AST RJ Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2026, 20(9), 327–342
https://doi.org/10.12913/22998624/224563
ISSN 2299-8624, License CC-BY 4.0

Received: 2026.05.11
Accepted: 2026.07.10
Published: 2026.08.01

Mapping of stability zones in Metal Active Gas surfacing using current-voltage and acoustic signatures for weld bead quality assessment

Przemysław Frankiewicz[1], Michał Bembenek[1*]

[1] Faculty of Mechanical Engineering and Robotics, AGH University of Krakow, A. Mickiewicza 30, 30-059 Krakow, Poland
* Corresponding author's e-mail: bembenek@agh.edu.pl

ABSTRACT

This study tests whether defective metal active gas (MAG) surfacing weld beads form distinct mechanism-specific clusters in multimodal sensor space, with implications for machine learning (ML) pipeline design. A synchronized acquisition platform integrating current and voltage signals at 10 kS/s, airborne acoustic sensing at 48–192 kHz, and high-resolution post-bead imaging was developed and validated, then used in a three-phase campaign yielding 105 weld beads on S355 steel. Current-based descriptors strongly separate acceptable and defective beads, with Cohen's $|d| = 1.34$ for the current coefficient of variation and 1.42 for the spike rate. The arc-stability response to voltage trim is strongly asymmetric, with current variability changing by a factor of 6.6 between trim = −3 V and +3 V. Defective beads form two distinct clusters in PCA and t-SNE projections, corresponding to electrical instability and insufficient material deposition. A baseline ML evaluation across five classifiers reached macro-F1 = 0.735 for the fused configuration. The findings provide quantitative pilot-scale evidence that defective weld beads do not form a homogeneous class, that the borderline class is similarly heterogeneous, and that supervised-learning pipelines for weld-bead quality assessment should be designed with mechanism-aware labels and sensor coverage rather than aggregated quality categories. The reported baseline macro-F1 = 0.735 should therefore be read as evidence that the extracted descriptors carry useful information for downstream supervised learning, not as a final model ranking.

Keywords: MAG surfacing, weld bead quality, current–voltage signals, acoustic monitoring, multimodal sensing, machine learning, sensor fusion.

INTRODUCTION

Quality assessment in arc welding and surfacing is still commonly based on post-process inspection methods, such as visual examination, radiographic testing, ultrasonic testing, or destructive evaluation [1]. Although these methods are effective, they provide delayed feedback and do not directly support data-driven process control or machine-learning-based quality prediction [2]. For this reason, in-process monitoring using current, voltage, acoustic, and visual signals has become a direction in automated weld quality assessment [3, 4]. Machine-learning models for weld quality classification require sensor features that are physically meaningful, repeatable, and separable between quality classes [5, 6]. Current and voltage signals are directly related to arc stability, short-circuiting, and metal transfer, while acoustic signals may reflect arc disturbances, droplet transfer, spatter, and other dynamic process phenomena [7, 8]. Optical data can additionally support visual assessment of bead geometry and surface condition [9]. However, before such signals are used for supervised learning, it is necessary to verify that the assumed quality labels correspond to distinct, consistent regions in the feature space. A key limitation of many classification-oriented studies is that defective welds are often treated as one homogeneous class. This assumption is convenient for binary or multi-class machine learning, but it may be physically incorrect. Different process deviations, such as insufficient material deposition, excessive short-circuiting, unstable arc behavior, or geometric irregularities, can result in the same final label while producing different signal signatures [10]. As a result, a classifier trained only on the final quality class may learn dataset-specific correlations instead of mechanism-related process patterns. This study addresses this problem through an experiment on Metal Active Gas (MAG) surfacing of S355 steel. A multimodal acquisition stand was developed to record synchronized current, voltage, acoustic, and visual data during automated bead deposition. The experimental campaign included stand validation, parameter exploration, and design-of-experiments refinement, yielding 105 weld beads. The aim was to map process stability zones and to evaluate whether the resulting sensor features provide a suitable basis for machine-learning-based weld-bead quality assessment [11]. The contribution of this work is a mechanism-oriented analysis of data for subsequent ML classification. The study identifies current–voltage and acoustic descriptors that separate acceptable, borderline, and defective beads, examines the asymmetric response to voltage trim, and shows that defective beads do not form a single uniform class in the feature space. These results provide a basis for developing more robust, mechanism-aware machine-learning models for weld quality assessment.

METHODOLOGY

Measurement stand and synchronization validation

The experimental stand was designed for automated MAG surfacing of S355 steel. It included a Kemppi X8 400 A power source operating in synergic mode with G3Si1 (EN ISO 14341-A; formerly designated SG2) solid wire of 1.2 mm diameter and M21 shielding gas (82% Ar + 18% CO_2, EN ISO 14175) [12]. Torch motion was controlled by a numerically controlled station using G-code commands sent from the host computer. Beads were deposited on structural steel S355J2+N (normalising-rolled) plates of nominal dimensions 150 × 150 × 10 mm. All plates originated from a single heat (Heat No. 433763, batch H087446, QM batch 040011659892) produced by ArcelorMittal Poland S.A. in a basic-oxygen process and supplied with an EN 10204 type 3.1 inspection certificate (yield ReH = 399 MPa, tensile Rm = 543 MPa, elongation A5 = 28.2 %, Charpy KV at -20 deg C = 181 J average; CEV approx. 0.40). The plates were delivered in the as-delivered (+N, normalising-rolled), non-pickled state (plain black surface with mill scale); each plate was therefore surface-ground and degreased with Clino-3 solvent before any bead was deposited, providing a repeatable, oxide-free working surface. The same operator carried out the preparation for all plates, so the macroscopic surface condition is held constant across the campaign. The acquisition system recorded welding current, arc voltage, acoustic signal, RGB video, post-bead images, and plate temperature [13]. Current and voltage were sampled at 10 kHz using LEM transducers and a National Instruments USB-6341 unit [14]. Airborne acoustic signals were recorded using a miniDSP UMIK-2 sound-pressure measurement microphone at 48 kHz during stand validation and parameter exploration (Plates P0-P7) and at 192 kHz for the final DOE campaign (Plates P8-P13), which provided all data used in the quantitative acoustic and machine-learning analyses, while visual data were acquired using an OAK-D Pro Wide PoE camera. The main acquisition channels are summarized in Table 1.

The remaining welding-procedure parameters were held constant throughout the campaign and are summarised in Table 2. The contact-tip-to-work distance (CTWD), shielding-gas flow rate, torch work and travel angles, and electrode polarity were set on the bench, monitored by the operator, and unchanged between plates. Heat input was derived from the recorded current-voltage waveforms and travel speed; at nominal settings (WFS = 5.0 m/min, voltage trim = 0 V, travel speed = 600 mm/min) it averaged approximately 0.3 kJ/mm. The acoustic channel was sampled at 48 kHz during stand validation and parameter exploration (Plates P0-P7) and at 192 kHz for the final DOE campaign (Plates P8-P13); the three acoustic descriptors are RMS-envelope features evaluated over 100 ms windows and are therefore insensitive to this sampling-rate difference. Airborne acoustic sensing was selected over contact acoustic-emission transducers because it is non-intrusive and does not require coupling to the workpiece; the trade-off is sensitivity to ambient noise, addressed by the SNR descriptor.

