0.996	0.997	0.997

spatter counts can be an early indicator of crack formation. Additionally, a relatively long and straight plume is characteristic of low heat energy input, contributing to unstable weld formation. Consequently, plume and spatter images provide valuable insight into weld quality, which is supported by features such as spatter count, plume area, height, width, eccentricity, position, and plume perimeter.

By applying image processing techniques, quantitative data can be extracted from each frame of the laser welding process, enabling the development of ML classification models. Hence, the results of ML classification demonstrate that the RF model can accurately classify weld quality with 99 % accuracy based on the extracted features from captured images. This result is better than the results that have been reported in the literature for ensemble-supervised ML algorithms. Nevertheless, given that the time sliding window procedure is used to generate feature space using time-domain statistical features, the feature space may be difficult to adapt to practical scenarios, given that extensive data processing is required. Even so, after optimizing the feature space, we reduced the number of feature sets from 49 to only nine of the most significant features while maintaining performance accuracy. Among these features, the average X-coordinate of the plume position (*AVG_End_X*) has the most considerable impact both in improving classification accuracy and node impurity in the RF classifier, suggesting it contains critical information about laser beam positioning and serves as a dominant factor in predicting weld quality.

Implications and limitations

While laser welding parameters and processed windows are well-defined, uncontrollable factors such as material variability and differences in surface conditions, composition, or thermal properties affect laser absorption and heat distribution, which leads to defects and poor-quality welding. Process-related factors, including joint fit-up accuracy and clamping forces, can change the weld quality even when the parameters remain constant. This underscores the need for adaptive control mechanisms and advanced sensing technologies to further enhance process stability in mass production. By classifying weld types based on plume dynamics and their characteristics, this approach provides non-



	AVG_Plum_Perimeter	AVG_Plum_Height	SD_Plum_Width(mm)	AVG_Plum_WidACh(mm)	AVG_Plum_Eccentricity	AVG_Spatter_Count	AVG_End_X	MIN_End_X	VAR_End_X	AVG_Plum_Area
AVG_Plum_Perimeter		0.947	-0.295	-0.089	0.637	-0.029	-0.109	0.116	-0.298	0.811
AVG_Plum_Height	0.947		-0.536	-0.382	0.764	-0.189	-0.275	-0.036	-0.293	0.658
SD_Plum_Width(mm)	-0.295	-0.536		0.872	-0.733	0.391	0.285	0.193	0.127	0.151
AVG_Plum_WidACh(mm)	-0.089	-0.382	0.872		-0.663	0.418	0.322	0.281	0.06	0.38
AVG_Plum_Eccentricity	0.637	0.764	-0.733	-0.663		-0.155	-0.043	0.095	-0.276	0.142
AVG_Spatter_Count	-0.029	-0.189	0.391	0.418	-0.155		0.52	0.435	-0.042	0.066
AVG_End_X	-0.109	-0.275	0.285	0.322	-0.043	0.52		0.785	0.029	-0.195
MIN_End_X	0.116	-0.036	0.193	0.281	0.095	0.435	0.785		-0.469	0.03
VAR_End_X	-0.298	-0.293	0.127	0.06	-0.276	-0.042	0.029	-0.469		-0.214
AVG_Plum_Area	0.811	0.658	0.151	0.38	0.142	0.066	-0.195	0.03	-0.214	

Performance metric	Crack	Partial Weld	Quality weld	Unwelded	Average
Accuracy	1.00	0.96	0.98	0.98	0.98
Precision	1.00	0.91	0.93	0.98	0.96
Recall	1.00	0.93	0.92	0.97	0.96
False Positive Rate	0.00	0.03	0.01	0.02	0.01
False Discovery Rate	0.00	0.09	0.07	0.02	0.05
F1 Score	1.00	0.92	0.92	0.97	0.96
Area Under Curve (AUC)	1.00	0.99	1.00	1.00	1.00
Negative Predictive Value	1.00	0.98	0.99	0.97	0.99
True Negative Rate	1.00	0.97	0.99	0.98	0.99
False Negative Rate	0.00	0.07	0.08	0.03	0.05
False Omission Rate	0.00	0.02	0.01	0.03	0.01
Threat Score	522.8	3.6	4.0	12.5	135.7
Statistical Parity	0.22	0.23	0.11	0.44	1.00

Fig. 10. The analysis includes (A) a Decision boundary matrix, (B) a Total Increase in Node Purity, (C) a Mean Decrease in Accuracy, (D) OOB Results, (E) a Receiver Operating Characteristic Curve, and (F) RF performance metrics for each individual label.

contact monitoring insights into welding, enabling more effective defect detection and quality assessment for this sensitive weld joint.

