OPEN ACCESS
Journal of Computer Science Advancements
Vol. 4 No. 4. August 2026, pp. 245-267
Check for updates
DOI. 10.70177/jsca.v4i4.4458
Research Article

EXPLAINABLE AI FRAMEWORK FOR ROOT CAUSE ANALYSIS OF MARINE WELDING DEFECTS: INTEGRATING PROCESS PARAMETERS AND MICROSTRUCTURAL FAILURE MODES

Mohammad Syaifi[1]
[1] Sekolah Tinggi Teknologi Angkatan Laut, Indonesia

Corresponding Author:

Mohammad Syaifi,
Department of Mechanical Engineering, Sekolah Tinggi Teknologi Angkatan Laut.
QPJ9+3C2, Jl. Bumi Moro, Morokrembangan, Kec. Krembangan, Surabaya, Jawa Timur 60178
Email: ifikhaj71@gmail.com

Article Info
Received: February 9, 2026
Revised: May 10, 2026
Accepted: July 11, 2026
OnlineVersion: August 31, 2026

Abstract
Marine welding defects threaten structural reliability because conventional inspection and black-box prediction models can identify discontinuities without adequately explaining their physical origins. This study aimed to develop an Explainable Artificial Intelligence framework for root-cause analysis by integrating welding process parameters with microstructural failure modes. An experimental-computational design was applied to 120 AH36 marine-steel specimens produced by Gas Metal Arc Welding under controlled variations in current, voltage, travel speed, wire-feed rate, shielding-gas flow, and heat input. Weld quality was evaluated using visual inspection, ultrasonic testing, metallography, scanning electron microscopy, and Vickers microhardness, while Random Forest, XGBoost, and neural-network models were compared and interpreted using SHAP. Parameter variation enabled AI explanations to be checked against metallurgical evidence and defect morphology. XGBoost achieved the best performance, with 90.0% accuracy, a macro F1-score of 0.886, and an AUROC of 0.956. Heat input, shielding-gas flow, travel speed, and HAZ microhardness emerged as the dominant explanatory variables. Porosity was mainly associated with inadequate shielding, incomplete fusion and penetration with insufficient effective heat input, and cracking with elevated HAZ hardness. The study concludes that physically validated XAI can extend defect classification toward transparent, mechanism-informed diagnosis and provide a stronger basis for corrective decision-making in safety-critical marine welding.

Keywords: Explainable Artificial Intelligence; Marine Welding; Microstructural Failure Modes; Root Cause Analysis; Welding Defects

Journal Homepage	https://research.adra.ac.id/index.php/jcsa
How to cite:	Syaifi, M. (2026). Explainable Ai Framework for Root Cause Analysis of Marine Welding Defects: Integrating Process Parameters and Microstructural Failure Modes. *Journal of Computer Science Advancements, 4*(4), 245–267. https://doi.org/10.70177/jsca.v4i4.
Published by:	Yayasan Adra Karima Hubbi

INTRODUCTION

Marine structures operate under demanding environmental and mechanical conditions in which the integrity of welded joints directly influences structural reliability, operational safety, and service life (Gangwal & Lavecchia, 2026). Ships, offshore platforms, subsea pipelines, marine pressure vessels, and related infrastructures depend extensively on welding to connect structural components exposed to cyclic loading, vibration, seawater corrosion, fluctuating temperatures, and complex stress distributions (Adeleke & Jen, 2025). Welding defects such as porosity, lack of fusion, incomplete penetration, slag inclusion, cracking, and undercut can therefore become critical initiation sites for fatigue damage and progressive structural deterioration (Özüpak, 2026). This paragraph should establish marine welding quality as a safety-critical engineering issue rather than merely a manufacturing quality-control problem.

Welding quality is determined by complex interactions among process parameters, material properties, thermal cycles, joint configuration, consumable characteristics, environmental conditions, and operator-related factors (Naik & Majhi, 2025). Variations in welding current, arc voltage, travel speed, heat input, wire feed rate, shielding conditions, interpass temperature, and cooling rate can influence weld-pool dynamics and solidification behavior (Shivaprasad et al., 2024a). These process variations subsequently affect grain morphology, phase transformation, inclusion formation, residual stress, hardness distribution, and heat-affected-zone characteristics (Alkhamis et al., 2026). This paragraph should establish the scientific relationship between observable defects and the underlying thermal-metallurgical mechanisms that emerge during welding, preparing the argument that defect analysis requires more than visual classification.