Table 1. Signal acquisition channel configuration

Channel	Sensor / device	Range	Sample rate / fps	Resolution
Welding current	LEM HT 400-SBD (Hall)	±400 A	10 kHz	16-bit
Welding voltage	LEM CV3-200	0–200 V	10 kHz	16-bit
Airborne acoustic signal	miniDSP UMIK-2	calibrated SPL	48 kHz (P0-P7) / 192 kHz (P8-P13)	24-bit
RGB video and high-resolution image	OAK-D Pro Wide	1920 × 1080	30 fps	8-bit/channel
Plate temperature	BOSCH GIS 1000C	−40 to +1000 °C	manual, 3 points	0.1 °C

Table 2. Welding-procedure parameters held constant throughout the campaig

Parameter	Value
Wire designation	G3Si1 (EN ISO 14341-A; formerly SG2)
Wire diameter	1.2 mm
Shielding gas	M21 (82% Ar + 18% CO2, EN ISO 14175)
Gas flow rate	14 L/min (16 mm nozzle; matched to WFS and travel speed)
Contact-tip-to-work distance (CTWD)	15 mm
Torch work angle	90 deg (perpendicular to the workpiece)
Torch travel angle	0 deg (neutral, no push or pull)
Electrode polarity	DCEP
Metal-transfer mode	synergic (Kemppi X8 synergic curve for G3Si1 1.2 mm / M21)
Travel speed (nominal)	600 mm/min
Nominal heat input	approx. 0.3 kJ/mm (derived from recorded I, U and travel speed)

All modalities were synchronized in the Python acquisition environment using a high-resolution monotonic timestamp source. Recording began approximately 1 second before the welding command, providing a pre-buffer for post-hoc alignment. The synchronization was intended for bead-level feature extraction and ML analysis, not for droplet-level event matching. Timing performance was validated on Plate P0 using three nominal weld beads. All validation beads passed the diagnostic procedure: the NI-audio start offset ranged from 120 to 350 ms, audio timestamp jitter was 5.0 ms, the NI signal continuity ratio was 1.0000, and no NI samples, audio blocks, or buffer overflows were lost. The validation metrics are summarized in Table 3 and illustrated in Figure 1. These results confirm that the stand was suitable for preliminary multimodal data collection and subsequent feature analysis.

Experimental design

The experiment was organized in three stages. Phase I consisted of stand validation on Plate

Table 3. Synchronization and acquisition validation metrics measured on Plate P0

Metric	Unit	Measured value	Tolerance
Inter-modal start offset (NI–audio)	ms	120–350 (mean 227)	< 1000
Audio block timestamp jitter, σ	ms	5.0	< 10
NI–audio duration mismatch	ms	0–700	< 1000
Video frame rate (RGB packets/duration)	fps	30.0	30 ± 1
NI dropped samples (over 3 beads)	count	0	0
Audio dropped blocks (over 3 beads)	count	0	0
Audio buffer overflows (over 3 beads)	count	0	0
NI signal continuity ratio	—	1.0000	> 0.99
Baseline RMS (no arc, autozero applied)	A	0.315	< 1.0

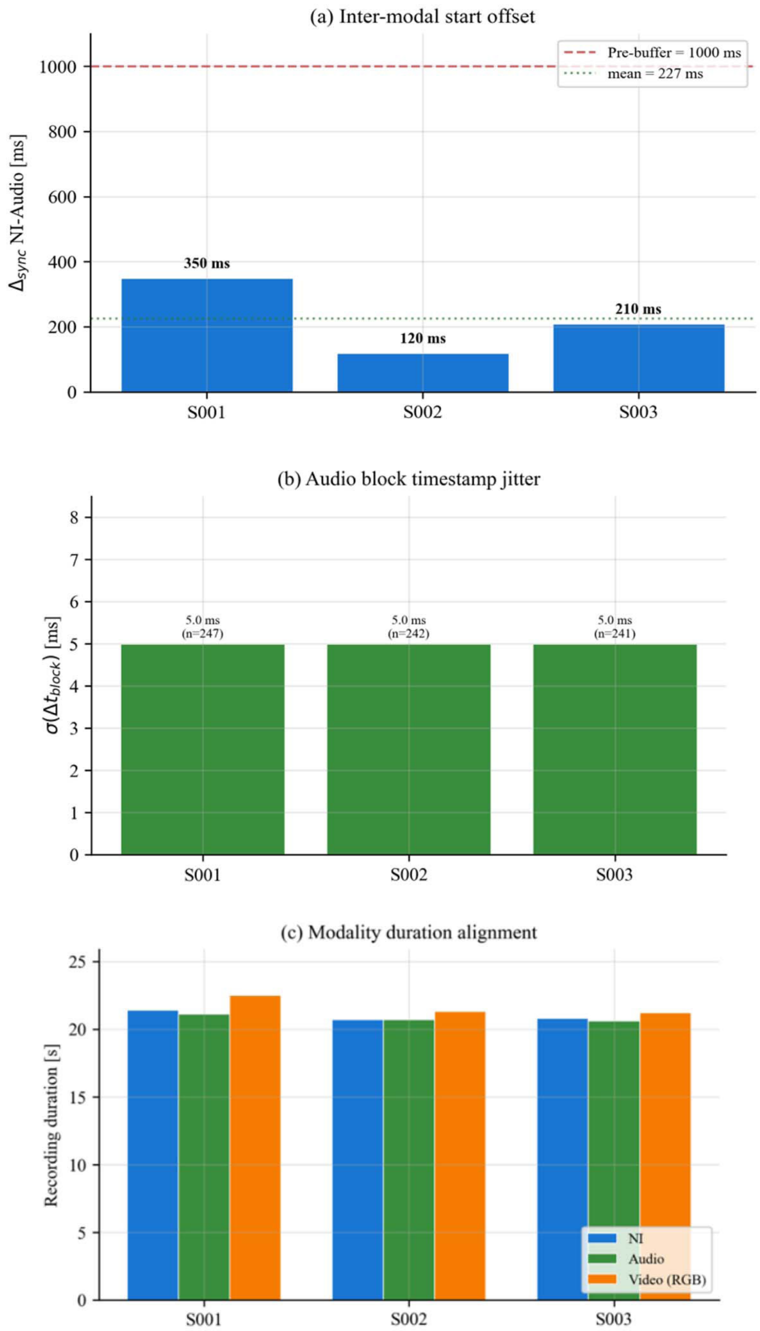

Figure 1. Acquisition timing validation on Plate P0. (a) Inter-modal start offset between NI and audio streams. (b) Audio block timestamp jitter. (c) Duration alignment of recorded modalities

P0 and included 3 weld beads. Phase II consisted of parameter exploration on Plates P1-P4 and included 32 weld beads. Travel speed, wire feed speed (WFS), and voltage trim were examined using single-parameter sweeps [15, 16]. Travel speed ranged from 300 to 1000 mm/min, WFS from 1.5 to 8.0 m/min, and voltage trim from -4 V to +3 V [17]. Phase III consisted of design-of-experiments validation with iterative refinement and included 70 weld beads. Based on the exploration results, a six-point design was selected to represent nominal, borderline, and defective process regions [18]. The final DOE points are shown in Table 4.

In total, the experimental phase yielded 105 weld beads: 3 in Phase I, 32 in Phase II, and 70 in Phase III. The final analysis focused mainly on the refined validation subset, while the exploration sweeps were used to interpret the process response to wire feed speed and voltage trim. All weld beads were labeled offline using the post-bead high-resolution images [19]. Each bead was assigned to one of three classes: acceptable, borderline, or defective [20]. Acceptable beads were continuous and geometrically regular. Borderline beads showed local imperfections, such as moderate spatter, irregularities at the start or end regions, or minor geometric deviations. Defective beads were characterized by discontinuity, severe instability, unacceptable geometry, excessive narrowing or widening, or visible surface defects [21]. The visual modality was therefore primarily used for quality labeling and interpretation, not as an input feature stream in the baseline ML models, as per standards [19–21].

Within each Phase III plate, the spatial position of each bead was randomised relative to the deposition order (position sequence [1, 4, 8, 2, 5, 7, 3, 6]) to decorrelate plate-edge and local thermal-neighbourhood effects from the order of deposition. Plate temperature was measured before every bead at three points (start, middle, end of the future bead trace) with a BOSCH GIS 1000C; in Phase III-final the inter-pass temperature ranged from 19 to 205 °C. These readings are stored in the bead-level metadata and are available as a potential covariate for downstream analysis but were not used as input features in the baseline supervised models reported here (Figures 2 and 3).