The limitations of the study are the following. Extensive data (pre) processing is required to obtain high classification accuracy. This considers generating a feature subspace from raw processed data using HSI. Even so, after receiving a feature subset, the transformation from raw variables to a feature set is required using time-sliding windows to generate statistical feature space (e.g., mean, standard deviation, variance) for performing ML classification. Hence, this certainly does not downplay the obtained results, as, in fact, we add a more nuanced understanding of the underlying features affecting the weld quality using rigorous HSI and data analysis procedures leveraging ML algorithms.

Future research

Future research will explore advanced monitoring techniques using more sensors, closed-loop feedback systems for real-time non-destructive condition monitoring, and process optimization methods to enhance the stability and reliability of laser welding in these material joints. As a part of our future research endeavours, we are considering utilizing different process parameters to obtain more information about weld quality. Also, another study will include more advanced ML algorithms that will consist of Deep Learning (DL) techniques, such as Convolutional Neural Networks, Deep Belief Networks, and Recurrent Neural Networks, to overcome the step of extensive image and data (pre) processing.

CRediT authorship contribution statement

Mohammadhossein Norouzian: Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Mahan Khakpour:** Writing – review & editing, Software, Investigation, Conceptualization. **Marko Orosnjak:** Writing – review & editing, Visualization, Methodology, Formal analysis, Data curation. **Atal Anil Kumar:** Writing – review & editing, Validation, Formal analysis. **Slawomir Kedziora:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Acknowledgment

The presented work is based on the Project " BRIGHT" (AFR-PPP grant, Reference 16663291). The authors would like to thank the support of the Luxembourg National Research Fund (FNR) and acknowledge Ceratizit Luxembourg as the project’s industrial partner.

Appendix A. Random Forest Cross Classification Matrix

Fig. A1.