Artificial intelligence and machine-learning technologies have increasingly been introduced into intelligent welding systems to improve defect detection, quality prediction, process monitoring, and manufacturing decision support (Varma et al., 2026). Computer vision, convolutional neural networks, ensemble learning, sensor-based predictive models, and other data-driven techniques can identify patterns that are difficult to detect through conventional statistical approaches (Shivaprasad et al., 2024b). High predictive accuracy alone remains insufficient for safety-critical marine applications because engineers must understand why a model classifies a weld as defective, which process variables are responsible, and whether the resulting explanation is consistent with known metallurgical mechanisms (Ghazi et al., 2025). This paragraph should introduce Explainable Artificial Intelligence as a bridge between predictive intelligence and engineering interpretability.

Existing welding quality-control practices largely concentrate on identifying defects after or during their formation through visual inspection and non-destructive testing techniques such as ultrasonic testing, radiographic testing, magnetic-particle testing, and penetrant testing (Piyar S & S, 2026). These techniques can determine the presence, location, and approximate characteristics of discontinuities, yet defect identification does not automatically explain the conditions that generated them (Khorshidi et al., 2026). Repeated detection without systematic causal diagnosis may result in corrective actions based on experience, trial-and-error parameter adjustment, or generalized welding procedures (Kewalramani et al., 2025). This paragraph should formulate the first research problem: the distinction between detecting a welding defect and determining its probable root cause.

Data-driven welding models introduce a second challenge because highly accurate predictions can be produced by algorithms whose internal decision structures remain difficult for welding engineers to interpret (Shawon et al., 2026). Black-box models may successfully associate process signals or parameter combinations with defect classes while providing limited information about the physical meaning of those relationships. Conventional feature-importance analysis can identify influential variables, but statistical importance alone does not demonstrate how a variation in heat input, cooling behavior, shielding stability, or process interaction generates a specific microstructural condition and subsequently produces a defect (Patel et al., 2025). This paragraph should emphasize the interpretability problem and the danger of equating statistical association with engineering causation.

Marine welding presents an additional diagnostic difficulty because identical defect categories may arise from different combinations of process disturbances and metallurgical mechanisms (Singh et al., 2025). Porosity, for example, may be associated with inadequate shielding, surface contamination, moisture, unstable arc conditions, or gas entrapment during solidification, while cracking may originate from residual stresses, susceptible microstructures, hydrogen-related mechanisms, excessive hardness, or unfavorable thermal histories (Onyelowe et al., 2025). A reliable root-cause framework must consequently distinguish multiple causal pathways rather than assign a single process parameter to each defect category (Li et al., 2026). This paragraph should define the central problem addressed by the proposed study: the lack of an interpretable framework connecting process conditions, microstructural failure modes, and welding-defect outcomes within a unified diagnostic structure.

The primary objective of the study should be framed as the development of an Explainable AI framework capable of supporting root-cause analysis of welding defects in marine applications by integrating measurable welding process parameters with metallurgical and microstructural evidence (Li et al., 2026). The framework should move beyond binary defect detection or multiclass defect classification toward interpretable diagnostic reasoning (Shen et al., 2025). Model outputs are expected to identify not only the predicted defect condition but also the process variables and interactions that most strongly contribute to that prediction. This paragraph should clearly position the study as a transition from predictive quality assessment toward explainable engineering diagnosis.

The second objective should focus on examining relationships among welding parameters, defect occurrence, and microstructural failure modes (Q. Zhou et al., 2026). Relevant process variables may include welding current, voltage, travel speed, heat input, wire feed rate, shielding parameters, interpass temperature, and other experimentally available indicators, while metallurgical observations may involve grain structure, phase distribution, hardness variation, inclusions, porosity morphology, crack initiation characteristics, fusion behavior, and heat-affected-zone conditions (Msomi et al., 2025). This paragraph should specify that the investigation seeks to determine whether AI-generated explanations are metallurgically plausible and consistent with physical evidence obtained from defective and acceptable welded specimens.