Feature extraction and dataset preparation

For each weld bead, features were extracted from the arc window, defined as the interval in which the absolute current exceeded 20 A. The first and last 10% of this interval were excluded to reduce the influence of ignition and extinction transients. Electrical descriptors included mean current, current standard deviation, maximum current, current coefficient of variation, mean voltage, voltage standard deviation, and mean electrical power. Spike rate was calculated as the number of rapid current peaks exceeding twice the mean current per second of arc time, and was

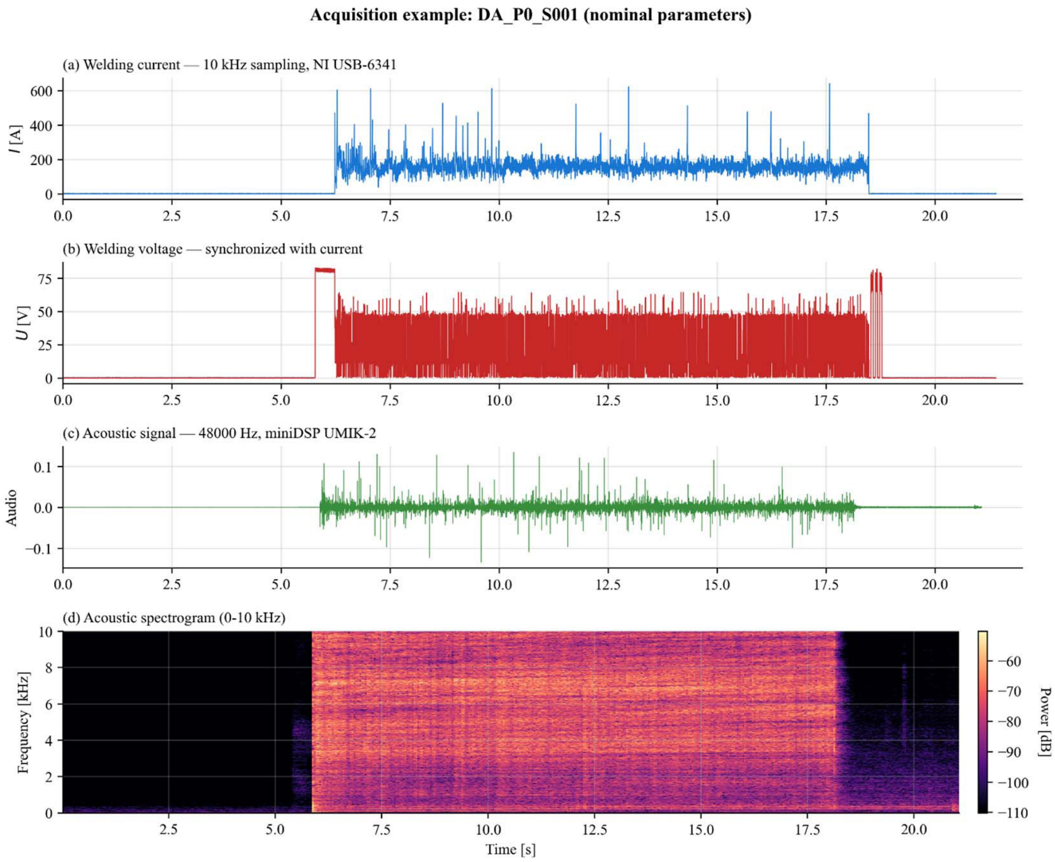

Figure 2. Example of synchronized multimodal acquisition for a nominal weld bead from Plate P0 sample S1: (a) welding current signal sampled at 10 kHz, (b) synchronized welding voltage signal, (c) acoustic waveform recorded by the measurement microphone, (d) acoustic spectrogram in the 0–10 kHz range. The figure illustrates the temporal alignment of the recorded modalities and the definition of the arc-active interval used in subsequent feature extraction (Plate P0 validation sample; airborne acoustic acquired at 48 kHz)

Figure 3. Example S355 steel plate after the MAG surfacing experiment, showing eight deposited weld beads produced under different process parameters. Post-bead images were used for offline quality labeling into acceptable, borderline, and defective classes

used as an arc-stability descriptor [22, 23]. For samples with valid acoustic recordings, three acoustic descriptors were extracted: arc-window RMS amplitude, signal-to-noise ratio, and RMS-amplitude variability over 100 ms windows. The resulting dataset contained process parameters, quality labels, current-voltage features, acoustic features where available, and metadata identifying the experimental phase and plate. Analyses involving acoustic or fused feature sets used only samples with the required acoustic features.

Machine-learning baseline

The extracted descriptors were used in a baseline machine learning analysis [24]. The aim was not to develop a production-ready classifier, but to verify whether the pilot features contain useful information for supervised classification of weld bead quality [25]. The input space for the baseline classifiers was restricted to sensor-derived descriptors only. Process parameters (WFS, voltage trim, travel speed) and experiment-management metadata (phase, plate identifier, bead index, DOE point) were excluded from the predictor set and used only for sub-setting and interpretation. This was a deliberate decision to prevent the experimental design from being encoded as a feature, given that several DOE points were chosen to target a specific intended quality class and could otherwise act as a near-perfect label proxy. Three input configurations were compared: current-voltage features only, acoustic features only, and fused current-voltage plus acoustic features [26]. Five standard classifiers were evaluated: Logistic regression, k-nearest neighbors, random forest, gradient boosting, and support vector machine with a radial basis function kernel [27]. Class weighting (class_weight = “balanced”) was applied uniformly to every model whose scikit-learn implementation supports it natively (logistic regression, random forest, SVM). k-Nearest neighbours and gradient boosting do not expose this parameter in the configuration used here and were therefore run without explicit class balancing; the imbalance in the dataset is moderate (Acceptable 13 / Borderline 12 / Defective 21) and the qualitative ordering of classifiers is unaffected by this asymmetry. Each classifier was implemented in scikit-learn (1.5.x) and wrapped in a Pipeline whose first stage is StandardScaler; the scaler was therefore re-fitted inside every cross-validation fold on the training partition only, preventing leakage of test-fold statistics into preprocessing. Cross-validation used StratifiedKFold with n_splits = min(5, n_min_class), shuffle = True, random_state = 42; a single set of out-of-fold predictions was produced via cross_val_predict and macro-F1 was computed once from those predictions. Classifier hyperparameters were the scikit-learn defaults except: Logistic Regression (max_iter = 2000); k-Nearest

Table 4. Final six-point design of experiments used in the validation phase

DOE point	WFS [m/min]	Voltage trim [V]	Travel speed [mm/min]	Intended class
1	5.0	0	600	Acceptable
2	3.5	0	600	Acceptable/Borderline
3	6.5	0	600	Borderline(geometric)
4	5.0	−3	600	Defective (electrical)
5	5.0	+5	600	Borderline
6	1.5	0	600	Defective (geometric)

neighbours (n_neighbors = 5); random forest (n_estimators = 200); gradient boosting (n_estimators = 200, max_depth = 3); SVM with RBF kernel (default C, gamma = "scale"). All models used random_state = 42 where applicable. No hyperparameter tuning was performed; this baseline is intentionally untuned to match the pilot scale of the dataset. The ML analysis was treated as a baseline evaluation supporting the feature analysis, not as a final classifier deployment.

RESULTS

Experimental campaign overview

The experimental campaign yielded 105 weld beads across 14 plates. The dataset was distributed across three experimental phases: stand validation, parameter exploration, and design-of-experiments validation. The final validation subset, referred to as Phase III-final, contained 46 weld beads: 13 acceptable, 12 borderline, and 21 defective (Figure 4). This subset was used for the main feature-space and machine-learning analyses. Table 5 summarizes the class distribution and data availability across the experimental phases. The earlier Phase III-init subset was used mainly for design refinement, whereas the refined Phase III-final subset was used for the final stability-zone and classification analysis.

Stability zones and parameter response

The distribution of weld bead quality labels in the wire-feed-speed and voltage-trim space, the same as in [28], is shown in Figure 5. The nominal region around WFS = 5.0 m/min and trim = 0 V produced predominantly acceptable beads [29]. Borderline labels appeared mainly near the edges of the operating window, particularly at reduced or increased WFS and at positive voltage trim [30]. Defective beads were concentrated in two regions: negative voltage trim and extremely low WFS [31]. This indicates that the final quality labels are not randomly distributed across the parameter space but instead correspond to repeatable process regions.

The parameter sweeps from Phase II further explain these regions. As shown in Figure 6a, the response to wire feed speed was not monotonic [32]. Very low wire feed speed produced high current variability and a high number of rapid current peaks per second, indicating unstable deposition. At high wire feed speed, the current signal became more stable, but the bead geometry was no longer acceptable [33]. This shows that electrical stability alone is not sufficient to guarantee acceptable bead quality [34]. The voltage-trim response was strongly asymmetric, as shown in Figure 6b. Negative voltage trim caused a rapid increase in current variability and spike rate.