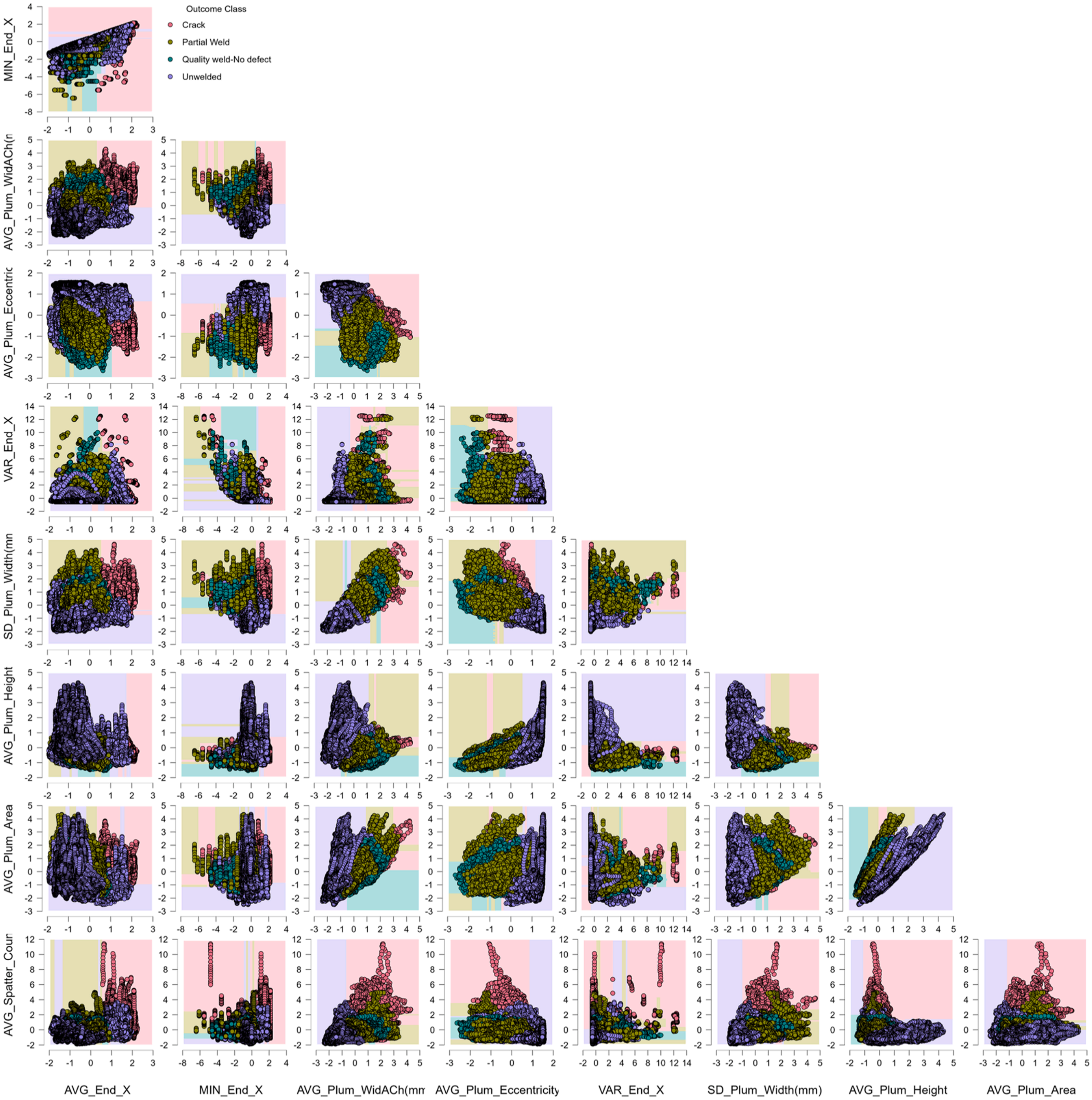


Fig. A1. RF Decision boundary matrix of reduced feature space.

Data availability

The data that has been used is confidential.

References

Asif, Kaiser, Zhang, Lu, Derrible, Sybil, Indacochea, J.Ernesto, Ozevin, Didem, Ziebart, Brian, 2022. Machine learning model to predict welding quality using air-coupled acoustic emission and weld inputs. J. Intell. Manuf. 33 (3), 881–895. https://doi.org/10.1007/s10845-020-01667-x.

Barbatti, C., Garcia, J., Liedl, G., Pyzalla, A., 2007. Joining of cemented carbides to steel by laser beam welding. Materwiss. Werksttech. 38 (11), 907–914. https://doi.org/10.1002/mawe.200700196.

Blumer, Anselm, Ehrenfeucht, Andrzej, Haussler, David, Warmuth, Manfred K., 1987. Occam's razor. Inf. Process. Lett. 24 (6), 377–380. https://doi.org/10.1016/0020-0190(87)90114-1.

Bohning, Dankmar., 1992. Multinomial logistic regression algorithm. Ann. Inst. Stat. Math. 44 (1), 197–200. https://doi.org/10.1007/BF00048682.

Breiman, Leo., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Chianese, Giovanni, Franciosa, Pasquale, Sun, Tianzhu, Ceglarek, Dariusz, Patalano, Stanislao, 2022. Using photodiodes and supervised machine learning for automatic classification of weld defects in laser welding of thin foils copper-to-steel battery tabs. J. Laser Appl. 34 (4). https://doi.org/10.2351/7.0000800.

Costa, A., Miranda, R.M., Quintino, L., 2006. Materials behavior in laser welding of hardmetals to steel. Mater. Manuf. Process. 21 (5), 459–465. https://doi.org/10.1080/10426910500471458.

Costa, A., Quintino, L., Miranda, R.M., 2004. Microstructural aspects of laser dissimilar welds of hard metals to steels. J. Laser Appl. 16 (4), 206–211. https://doi.org/10.2351/1.1809635.