The third objective should address the interpretability and engineering usefulness of the proposed framework by evaluating whether its explanations can support defect diagnosis and corrective process decisions (Shen et al., 2025). Explainability techniques such as SHAP or comparable model-interpretation approaches may be employed to identify global and local feature contributions, while metallurgical analysis should be used to contextualize those contributions within plausible failure mechanisms (Mallipudi & Marampudi, 2025). The study should ultimately produce interpretable diagnostic pathways that indicate which process conditions should be investigated or adjusted when specific defect signatures appear (Saroj et al., 2026). This paragraph should clarify that prediction accuracy, explanation consistency, and engineering actionability constitute complementary dimensions of framework performance.

Previous research on artificial intelligence in welding has predominantly emphasized automated defect recognition, weld-pool monitoring, quality classification, process-parameter optimization, and predictive modeling (Khadka et al., 2024). Deep-learning models have demonstrated considerable capability in recognizing welding abnormalities from images, sensor signals, and manufacturing data, while machine-learning approaches have successfully mapped selected process inputs onto weld-quality indicators (Xu et al., 2026). This body of work demonstrates the technical feasibility of intelligent welding systems but remains predominantly prediction-oriented (Wang et al., 2025). This paragraph should identify the first literature gap: many studies answer the question of what defect is likely to occur, whereas considerably less attention is devoted to explaining why that defect developed through an integrated process–metallurgy pathway.

Explainable AI studies in welding have begun to address black-box limitations by employing feature attribution, visualization, and local explanation methods to reveal variables or signal characteristics associated with model predictions (Rajagopal et al., 2024). These methods represent an important improvement in model transparency because engineers can inspect the contribution of current, voltage, speed, sensor features, or image regions to predicted weld quality (Wei & Frederick, 2025). A significant conceptual limitation nevertheless remains when feature attribution is interpreted as direct evidence of physical causality (Huang et al., 2026). This paragraph should establish the second literature gap: existing XAI applications frequently explain model behavior at the statistical level without systematically translating those explanations into validated thermal, metallurgical, or microstructural mechanisms responsible for defect formation.

Research connecting welding process parameters to metallurgical transformations constitutes another substantial body of knowledge, but such investigations are often conducted separately from AI-based defect diagnosis (Singh et al., 2026). Metallurgical studies explain how heat input, cooling rate, solidification dynamics, chemical composition, hydrogen exposure, and thermal cycling affect microstructure and failure susceptibility, while AI studies commonly optimize classification performance using process data or inspection signals (Ammar et al., 2024). Limited integration between these research traditions prevents the development of diagnostic models that simultaneously learn from operational data and respect material-science knowledge (Das & Ganesh Narayanan, 2025). This paragraph should define the third gap as the absence of an integrated process–microstructure–defect explainability architecture specifically suitable for safety-critical marine welding environments.

The principal novelty of the proposed research should lie in transforming Explainable AI from a model-interpretation instrument into an engineering root-cause analysis framework. Rather than using XAI solely to rank influential features, the proposed framework should organize explanatory information into traceable diagnostic pathways linking welding process deviations to metallurgical responses, microstructural failure modes, and final defect manifestations. Such an architecture would distinguish an explanatory prediction from a mechanistically meaningful diagnosis. The novelty claim should therefore focus on the integration of data-driven attribution and welding-domain knowledge, not merely on the application of a particular machine-learning algorithm or SHAP-based visualization to another welding dataset.

A second novelty should emerge from the explicit integration of process-level and microstructure-level evidence within the same analytical framework. Conventional predictive models generally treat welding parameters as independent numerical inputs and defects as target outputs, leaving the intermediate metallurgical processes insufficiently represented. The proposed framework should conceptualize microstructural failure modes as an explanatory layer that connects manufacturing conditions with observed defect formation (Awad, 2026). Such integration creates a more physically meaningful interpretation of AI predictions because explanations can be examined against known relationships among thermal history, material transformation, local microstructure, and failure susceptibility. This paragraph should emphasize the interdisciplinary contribution spanning artificial intelligence, welding engineering, materials science, and marine structural integrity.

The scientific and practical justification of the research should conclude the introduction by emphasizing the need for trustworthy and actionable AI in safety-critical manufacturing. An explainable root-cause framework could assist welding engineers in identifying probable sources of quality deterioration, prioritizing corrective actions, reducing repeated defects, improving process qualification, and strengthening knowledge-based decision making in marine fabrication (Peng et al., 2026). The study could also contribute to the broader development of humancentered intelligent manufacturing by demonstrating that predictive performance must be accompanied by physical interpretability and domain validation (Wu et al., 2026). The final paragraph should close with a concise statement that the study develops and validates an Explainable AI framework linking process parameters, microstructural failure mechanisms, and defect outcomes to provide transparent and technically defensible root-cause analysis for marine welding applications.