In contrast, positive voltage trim reduced current variability [35]. The current coefficient of variation, defined as the ratio of current standard deviation to mean current, increased from approximately 0.15 at trim = +3 V to approximately 0.97 at trim = −3 V. The ratio between these values was approximately 6.6, which quantifies the asymmetry of the process response. This result indicates that shortening the arc by negative trim destabilizes the process much more strongly than increasing the trim by a comparable magnitude [36].

Class separability based on current–voltage features

The class separability analysis is summarized in Table 6 and Figure 7. The strongest separation was observed between acceptable and defective beads [37]. The largest effect sizes were obtained for spike rate, current coefficient of variation, mean current, and mean electrical power [38]. For acceptable versus defective beads, Cohen's absolute effect size reached 1.42 for spike rate, 1.34 for current coefficient of variation, 1.21 for

Table 5. Per-phase breakdown of the experimental campaign with class distribution and audio data availability

Phase	Description	Plates	n	Acceptable	Borderline	Defective	Audio valid
I	Stand validation	P0	3	3	0	0	3
II	Parameter exploration	P1–P4	32	12	11	9	32
III-init	Initial DOE	P5–P7	24	11	8	5	0
III-final	Final DOE	P8–P13	46	13	12	21	46
Total		P0–P13	105	39	31	35	81

Figure 4. Overview of the experimental campaign. (a) Per-plate distribution of weld bead quality labels across the experimental phases. (b) Aggregate sample counts per phase

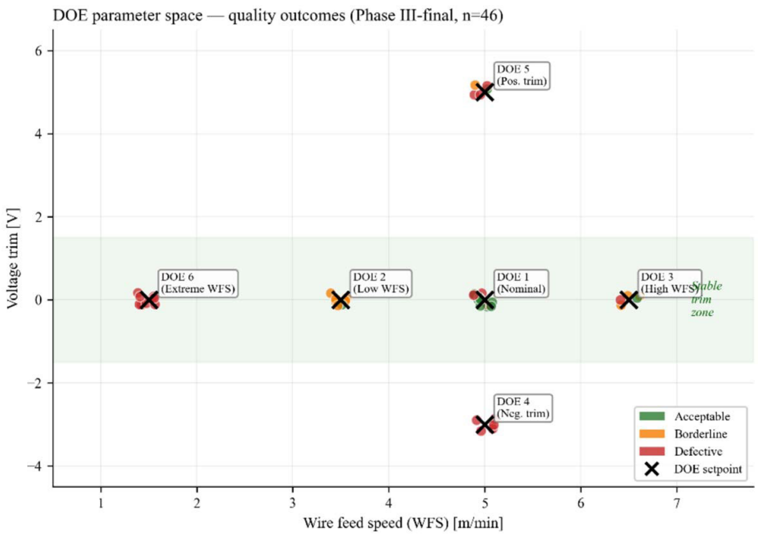

Figure 5. Stability zones in the wire-feed-speed and voltage-trim parameter space for Phase III-final. Each marker represents one weld bead and is coloured by quality class

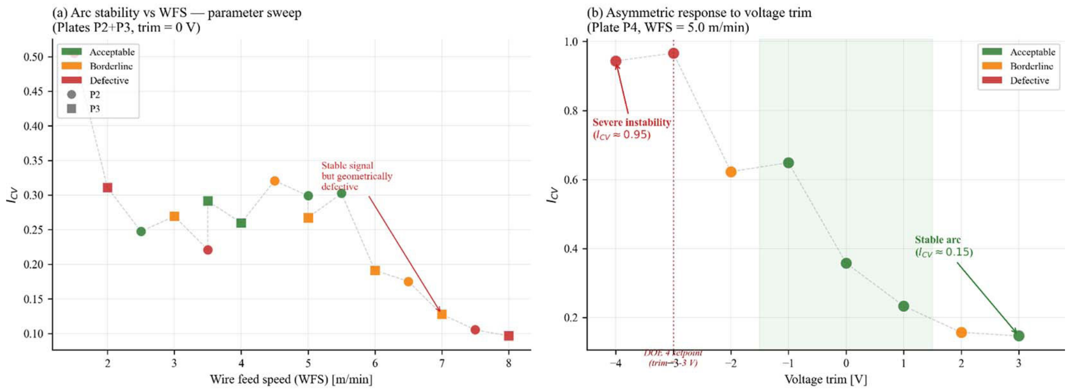

Figure 6. Parameter response curves from the first phase. (a) Current coefficient of variation as a function of wire feed speed. (b) Asymmetric response of current variability and spike rate to voltage trim

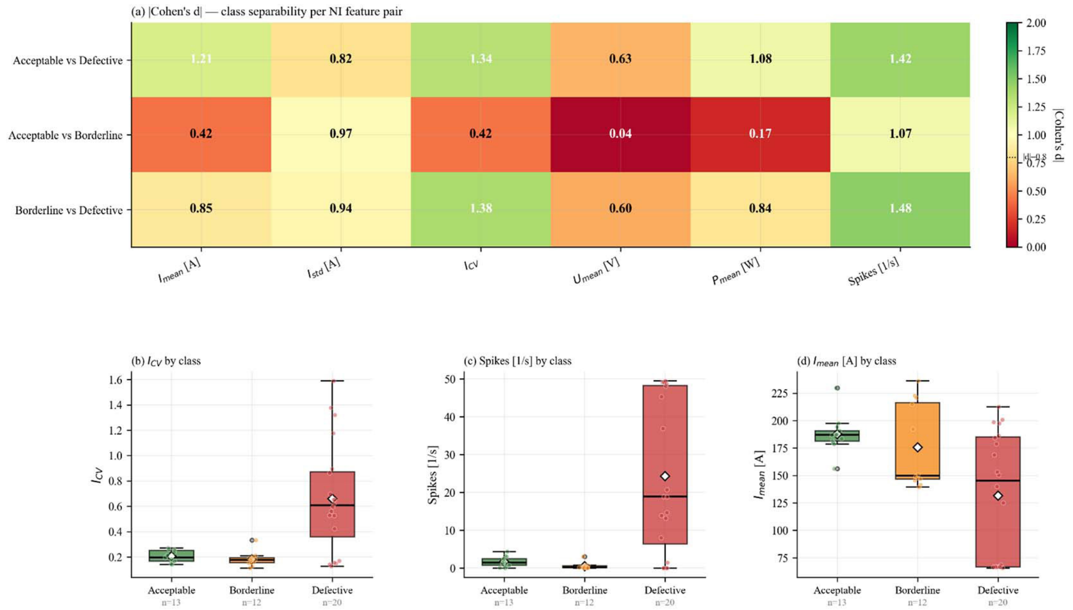

Figure 7. Class separability analysis. (a) Heatmap of absolute Cohen's d effect sizes for pairwise class separability. (b–d) Box plots of selected current-based descriptors per class

mean current, and 1.08 for mean electrical power. In contrast, acceptable and borderline beads were much less clearly separated by current–voltage descriptors. The absolute effect size for the current coefficient of variation was 0.42, and the corresponding effect sizes for the mean current and mean electrical power were also small. This suggests that the borderline class contains cases whose quality degradation is not dominated by electrical instability, but rather by geometric or local visual imperfections [39]. The only notable exception was spike rate, which may reflect local start- or end-region instabilities.

Mechanism-dependent defective signatures

The defective class did not correspond to a single signal pattern. Defective beads produced under negative-voltage trim showed high current variability and frequent spike events, consistent with unstable short-circuit transfer. In contrast, defective beads produced at extremely low WFS showed a different pattern: reduced mean current, unstable deposition, and high spike rate associated with insufficient material deposition. These two cases share the same final quality label but differ in the degradation mechanisms. This is also visible in the low-dimensional projections of the current–voltage feature space shown in Figure 8. Defective beads tend to form separate groups associated with negative trim and low-WFS conditions, whereas acceptable and borderline beads partially overlap. This supports the interpretation that the defective class should not be treated as a single, homogeneous distribution in supervised learning pipelines.