Costa, Alexandra P., Quintino, Luísa, Greitmann, Martin, 2003. Laser beam welding hard metals to steel. J. Mater. Process. Technol. 141 (2), 163–173. https://doi.org/10.1016/S0924-0136(02)00954-8.

Cristea, Daniela-Maria, Ioan Sima, and László-Barna Iantovics. 2024. "How good perform logistic regression algorithm for complex gastroenterological image analysis. Comparative analysis with Physicians Performace." https://doi.org/10.20944/preprints202410.0683.v1.

Das, Jiten, Rao, G.Appa, Pabi, S.K., 2010. Microstructure and mechanical properties of tungsten heavy alloys. Mater. Sci. Eng.: A 527 (29–30), 7841–7847. https://doi.org/10.1016/j.msea.2010.08.071.

Deng, Fuqin, Huang, Yongshen, Lu, Song, Chen, Yingying, Chen, Jia, Feng, Hua, Zhang, Jianmin, et al., 2020. A multi-sensor data fusion system for laser welding process monitoring. IEEe Access. 8, 147349–147357. https://doi.org/10.1109/ACCESS.2020.3015529.

Friedman, Jerome H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.

Goh, Ta Yang, Basah, Shafriza Nisha, Yazid, Haniza, Safar, Muhammad Juhairi Aziz, Saad, Fathinul Syahir Ahmad, 2018. Performance analysis of image thresholding: otsu technique. Measurement 114 (January), 298–307. https://doi.org/10.1016/j.measurement.2017.09.052.

Gonzalez, Rafael C., Woods, Richard E., 2018. Digital Image Processing. Pearson.

Amne Elahi, Mahdi, Norouzian, Mohammadhossein, Plapper, Peter. 2023. Improving the absorption of copper for near-infrared laser beams to optimize laser spot welding quality of copper to aluminum. https://hdl.handle.net/10993/55724.

Gosiewska, Alicja, and Przemyslaw Biecek. 2019. "Do not trust additive explanations," March.

Gu, Quanquan, Zhenhui Li, and Jiawei Han. 2011. "Linear discriminant dimensionality reduction." In, 549–64. https://doi.org/10.1007/978-3-642-23780-5_45.

Hake, C., Omlor, M., Breitbarth, A., Notni, G., Dilger, K., 2023. Artificial intelligence methods for In-process high-speed image analysis in laser beam welding of hairpins. IOP Conf. Ser.: Mater. Sci. Eng. 1296 (1), 012007. https://doi.org/10.1088/1757-899X/1296/1/012007.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-84858-7.

Katayama, Seiji., 2020. Fundamentals and Details of Laser Welding. Springer Singapore, Singapore. https://doi.org/10.1007/978-981-15-7933-2.

Kaur, Arvinder, Malhotra, Ruchika, 2008. Application of random forest in predicting fault-prone classes. In: 2008 International Conference on Advanced Computer Theory and Engineering. IEEE, pp. 37–43. https://doi.org/10.1109/ICACTE.2008.204.

Kawahito, Y., Mizutani, M., Katayama, S., 2009. High quality welding of stainless steel with 10 KW High power fibre laser. Sci. Technol. Weld. Join. 14 (4), 288–294. https://doi.org/10.1179/136217108×372531.

Kim, Hyeongwon, Nam, Kimoon, Oh, Sehyeok, Ki, Hyungson, 2021. Deep-learning-based real-time monitoring of full-penetration laser keyhole welding by using the synchronized coaxial observation method. J. Manuf. Process. 68, 1018–1030. https://doi.org/10.1016/j.jmapro.2021.06.029. August.

Levin, Zachary S., Hartwig, K.Ted, 2015. Hardness and microstructure of tungsten heavy alloy subjected to severe plastic deformation and post-processing heat treatment. Mater. Sci. Eng.: A 635, 94–101. https://doi.org/10.1016/j.msea.2015.02.025. May.

Li, Guohua, Cai, Yan, Wu, Yixiong, 2009. Stability information in plasma image of high-power CO2 laser welding. Opt. Lasers. Eng. 47 (9), 990–994. https://doi.org/10.1016/j.optlaseng.2009.04.013.