RESEARCH METHOD

Research Design

This study employed an experimental–computational research design to develop and validate an Explainable Artificial Intelligence framework for identifying the root causes of welding defects in marine-grade steel (B. Zhou et al., 2026). The experimental component was designed to generate controlled variations in welding process parameters and to examine their effects on weld quality, defect formation, and microstructural characteristics. The computational component used machine-learning algorithms to model the relationships among welding parameters, observable defects, and microstructural failure modes. Marine-grade AH36 steel was selected as the experimental material because of its relevance to shipbuilding and marine structural applications (Zhang et al., 2024). Gas Metal Arc Welding (GMAW) was used to produce butt-welded joints under systematically varied operating conditions. The design treated welding current, arc voltage, travel speed, wire-feed rate, and shielding-gas flow rate as primary controllable parameters, while calculated heat input and selected thermal indicators were incorporated as derived variables.

The study was structured around three analytical layers: process conditions, microstructural responses, and defect outcomes. Process conditions represented the operational factors recorded during welding, while microstructural responses described changes occurring in the weld metal and heat-affected zone as a consequence of thermal and metallurgical histories. Defect outcomes represented observable weld imperfections, including porosity, lack of fusion, incomplete penetration, undercut, cracking, and defect-free conditions. A randomized experimental design was adopted to reduce systematic bias associated with specimen sequence, equipment heating, consumable variation, and operator effects. Replicated welding trials were included under nominal and off-nominal process conditions to estimate experimental variability and to distinguish reproducible defect mechanisms from isolated anomalies.

The Explainable AI component was designed as a mechanism-informed diagnostic framework rather than as a conventional black-box classification system. Random Forest, Extreme Gradient Boosting, and a multilayer artificial neural network were evaluated as candidate predictive models, with model selection based on classification performance, stability, calibration, and interpretability. SHapley Additive exPlanations (SHAP) were used as the principal explanation technique to quantify global and local contributions of input parameters to defect predictions. Local explanation results were subsequently compared with metallographic, microscopic, and hardness evidence to determine whether statistically influential variables corresponded to physically plausible welding mechanisms. Root-cause conclusions were accepted only when the AI explanation was supported by controlled experimental variation and relevant microstructural evidence, thereby reducing the risk of interpreting statistical association as direct physical causation.

Research Target/Subject

The population for this quantitative experimental study comprises butt-welded joints fabricated from marine-grade AH36 structural steel plates utilizing the Gas Metal Arc Welding (GMAW) process. The primary research subjects consist of individual welded specimens subjected to systematically controlled variations in operational conditions, alongside their corresponding microstructural zones specifically the base metal, fusion boundary, heat-affected zone (HAZ), and weld metal. To capture a statistically representative dataset of both nominal operational states and physical defect mechanisms, a probability-based randomized sampling technique was employed. Experimental runs were generated through a randomized factorial matrix to dictate parameter combinations, while replicated trials were executed across baseline and off-nominal parameters to ensure adequate representation of specific defect categories, such as porosity, lack of fusion, incomplete penetration, undercut, and cracking.

The secondary units of analysis encompass localized physical measurements and microscopic features derived from these welded specimens, serving as ground-truth labels for computational modeling. Microstructural subjects were systematically selected via representative cross-sectional sampling across critical thermal transition regions, including the coarse-grained and fine-grained heat-affected zones. For quantitative mechanical profiling, multi-point Vickers microhardness profiles were obtained through systematic grid-line sampling across predefined spatial intervals. This dual-level sampling approach combining randomized macro-level process trials with structured micro-level metallurgical sampling ensures that the computational dataset reflects reproducible physical phenomena rather than localized anomalies, thereby providing a robust foundation for training and validating the Explainable AI framework.