Acoustic signatures

Representative acoustic RMS envelopes are shown in Figure 9. The acoustic signal provided

Table 6. Feature statistics per class for Phase III-final, with Cohen's d effect size between acceptable and defective classes

Feature	Unit	Acceptable (n=13)	Borderline (n=12)	Defective (n=21)	|d| Acc–Def
I_mean	A	187.8 ± 15.5	175.7 ± 36.6	131.5 ± 56.6	1.21
I_std	A	38.0 ± 7.8	30.7 ± 6.4	79.9 ± 63.8	0.82
I_CV	—	0.20 ± 0.05	0.18 ± 0.05	0.66 ± 0.42	1.34
U_mean	V	20.30 ± 1.57	20.38 ± 2.38	18.56 ± 3.21	0.63
P_mean	W	3889 ± 577	3726 ± 1197	2504 ± 1524	1.08
Spike rate	events/s	1.7 ± 1.3	0.5 ± 0.8	24.3 ± 19.7	1.42

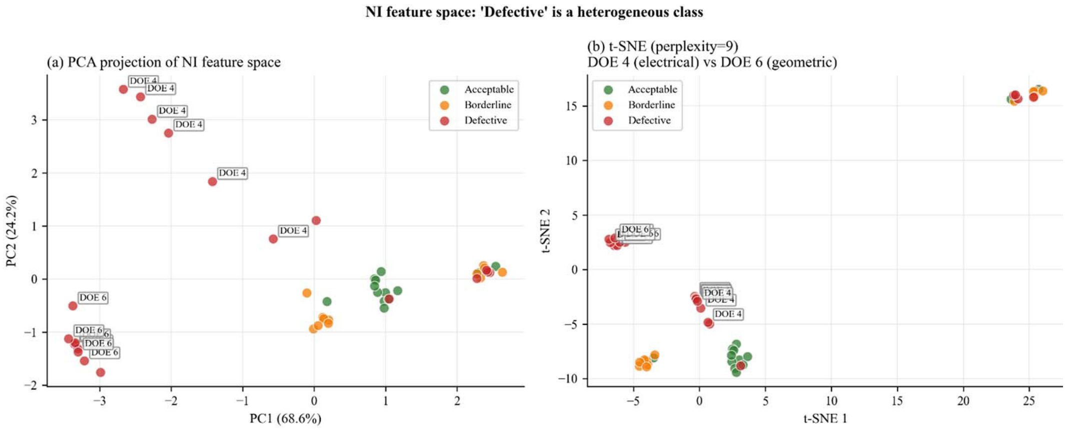

Figure 8. Two-dimensional projections of the current–voltage feature space using (a) Principal component analysis and (b) t-distributed stochastic neighbor embedding. Defective beads form separate groups associated with negative voltage trim and low wire feed speed

an additional view of process dynamics [40]. Acceptable beads showed relatively stable acoustic envelopes, whereas defective beads associated with electrical instability showed increased RMS variability [41]. Low-WFS defective beads showed lower acoustic amplitude, consistent with reduced arc intensity and insufficient material deposition. The acoustic trends were qualitatively consistent with the current-based observations [42]. However, in this dataset, the acoustic modality should be interpreted as complementary rather than primary, because the strongest quantitative class separation was obtained from current–voltage descriptors.

Machine-learning baseline

The baseline ML results are summarized in Table 7 and Figure 10. Among the current–voltage-only models, Random Forest achieved the highest performance, with an accuracy of 0.756 and a macro-F1 of 0.725. The best acoustic-only model was the SVM with a radial basis function kernel, achieving an accuracy of 0.641 and a macro-F1 of 0.624.

The highest macro-F1 score overall was obtained by the fused current–voltage and acoustic feature set with Random Forest, reaching 0.735. The improvement of the fused configuration over the current–voltage-only Random Forest baseline was small, with Δmacro-F1 = +0.010. Therefore, the ML results should not be interpreted as strong evidence that fusion is superior in this dataset. Instead, they show that the extracted features contain useful information for supervised classification and that acoustic descriptors may provide limited complementary information. The reported accuracy and macro-F1 values should be

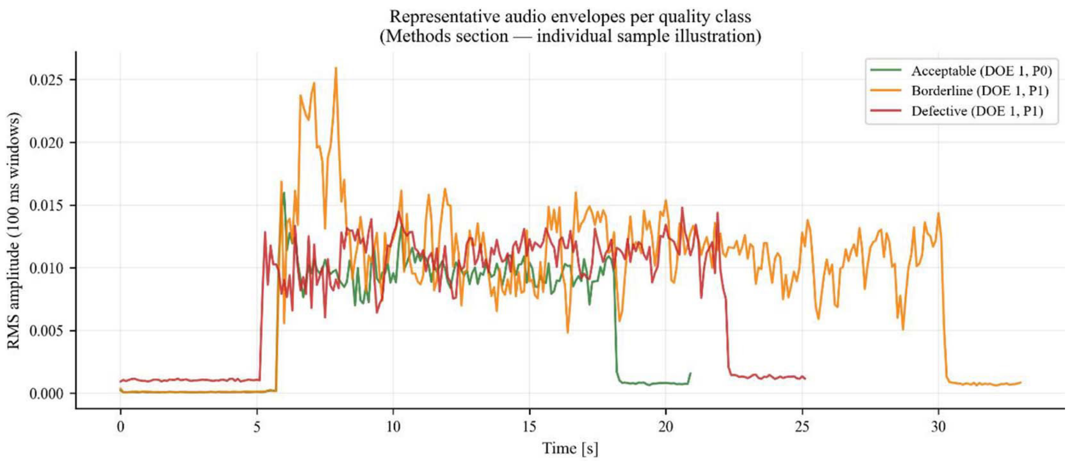

Figure 9. Representative acoustic RMS envelopes calculated using 100 ms windows. The acoustic signatures illustrate differences between stable deposition, electrical instability, and insufficient material deposition

Table 7. Cross-validation performance summary for all classifiers and feature-configuration combinations

Classifier	NI only – Acc	NI only – F1	Audio only – Acc	Audio only – F1	Combined – Acc	Combined – F1
Logistic Regression	0.689	0.624	0.590	0.543	0.692	0.661
KNN (k=5)	0.733	0.698	0.615	0.600	0.615	0.582
Random Forest	0.756	0.725	0.615	0.564	0.761	0.735
Gradient Boosting	0.711	0.686	0.590	0.538	0.692	0.688
SVM (RBF)	0.667	0.626	0.641	0.624	0.692	0.658

interpreted as within-campaign estimates. They were obtained with stratified k-fold cross-validation on a single material heat, a single power source, a single shielding gas mixture and a single operator, and they do not constitute evidence of generalisation to new plates, new material batches, or new process configurations. Leave-one-plate-out (LOPO) validation, which would offer a more realistic estimate of cross-plate generalisation, is reserved for the full-scale study.

DISCUSSION

Interpretation of process signatures

The results show that the quality labels obtained in the experimental campaign correspond to repeatable regions in the process-parameter space. The strongest separation was observed between acceptable and defective beads, mainly for current-based descriptors related to arc instability. Current coefficient of variation, spike rate, mean current, and mean electrical power showed large effect sizes, confirming that electrical signals are highly informative for detecting severe process degradation. At the same time, acceptable and borderline beads were weakly separated by current–voltage descriptors. This suggests that many borderline cases were due to local geometric imperfections, such as start- or end-region irregularities, moderate spatter, or deviations in bead shape, rather than to strong arc instability. Therefore, current-based features are effective for detecting severe defective conditions, but less reliable for distinguishing acceptable and borderline beads. The voltage-trim sweep also revealed an asymmetric process response. Negative voltage trim rapidly increased current variability and spike rate, whereas positive trim reduced current variability. This can be explained by the arc shortening under negative trim, which increases the likelihood of unstable short-circuiting events. Positive trim lengthens the arc and may reduce short-circuit-dominated instability, although it can still affect bead geometry. As a result, current-based stability descriptors are especially sensitive to negative deviations from the nominal operating region.