Liu, Guiqian, Zhang, Zhanhui, Wang, Honghai, Gui, Yan, Huang, Xuefei, Li, Yanfeng, Tan, Yicheng, 2023. Influence of laser welding defocus and penetration monitoring based on advanced optical sensors. Opt 280 (June). https://doi.org/10.1016/j.ijleo.2023.170811.

Jordan, M., J. Kleinberg, B. Schölkopf., 2006. Pattern recognition and machine learning.

Louppe, Gilles. 2014. "Understanding random forests: from theory to practice," July.

Lv, Jiajun, Lang, Xiaolei, Xu, Jinhong, Wang, Mengmeng, Liu, Yong, Zuo, Xingxing, 2023. Continuous-time fixed-lag smoothing for LiDAR-inertial-camera SLAM. IEEE/ASME Trans. Mechatron. 28 (4), 2259–2270. https://doi.org/10.1109/TMECH.2023.3241398.

Malarvel, Muthukumaran, Singh, Harjeet, 2021. An autonomous technique for weld defects detection and classification using multi-class support vector machine in X-radiography image. Opt 231 (April), 166342. https://doi.org/10.1016/j.ijleo.2021.166342.

Mir, Fayaz Ahmad, Khan, Noor Zaman, Parvez, Saad, 2021. Recent advances and development in joining ceramics to metals. Mater. Today: Proc. 46, 6570–6575. https://doi.org/10.1016/j.matpr.2021.04.047.

Miranda, R.M., Quintino, L., Costa, A., Pina, J.C.P., Rosa, T., Catarino, P., Rodrigues, J.P., 2008. Analysis of different laser welding processes for joining hardmetals to steel. Weld. World 52 (7–8), 42–51. https://doi.org/10.1007/BF03266651.

Mirski, Zbigniew, Granat, Kazimierz, Stano, Sebastian, 2016. Possibilities of laser-beam joining cemented carbides to steel. Weld. Int. 30 (3), 187–191. https://doi.org/10.1080/09507116.2014.937620.

Mu, Weilei, Gao, Jianmin, Jiang, Hongquan, Wang, Zhao, Chen, Fumin, Dang, Changying, 2013. Automatic classification approach to weld defects based on PCA and SVM. Insight - Non-Destr. Test. Cond. Monit. 55 (10), 535–539. https://doi.org/10.1784/insi.2012.55.10.535.

Nam, Kimoon, Ki, Hyungson, 2023. One camera-based laser keyhole welding monitoring system using deep learning. J. Manuf. Process. 104 (October), 17–27. https://doi.org/10.1016/j.jmapro.2023.08.056.

Natekin, Alexey, Knoll, Alois, 2013. Gradient boosting machines, a tutorial. Front. Neurorobot. 7. https://doi.org/10.3389/fnbot.2013.00021.

Nogami, Shuhei, Hasegawa, Akira, Fukuda, Makoto, Rieth, Michael, Reiser, Jens, Pintsuk, Gerald, 2021. Mechanical properties of tungsten: recent research on modified tungsten materials in Japan. J. Nucl. Mater. 543 (January), 152506. https://doi.org/10.1016/j.jnucmat.2020.152506.

Norouzian, Mohammadhossein, Elahi, Mahdi Amne, Koch, Marcus, Zaeem, Reza Mahin, Kedziora, Slawomir, 2024. Data-driven analysis of surface roughness influence on weld quality and defect formation in laser welding of Cu–Al." *Proceedings of the Institution of Mechanical Engineers*. L: J. Mater.: Des. Appl. December. https://doi.org/10.1177/14644207241236138.

Norouzian, Mohammadhossein, Amne Elahi, Mahdi, Plapper, Peter, 2023. A review: suppression of the solidification cracks in the laser welding process by controlling the grain structure and chemical compositions. J. Adv. Join. Process. Elsevier B.V. https://doi.org/10.1016/j.jajp.2023.100139.

Ren, Chai, Fang, Z.Zak, Koopman, Mark, Butler, Brady, Paramore, James, Middlemas, Scott, 2018. Methods for improving ductility of tungsten - A review. Int. J. Refract. Met. Hard Mater. 75 (September), 170–183. https://doi.org/10.1016/j.ijrmhm.2018.04.012.