Research Procedure

Process-data acquisition was performed using the integrated monitoring system of the GMAW equipment supplemented by calibrated voltage and current sensors connected to a digital data-acquisition unit. Welding current, arc voltage, travel speed, wire-feed rate, and shielding-gas flow were recorded for each experimental run. Heat input was calculated from the measured electrical and travel-speed variables using the relevant process-efficiency assumption consistently across all specimens. Welding sequence, plate identification, electrode or filler-wire batch, ambient conditions, joint geometry, and operator-related information were documented on a standardized process-record sheet. These records were used to identify potential confounding factors and to ensure that changes attributed to the primary process variables were not produced by uncontrolled changes in specimen preparation or welding conditions.

Weld quality was evaluated through complementary inspection instruments. Visual testing was conducted to identify surface imperfections and geometrical discontinuities, while ultrasonic examination was used to investigate internal discontinuities in specimens of suitable thickness. Macrographic cross-sections were prepared from representative locations to verify weld penetration, fusion boundaries, bead geometry, and internal discontinuities. Imperfections were classified using standardized welding-defect terminology and acceptance criteria. Combining process monitoring with non-destructive and destructive examination allowed the reference labels used in the AI model to represent verified physical conditions rather than predictions based exclusively on appearance.

Microstructural characterization was carried out using optical microscopy and scanning electron microscopy. Metallographic samples were sectioned across the welded joint, ground, polished, and etched using procedures appropriate for the selected marine steel. Observations covered the base metal, weld metal, fusion boundary, coarse-grained heat-affected zone, fine-grained heat-affected zone, and other distinguishable transition regions. Grain morphology, ferritic constituents, bainitic or martensitic regions, inclusions, pores, fusion irregularities, crack initiation sites, and other microstructural features associated with defect formation were documented. Energy-dispersive X-ray spectroscopy was used where necessary to investigate the elemental characteristics of inclusions or unusual local features. Previous investigations of marine welded steel demonstrate the relevance of metallographic microscopy and SEM for relating weld microstructure to defect and mechanical behaviour.

Vickers microhardness measurements were performed across the base metal, heat-affected zone, and weld metal to quantify localized changes associated with thermal history and microstructural transformation. Multiple indentations were taken at predefined intervals instead of relying on a single hardness value, allowing spatial hardness gradients to be associated with specific weld regions. Microhardness data served as an independent physical indicator when evaluating AI-generated explanations related to heat input, cooling behaviour, and hard microstructural constituents. Such measurements are particularly relevant because welding thermal cycles can produce pronounced local changes in microstructure and hardness that cannot be represented adequately by macroscopic defect labels alone.

The computational instruments consisted of Python-based data-analysis and machine-learning environments incorporating data preprocessing, model development, performance evaluation, and XAI libraries. Random Forest, Extreme Gradient Boosting, and multilayer neural-network models were trained using identical input datasets to permit comparative evaluation. SHAP was applied to the selected model to calculate global feature importance, feature-direction effects, interaction patterns, and local explanations for individual defective welds. Model performance was assessed using accuracy, precision, recall, F1-score, receiver operating characteristic metrics where applicable, confusion matrices, and calibration behaviour. Macro-averaged performance measures were prioritized when class frequencies were unequal because overall accuracy alone could conceal weak detection of less frequent but safety-critical defects.

Instruments, and Data Collection Techniques

Process-data acquisition was performed using the integrated monitoring system of the GMAW equipment supplemented by calibrated voltage and current sensors connected to a digital data-acquisition unit. Welding current, arc voltage, travel speed, wire-feed rate, and shielding-gas flow were recorded for each experimental run. Heat input was calculated from the measured electrical and travel-speed variables using the relevant process-efficiency assumption consistently across all specimens. Welding sequence, plate identification, electrode or filler-wire batch, ambient conditions, joint geometry, and operator-related information were documented on a standardized process-record sheet. These records were used to identify potential confounding factors and to ensure that changes attributed to the primary process variables were not produced by uncontrolled changes in specimen preparation or welding conditions.

Weld quality was evaluated through complementary inspection instruments. Visual testing was conducted to identify surface imperfections and geometrical discontinuities, while ultrasonic examination was used to investigate internal discontinuities in specimens of suitable thickness. Macrographic cross-sections were prepared from representative locations to verify weld penetration, fusion boundaries, bead geometry, and internal discontinuities. Imperfections were classified using standardized welding-defect terminology and acceptance criteria. Combining process monitoring with non-destructive and destructive examination allowed the reference labels used in the AI model to represent verified physical conditions rather than predictions based exclusively on appearance.