Implications for machine-learning-based quality assessment

The main implication for machine-learning-based weld quality assessment is that the defective class should not be treated as a single uniform distribution. Defective beads produced by negative-voltage trim and by extremely low wire-feed speed received the same final label, but their signal signatures differed. Negative trim was associated with high current variability and frequent spike events, whereas extreme low wire feed speed was associated with reduced mean current and insufficient deposition. This has direct consequences for supervised learning. A classifier trained only to separate acceptable from defective beads may perform well on a dataset, but still fail

Table 8. Per-class classification report for the best model - Random Forest

Class	Precision	Recall	F1-score	Support
Acceptable	0.846	0.846	0.846	13
Borderline	0.545	0.500	0.522	12
Defective	0.818	0.857	0.837	21
Macro avg	0.737	0.734	0.735	46
Weighted avg	0.755	0.761	0.757	46

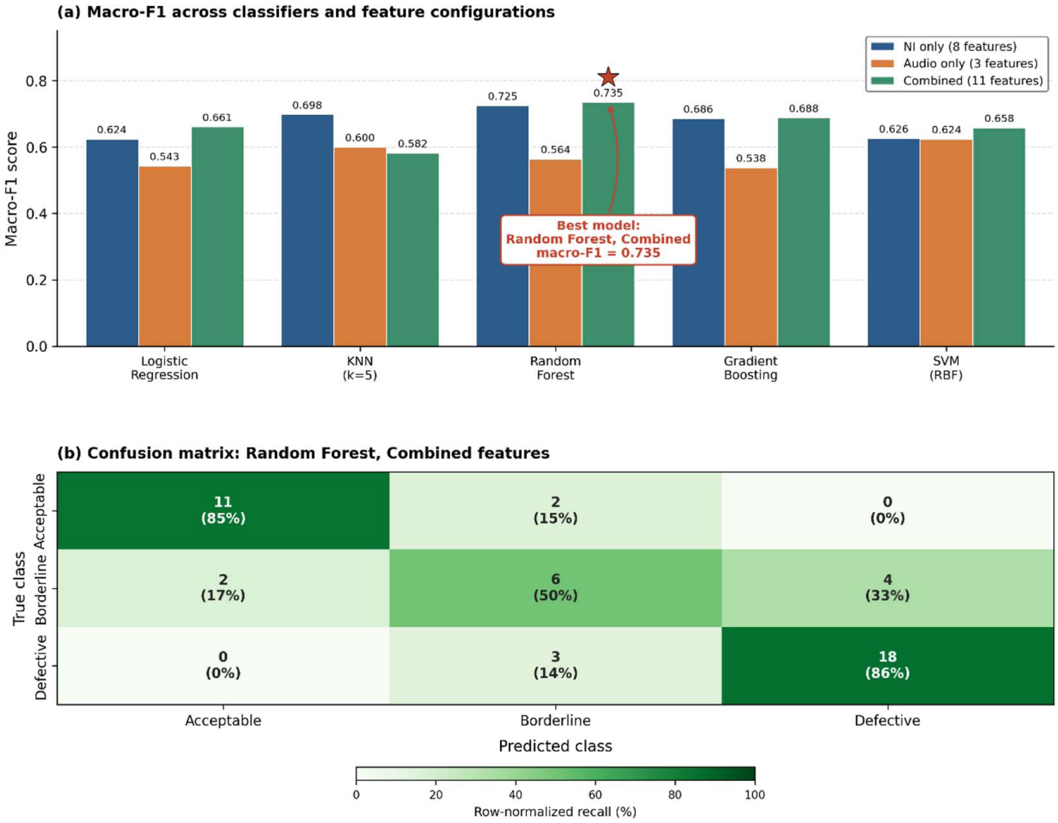

Figure 10. Machine-learning baseline performance. (a) Macro-F1 scores for all five classifiers across the three feature configurations (NI only, audio only, fused). (b) Confusion matrix for the best fused-configuration model (Random Forest, macro-F1 = 0.735), with absolute counts and row-normalized percentages indicating class-conditional recall

to generalize if future defective samples originate from mechanisms underrepresented in the training set. In such cases, the model may learn mechanism-specific correlations rather than general indicators of poor weld bead quality. For this reason, the results support a mechanism-aware classification approach. Future datasets should distinguish between degradation mechanisms such as electrical instability, insufficient material deposition, excessive deposition, and geometric irregularity [43]. A practical solution may be a hierarchical model that first separates acceptable from unacceptable beads and then identifies the likely defect mechanism [44].

The weak classifiability of the borderline class is therefore not primarily a property of the classifier but a property of the label itself. The borderline category aggregates local, geometric, and transient imperfections (start- and end-region irregularities, moderate spatter, narrow geometric deviations) that need not produce strong anomalies in the current-voltage signals. This is a sensor-coverage gap, not a model deficiency, and it provides a direct argument for adding quantitative image-based descriptors (bead width, spatter coverage, geometric roughness) in the full-scale campaign. The per-class classification report (Table 8) confirms this picture: Acceptable and Defective beads are recognised with high F1 (0.846 and 0.837), whereas the Borderline class is the principal source of error (F1 = 0.522, recall = 0.500). Six of twelve borderline beads were misclassified, mostly into the adjacent Acceptable and Defective classes, while only two of thirteen acceptable and three of twenty-one defective beads were misclassified. The bottleneck is therefore the separation of the borderline class from its neighbours, not the detection of defective beads.

Role of multimodal sensing

The baseline machine-learning results showed that current–voltage features alone already provide useful classification performance. The fused current–voltage and acoustic feature set achieved the highest macro-F1 score, but the improvement over the current–voltage-only baseline was small. Therefore, the results should not be interpreted as strong evidence that acoustic fusion is universally superior. Rather, they indicate that acoustic descriptors may contain complementary information for selected transient or mechanism-specific phenomena [45]. In this study, the strongest quantitative class separation was obtained from current–voltage descriptors. Acoustic monitoring should therefore be treated as a complementary modality. The visual modality was primarily used for offline labeling and interpretation, not as an input feature stream in the baseline classifiers [46]. This is important because many borderline cases appeared to be related to bead geometry rather than electrical instability. Incorporating quantitative image-based descriptors is therefore a natural next step for improving the classification of borderline and geometric defects [47].

The principal contribution of the study is therefore not the macro-F1 number or the relative ordering of the five classifiers, but the structural observation that defective beads and borderline beads each aggregate at least two distinct degradation mechanisms with separable sensor signatures. This argues for redesigning the labelling scheme along mechanism lines, rather than for replacing one classifier family with another, and motivates the inclusion of quantitative image-based descriptors in the full-scale campaign.

Limitations and future work

This work should be interpreted as a preliminary pilot validation rather than a production-ready evaluation, and several limitations follow directly from this scope. The dataset of 105 weld beads is sufficient to characterize signal–quality relationships and to identify candidate features for a full-scale study. It is, however, too small for definitive ML benchmarking: confidence intervals on classifier performance metrics span approximately ±15 percentage points at this sample size, and the +0.010 macro-F1 advantage of the fused configuration is well within the statistical noise of the evaluation. The classifier comparison should therefore be read as evidence that hand-crafted features carry useful information for supervised learning, not as a definitive ranking of models or sensing modalities. The study is restricted to a single MAG surfacing setup, a single power source, a single shielding gas mixture, and S355 steel. Generalization to other materials, power sources, wire types, or process modes requires explicit cross-domain validation [48]. A single rater from post-bead images assigned quality labels; inter-rater agreement was not quantified, and visual information was used for labeling rather than as a quantitative feature stream. The acoustic modality was unavailable for plates P5–P7 due to a transient hardware issue, and acoustic results should be considered preliminary [49]. The cross-validation scheme used stratified k-fold rather than leave-one-plate-out (LOPO); the latter is the more realistic estimator of cross-plate generalization and is reserved for the full-scale study. The findings are therefore best read as a pilot characterization that identifies signal signatures, quantifies the heterogeneity of the defective class, maps the asymmetric trim response, and motivates the experimental design of a full-scale campaign with stricter cross-validation, multi-rater labeling, inclusion of quantitative visual descriptors, and explicit mechanism-aware classifier evaluation [50].

The six DOE points were deposited in a fixed temporal order within each plate, so the inter-pass temperature is partially confounded with the DOE point in this pilot; this confound was not fully decorrelated. The recorded three-point temperature log nevertheless allows the inter-pass temperature to be treated as a covariate in subsequent analysis. The full-scale campaign will either enforce a controlled inter-pass temperature window (forced cooling between beads) or randomise the temporal order of DOE points within each plate.