Tang, Jun, Banghua He, Bing Zhou, Bin Yi, Yanchao Yin, and Xiaobao Liu. 2023. "Process quality prediction method based on CNN_LSTM Hybrid neural network model." In, 621–32. https://doi.org/10.1007/978-981-19-9338-1_74.

Schiry, Marc, Plapper, Peter, 2019. Elucidation of influencing parameters of the laser butt welding process of dissimilar steel to tungsten alloy sheets. https://hdl.handle.net/10993/39387.

Tharwat, Alaa, Gaber, Tarek, Ibrahim, Abdelhameed, Ella Hassanien, Aboul, 2017. Linear discriminant analysis: a detailed tutorial. AI Commun. 30 (2), 169–190. https://doi.org/10.3233/AIC-170729.

Tuncel, Oguz, Aydin, Hakan, Davut, Kemal, 2023. Effect of heat input on HAZ softening in Fiber laser welding of 22MnB5 steel. Opt. Laser. Technol. 164 (September), 109560. https://doi.org/10.1016/j.optlastec.2023.109560.

Vasan, Vinod, Venkatesh Sridharan, Naveen, Balasundaram, Rebecca Jeyavadhanam, Vaithiyanathan, Sugumaran, 2024. Ensemble-based deep learning model for welding defect detection and classification. Eng. Appl. Artif. Intell. 136 (October), 108961. https://doi.org/10.1016/j.engappai.2024.108961.

Wang, Teng, Gao, Xiangdong, Seiji, Katayama, 2013. Analysis of laser-induced plume during disk laser welding at different speeds. Plasma Sci. Technol. 15 (8), 821–824. https://doi.org/10.1088/1009-0630/15/8/20.

Widodo, Slamet, Ermatita, Deris Stiawan, Wayan Widi Pradnyana, I., 2023. The one-vs-rest method for a multilabel patent classification machine learning approach using a regression model. In: 2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS). IEEE, pp. 688–693. https://doi.org/10.1109/ICIMCIS60089.2023.10348993.

Xu, Peiquan, and Leijun Li. 2017. "Fiber laser welding of WC-Co and carbon steel dissimilar materials." https://www.researchgate.net/publication/312901977.

Xue, Boce, Chang, Baohua, Du, Dong, 2022. Monitoring of high-speed laser welding process based on vapor plume. Opt. Laser. Technol. 147 (March). https://doi.org/10.1016/j.optlastec.2021.107649.

You, Deyong, Gao, Xiangdong, Katayama, Seiji, 2014. Monitoring of high-power laser welding using high-speed photographing and image processing. Mech. Syst. Signal. Process. 49 (1–2), 39–52. https://doi.org/10.1016/j.ymssp.2013.10.024.

Zhang, Chongsheng, Liu, Changchang, Zhang, Xiangliang, Almpanidis, George, 2017. An up-to-date comparison of State-of-the-art classification algorithms. Expert. Syst. Appl. 82 (October), 128–150. https://doi.org/10.1016/j.eswa.2017.04.003.

Zhang, Jiayi, Ce, Yin, Xu, Yiyang, Swee Leong, Sing, 2024. Machine learning applications for quality improvement in laser powder bed fusion: a State-of-the-art review. Int. J. AI Mater. Des. 1 (1), 26. https://doi.org/10.36922/ijamd.2301.

Zhang, Yanxi, Gao, Xiangdong, You, Deyong, Zhang, Nanfeng, 2019. Data-driven detection of laser welding defects based on real-time spectrometer signals. IEEe Sens. J. 19 (20), 9364–9373. https://doi.org/10.1109/JSEN.2019.2927268.

Zhang, Yanxi, You, Deyong, Gao, Xiangdong, Wang, Congyi, Li, Yangjin, Gao, Perry P., 2020. Real-time monitoring of high-power disk laser welding statuses based on deep learning framework. J. Intell. Manuf. 31 (4), 799–814. https://doi.org/10.1007/s10845-019-01477-w.

Zhang, Yongli. 2012. "Support vector machine classification algorithm and its application." In, 179–86. https://doi.org/10.1007/978-3-642-34041-3_27.