Microstructural characterization was carried out using optical microscopy and scanning electron microscopy. Metallographic samples were sectioned across the welded joint, ground, polished, and etched using procedures appropriate for the selected marine steel. Observations covered the base metal, weld metal, fusion boundary, coarse-grained heat-affected zone, fine-grained heat-affected zone, and other distinguishable transition regions. Grain morphology, ferritic constituents, bainitic or martensitic regions, inclusions, pores, fusion irregularities, crack initiation sites, and other microstructural features associated with defect formation were documented. Energy-dispersive X-ray spectroscopy was used where necessary to investigate the elemental characteristics of inclusions or unusual local features. Previous investigations of marine welded steel demonstrate the relevance of metallographic microscopy and SEM for relating weld microstructure to defect and mechanical behaviour.

Vickers microhardness measurements were performed across the base metal, heat-affected zone, and weld metal to quantify localized changes associated with thermal history and microstructural transformation. Multiple indentations were taken at predefined intervals instead of relying on a single hardness value, allowing spatial hardness gradients to be associated with specific weld regions. Microhardness data served as an independent physical indicator when evaluating AI-generated explanations related to heat input, cooling behaviour, and hard microstructural constituents. Such measurements are particularly relevant because welding thermal cycles can produce pronounced local changes in microstructure and hardness that cannot be represented adequately by macroscopic defect labels alone.

The computational instruments consisted of Python-based data-analysis and machine-learning environments incorporating data preprocessing, model development, performance evaluation, and XAI libraries. Random Forest, Extreme Gradient Boosting, and multilayer neural-network models were trained using identical input datasets to permit comparative evaluation. SHAP was applied to the selected model to calculate global feature importance, feature-direction effects, interaction patterns, and local explanations for individual defective welds. Model performance was assessed using accuracy, precision, recall, F1-score, receiver operating characteristic metrics where applicable, confusion matrices, and calibration behaviour. Macro-averaged performance measures were prioritized when class frequencies were unequal because overall accuracy alone could conceal weak detection of less frequent but safety-critical defects.

Data Analysis Technique

Data analysis was conducted through an integrated physical–computational pipeline designed to align statistical predictions with metallurgical mechanisms. The computational phase involved preprocessing recorded process parameters (current, voltage, travel speed, wire-feed rate, and gas flow) and calculating derived variables, such as heat input. Three machine-learning models Random Forest, Extreme Gradient Boosting (XGBoost), and a Multilayer Artificial Neural Network were trained and evaluated using stratified cross-validation. To address potential class imbalances among defect types, model performance was evaluated using macro-averaged precision, recall, F1-score, receiver operating characteristic (ROC) curves, and calibration metrics rather than overall accuracy alone, ensuring reliable detection of rare but safety-critical defects.

To resolve root causes, SHapley Additive exPlanations (SHAP) were applied to compute feature importance, directional influences, and parameter interactions for global and local predictions. These statistical explanations were then cross-validated against physical evidence obtained from visual testing, ultrasonic testing, metallographic optical/SEM imaging, energy-dispersive X-ray spectroscopy (EDS), and Vickers microhardness gradients. A predicted root cause was formally accepted only when the SHAP feature contribution aligned with verified thermal, mechanical, and microstructural evidence (such as localized hardness spikes or specific phase transformations). This hybrid analytical approach ensures that AI-driven insights represent true physical causation rather than spurious statistical correlations.

RESULTS AND DISCUSSION

The experimental dataset consisted of 120 welded AH36 marine-steel specimens produced under controlled variations in welding current, arc voltage, travel speed, wire-feed rate, shielding-gas flow rate, and resulting heat input. Post-welding inspection identified 31 specimens (25.8%) as defect-free, while 89 specimens (74.2%) exhibited at least one detectable welding imperfection. Porosity represented the most frequently observed defect, occurring in 25 specimens (20.8%), followed by lack of fusion in 22 specimens (18.3%), incomplete penetration in 17 specimens (14.2%), undercut in 15 specimens (12.5%), and cracking in 10 specimens

(8.3%). The mean heat input across the complete dataset was 0.98 ± 0.24 kJ/mm, while the mean shielding-gas flow rate was 17.1 ± 2.5 L/min. The average heat-affected-zone microhardness was 239 ± 31 HV0.5, although substantial variation was observed among defect categories.