Surface preparation was uniform across the campaign (grinding + Clino-3 degreasing on every plate, from a single material heat with an EN 10204 type 3.1 certificate), but residual local contamination cannot be fully excluded; in the present dataset one bead deposited at nominal parameters was classified as borderline specifically because of a localised surface defect. This is consistent with the heterogeneous nature of the borderline class (Section 4.2) and motivates stricter surface-condition control in the full-scale campaign.

CONCLUSIONS

The principal conclusion of the campaign is that the quality labels assigned to MAG-surfacing beads are not internally homogeneous in the sensor feature space: defective beads form at least two mechanism-specific clusters (electrical instability vs insufficient deposition) and the borderline class aggregates further local, geometric and transient imperfections that are not well captured by current-voltage signals. Supervised pipelines for weld-bead quality assessment should therefore use mechanism-aware labels and sensor coverage rather than aggregated quality categories.

The experimental campaign yielded 105 weld beads and provided the following conclusions. Current-based descriptors were the most informative features in the dataset. Current coefficient of variation and spike rate strongly separated acceptable and defective beads, with Cohen's effect sizes of 1.34 and 1.42, respectively. However, acceptable and borderline beads were weakly separated by electrical descriptors, suggesting that borderline cases are mainly related to local geometric imperfections. Arc stability showed a strongly asymmetric response to voltage trim. Negative trim rapidly increased current variability and spike rate, while positive trim reduced current variability. The current coefficient of variation at trim = −3 V was approximately 6.6 times higher than at trim = +3 V. Defective beads were not homogeneous in the feature space. Negative-trim defects and low-wire-feed defects exhibited distinct signal signatures, suggesting distinct degradation mechanisms underlying the same final quality label. This result is important for supervised ML, where treating all defective beads as one uniform class may reduce generalization and diagnostic value. The baseline ML evaluation confirmed that the extracted process features are useful for classifying weld bead quality. The highest macro-F1 score was 0.735 for the fused current–voltage and acoustic feature set, although the improvement over the current–voltage-only baseline was marginal. The results therefore support mechanism-aware feature design and hierarchical classification rather than the direct deployment of a production-ready classifier.

Acknowledgments

The article was financed from the Research Subsidy of the Faculty of Mechanical Engineering and Robotics, AGH University of Krakow. Research project partly supported by program "Excellence initiative – research university" for the AGH University of Krakow.

The research presented in this article was carried out as part of the doctoral dissertation of Przemysław Frankiewicz at the Faculty of Mechanical Engineering and Robotics, AGH University of Kraków.

REFERENCES

1. Czapiewski W, Pałubicki S, Plichta J, Nadolny K. State of knowledge in the field of regenerative hardfacing methods in the context of the circular economy. Applied Sciences [Internet]. 2026; 16(2): 792. https://doi.org/10.3390/app16020792
2. Madhvacharyula AS, Pavan AVS, Gorthi S, Chitral S, Venkaiah N, Kiran DV. In situ detection of welding defects: a review. Welding in the World. 2022; 66(4): 611–28. https://doi.org/10.1007/s40194-021-01229-6
3. Zhu Q, Huang Z, Li H. Research progress of acoustic monitoring technology in welding and additive manufacturing processes. Micromachines [Internet]. 2026; 17(2): 246. https://doi.org/10.3390/mi17020246.
4. Arjomandi M, Motley J, Ngo Q, Anees Y, Afzal MA, Mukherjee T. A Review on in-situ monitoring in wire arc additive manufacturing: Technologies, applications, challenges, and needs. Machines [Internet]. 2026; 14(1): 19. https://doi.org/10.3390/machines14010019
5. Tanwar RS, Jhavar S, Das D, Chethan Kumar N. Deep learning and machine learning approaches for acoustic-based defect monitoring in wire arc direct energy deposition. Journal of Intelligent Manufacturing. 2026. https://doi.org/10.1007/s10845-026-02855-x
6. Wang Z, Li L, Chen H, Wu X, Dong Y, Tian J, et al. Penetration recognition based on machine learning in arc welding: A review. The International Journal of Advanced Manufacturing Technology. 2023; 125(9): 3899–923. https://doi.org/10.1007/s00170-023-11035-7
7. Griffin JM, Jones S, Perumal B, Perrin C. Investigating the detection capability of acoustic emission monitoring to identify imperfections produced by the Metal Active Gas (MAG) welding process. Acoustics [Internet]. 2023; 5(3): 714–45. https://doi.org/10.3390/acoustics5030043
8. Kumar V, Parida MK, Albert SK. The state-of-the-art methodologies for quality analysis of arc welding process using weld data acquisition and analysis techniques. International Journal of System Assurance Engineering and Management. 2022; 13(1): 34–56. https://doi.org/10.1007/s13198-021-01282-w
9. Zhirnov I, Krakhmalev P. Enhanced quality prediction in laser powder bed fusion through data fusion-enabled regression analysis of spatter-related defects. Journal of Manufacturing Processes. 2025; 152: 293–307. https://doi.org/10.1016/j.jmapro.2025.08.013
10. Tajalli Nobari S, Hanning F, Mi Y, Volpp J. Mechanisms, sensors, and signals for defect formation and in situ monitoring in metal additive

manufacturing. Eng [Internet]. 2026; 7(3): 129. https://doi.org/10.3390/eng7030129

11. Gao H, Li H, Shao D, Fang N, Miao Y, Pan Z, et al. Towards quality controllable strategies in wire-arc directed energy deposition. International Journal of Extreme Manufacturing. 2025; 7(4): 042004. https://doi.org/10.1088/2631-7990/adb9a9
12. Norrish J. Recent gas metal arc welding (GMAW) process developments: the implications related to international fabrication standards. Welding in the World. 2017; 61(4): 755–67. https://doi.org/10.1007/s40194-017-0463-8
13. Pal K, Bhattacharya S, Pal SK. Multisensor-based monitoring of weld deposition and plate distortion for various torch angles in pulsed MIG welding. The International Journal of Advanced Manufacturing Technology. 2010; 50(5): 543–56. https://doi.org/10.1007/s00170-010-2523-8
14. Devakumaran K, Rajasekaran N, Ghosh PK. Process characteristics of inverter type GMAW power source under static and dynamic operating conditions. Materials and Manufacturing Processes. 2012; 27(12): 1450–6. https://doi.org/10.1080/10426914.2012.663149
15. Kumar V, Ranjan Sahu D, Mandal A. Parametric study and optimization of GMAW based AM process for Multi-layer bead deposition. Materials Today: Proceedings. 2022; 62: 255–61. https://doi.org/10.1016/j.matpr.2022.03.223
16. Kamble AG, Rao RV. Investigation on effects of parameters of GMAW process on bead geometry, hardness and microstructure of AISI 410 steel weldments. Advances in Materials and Processing Technologies. 2022; 8(3): 2450–64. https://doi.org/10.1080/2374068X.2021.1912537
17. Rao PS, Gupta OP, Murty SSN, Rao ABK. Effect of process parameters and mathematical model for the prediction of bead geometry in pulsed GMA welding. The International Journal of Advanced Manufacturing Technology. 2009; 45(5): 496–505. https://doi.org/10.1007/s00170-009-1991-1
18. Sreeraj P, Kannan T, Maji S. Optimization of GMAW process parameters using particle swarm optimization. International Scholarly Research Notices. 2013; 2013(1): 460651. https://doi.org/10.1155/2013/460651
19. PN-EN ISO 17637:2017-02 Badania nieniszczące spoin - Badania wizualne złączy spawanych. Warszawa: PKN; 2017.
20. PN-EN ISO 6520-1:2009 Spawanie i procesy pokrewne - Klasyfikacja geometrycznych niezgodności spawalniczych w metalach - Część 1: Spawanie. Warszawa: PKN; 2009.
21. PN-EN ISO 5817:2023-07 Spawanie - Złącza spawane ze stali, niklu, tytanu i ich stopów (z wyjątkiem spawanych wiązką) - Poziomy jakości według niezgodności spawalniczych. Warszawa: PKN; 2023.
22. Santhakumari A, Senthilkumar T, Ramasamy N, Mahadevan G. Influence of controlled and conventional short circuit waveforms on mechanical and microstructural effects in the gas metal arc welding processes. Case Studies in Thermal Engineering. 2024; 55: 104056. https://doi.org/10.1016/j.csite.2024.104056
23. Galeazzi D, Silva RHGe, Pigozzo IO, da Rosa AF, Pereira AS, Marques C. Analysis of current pulse during short-circuit phase in CMT version of GMAW process under a view of additive manufacturing. Welding in the World. 2022; 66(7): 1369–80. https://doi.org/10.1007/s40194-022-01298-1
24. Mattera G, Chozaki SP, Norrish J. Advances in machine learning for parameters optimisation and in-situ monitoring of wire arc additive manufacturing. Welding in the World. 2026; 70(4): 1173–202. https://doi.org/10.1007/s40194-025-02200-5
25. Eka S, Paton NW, Sherry AH, Pickering EJ, Martin M. Review of state-of-the-art machine learning research for structural integrity assessment of welded joints. Volume 2: Computer Technology & Bolted Joints 2025. https://doi.org/10.1115/PVP2025-155746
26. Thekkuden DT, Mourad A-HI. Investigation of feed-forward back propagation ANN using voltage signals for the early prediction of the welding defect. SN Applied Sciences. 2019; 1(12): 1615. https://doi.org/10.1007/s42452-019-1660-4
27. Huang Y, Yang D, Wang K, Wang L, Zhou Q. Stability analysis of GMAW based on multi-scale entropy and genetic optimized support vector machine. Measurement. 2020; 151: 107282. https://doi.org/10.1016/j.measurement.2019.107282
28. Cruz JG, Torres EM, Absi Alfaro SC. A methodology for modeling and control of weld bead width in the GMAW process. Journal of the Brazilian Society of Mechanical Sciences and Engineering. 2015; 37(5): 1529–41. https://doi.org/10.1007/s40430-014-0299-8
29. Ramos-Jaime D, Juárez IL, Perez P. Effect of process parameters on robotic GMAW bead area estimation. Procedia Technology. 2013; 7: 398–405. https://doi.org/10.1016/j.protcy.2013.04.050
30. Kannan T, Yoganandh J. Effect of process parameters on clad bead geometry and its shape relationships of stainless steel claddings deposited by GMAW. The International Journal of Advanced Manufacturing Technology. 2010; 47(9): 1083–95. https://doi.org/10.1007/s00170-009-2226-1
31. Nogueira RMU, Bohórquez CEN, Zanella IG. Comparison of arc stability and variability of geometry of welds obtained by MIG/MAG and FCAW processes. Welding International. 2017; 31(3): 161–72. https://doi.org/10.1080/09507116.2016.1218611

32. Mvola B, Kah P, Layus P. Review of current waveform control effects on weld geometry in gas metal arc welding process. The International Journal of Advanced Manufacturing Technology. 2018; 96(9): 4243–65. https://doi.org/10.1007/s00170-018-1879-z
33. Chaudhari R, Parmar H, Vora J, Patel VK. parametric study and investigations of bead geometries of GMAW-based wire–arc additive manufacturing of 316L stainless steels. Metals [Internet]. 2022; 12(7): 1232. https://doi.org/10.3390/met12071232
34. Hackenhaar W, Gonzalez AR, Machado IG, Mazzaferro JAE. Welding parameters effect in GMAW fusion efficiency evaluation. The International Journal of Advanced Manufacturing Technology. 2018; 94(1): 497–507. https://doi.org/10.1007/s00170-017-0888-7
35. Ferreira GRB, Filho RMM, Scotti A, Lagares ML. Exploring a locus of maximum metal transfer stability of the short-circuiting GMAW process based on the reignition voltage peaks. Journal of the Brazilian Society of Mechanical Sciences and Engineering. 2021; 43(11): 503. https://doi.org/10.1007/s40430-021-03223-x
36. Ohara M, Mizuguchi T, Miyata K, Tsuyama T, Fujiwara K. A new approach to controlling metal transfer by dynamic modification in gas composition of arc atmosphere: studies on pulsed gas MAG welding*. Welding International. 2020; 34(10–12): 430–54. https://doi.org/10.1080/09507116.2021.1936927
37. Sumesh A, Rameshkumar K, Raja A, Mohandas K, Santhakumari A, Shyambabu R. Establishing correlation between current and voltage signatures of the arc and weld defects in GMAW process. Arabian Journal for Science and Engineering. 2017; 42(11): 4649–65. https://doi.org/10.1007/s13369-017-2609-9
38. Kah P, Edigbe GO, Ndiwe B, Kubicek R. Assessment of arc stability features for selected gas metal arc welding conditions. SN Applied Sciences. 2022; 4(10): 268. https://doi.org/10.1007/s42452-022-05150-5
39. Adak DK, Mukherjee M, Pal TK. Development of a direct correlation of bead geometry, grain size and HAZ width with the GMAW process parameters on bead-on-plate welds of mild steel. Transactions of the Indian Institute of Metals. 2015; 68(5): 839–49. https://doi.org/10.1007/s12666-015-0518-8
40. Norouzian M, Orosnjak M, Kedziora SA, Useldinger R, Aslan F. acoustic emission signatures as predictors of laser weld integrity: A machine learning framework for real-time laser welding quality assessment. SSRN, 2025. https://dx.doi.org/10.2139/ssrn.5556630
41. Hauser T, Reisch RT, Kamps T, Kaplan AFH, Volpp J. Acoustic emissions in directed energy deposition processes. The International Journal of Advanced Manufacturing Technology. 2022; 119(5): 3517–32. https://doi.org/10.1007/s00170-021-08598-8
42. Hamidi Nasab M, Masinelli G, de Formanoir C, Schlenger L, Van Petegem S, Esmaeilzadeh R, et al. Harmonizing sound and light: X-ray imaging unveils acoustic signatures of stochastic inter-regime instabilities during laser melting. Nature Communications. 2023; 14(1): 8008. https://doi.org/10.1038/s41467-023-43371-3
43. Mattera G, Nele L, Pan Z. Investigation of contact acoustic emission-based in-situ anomaly detection in wire arc additive manufacturing using deep unsupervised learning. The International Journal of Advanced Manufacturing Technology. 2026; 143(3): 2017–40. https://doi.org/10.1007/s00170-026-17525-8
44. Xu J, Liu Q, Xu Y, Xiao R, Hou Z, Chen S. Review on the application of the attention mechanism in sensing information processing for dynamic welding processes. Journal of Manufacturing and Materials Processing [Internet]. 2024; 8(1): 22. https://doi.org/10.3390/jmmp8010022
45. Pandiyan V, Baganis A, Axel Richter R, Wróbel R, Leinenbach C. 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 and Physical Prototyping. 2024; 19(1): e2356080. https://doi.org/10.1080/17452759.2024.2356080
46. Jeong YC, Lee SJ, Park TW, Kim HK, Cho YT. Real-time monitoring of immobilized oxide defects on bead surface in welding of high-strength galvanized steel. The International Journal of Advanced Manufacturing Technology. 2020; 107(5): 2409–16. https://doi.org/10.1007/s00170-020-05070-x
47. Pinto-Lopera JE, S. T. Motta JM, Absi Alfaro SC. Real-time measurement of width and height of weld beads in GMAW processes. Sensors [Internet]. 2016; 16(9): 1500. https://doi.org/10.3390/s16091500
48. Assunção PDC, Ribeiro RA, Dos Santos EBF, Braga EM, Gerlich AP. Comparing CW-GMAW in direct current electrode positive (DCEP) and direct current electrode negative (DCEN). The International Journal of Advanced Manufacturing Technology. 2019; 104(5): 2899–910. https://doi.org/10.1007/s00170-019-04175-2
49. He F, Wu Z, Zhang Z, Ding D, Tong J, Li H, et al. Investigation of the mechanisms of structure-induced acoustic modulation in WA-DED metal additive manufacturing with an adaptive modelling approach. Journal of Materials Processing Technology. 2026; 347: 119160. https://doi.org/10.1016/j.jmatprotec.2025.119160
50. Zhang J, Yin C, Farbiz F, Jafary-Zadeh M, Sing SL. Advancing machine learning applications for in-situ monitoring and control in laser-based metal additive manufacturing: a state-of-the-art review. Virtual and Physical Prototyping. 2025; 20(1): e2592732. https://doi.org/10.1080/17452759.2025.2592732