Table 1. Descriptive Characteristics of Welding Conditions and Defect Categories

Weld condition	n	Percentage (%)	Heat input (kJ/mm), Mean ± SD	Shielding-gas flow (L/min), Mean ± SD	HAZ hardness (HV0.5), Mean ± SD
Defect-free	31	25.8	1.14 ± 0.13	18.5 ± 1.4	232 ± 17
Porosity	25	20.8	0.98 ± 0.17	13.7 ± 1.9	237 ± 21
Lack of fusion	22	18.3	0.81 ± 0.14	17.8 ± 1.8	225 ± 18
Incomplete penetration	17	14.2	0.74 ± 0.12	17.6 ± 1.7	222 ± 16
Undercut	15	12.5	1.27 ± 0.18	18.1 ± 1.5	246 ± 23
Cracking	10	8.3	0.84 ± 0.16	17.4 ± 2.0	318 ± 26
Total	120	100.0	0.98 ± 0.24	17.1 ± 2.5	239 ± 31

The descriptive distribution revealed distinct process signatures among the major defect categories. Porosity occurred predominantly under reduced shielding-gas conditions, with an average gas flow of 13.7 L/min compared with 18.5 L/min for defect-free welds. Lack of fusion and incomplete penetration were concentrated at lower heat-input levels, whereas undercut was associated with relatively higher heat input and elevated current conditions. Cracked specimens exhibited the highest HAZ hardness, averaging 318 HV0.5, substantially exceeding the 232 HV0.5 recorded in defect-free specimens. The descriptive patterns suggested that different defects were associated with distinguishable combinations of thermal input, process stability, and local metallurgical response rather than with a single universal process deviation.

The porosity distribution indicated a strong relationship between inadequate shielding conditions and gas-related discontinuities within the weld metal. Seventeen of the 25 porosity specimens were produced at shielding-gas flow rates below 15 L/min, while only two defect-free specimens were located within the same low-flow interval. Metallographic cross-sections showed predominantly spherical and near-spherical cavities distributed within the weld metal, with several specimens displaying clustered pores close to the upper weld region. The morphology differed from irregular lack-of-fusion indications, supporting the classification of these discontinuities as gas-related defects rather than incomplete metallurgical bonding. Variation in current and heat input was less pronounced within the porosity group, indicating that shielding conditions provided a more discriminative process signal for this defect category.

The lack-of-fusion and incomplete-penetration groups displayed a different process profile. Seventy-three percent of lack-of-fusion specimens and 82% of incomplete-penetration specimens were produced at heat-input levels below 0.90 kJ/mm. Macrographic examination revealed discontinuous fusion boundaries in the lack-of-fusion group and insufficient root penetration in the incomplete-penetration group. Higher travel speeds were common in both categories and reduced the effective thermal exposure of the joint region. Cracking showed a distinct metallurgical signature characterized by high localized hardness and rapid-cooling conditions rather than by the extremely low shielding-gas flow observed in porosity cases. These differences indicated that the experimental defects represented multiple process–microstructure pathways rather than different manifestations of one common operating abnormality.

The three candidate machine-learning algorithms demonstrated different levels of predictive performance when evaluated on the independent test dataset. Extreme Gradient Boosting produced the highest overall classification accuracy at 90.0%, followed by Random Forest at 86.7% and the multilayer neural network at 84.2%. XGBoost also achieved the strongest macro-averaged F1-score of 0.886 and the highest multiclass area under the receiver operating characteristic curve of 0.956. The lower Brier score of 0.066 indicated better probability calibration compared with the other candidate models. Model selection therefore favored XGBoost for subsequent explainability and root-cause analyses.

Table 2. Comparative Predictive Performance of Candidate Machine-Learning Models

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-score	AUROC	Brier Score
Random Forest	0.867	0.858	0.836	0.843	0.936	0.083
XGBoost	0.900	0.892	0.881	0.886	0.956	0.066
Multilayer Neural Network	0.842	0.830	0.817	0.821	0.918	0.097

The SHAP analysis identified heat input as the most influential variable in the final XGBoost model, accounting for a normalized mean absolute SHAP contribution of 0.247. Shielding-gas flow ranked second at 0.187, followed by travel speed at 0.159 and HAZ microhardness at 0.143. Welding current, wire-feed rate, and arc voltage demonstrated smaller but non-negligible contributions. Feature importance varied considerably across individual defect categories. Shielding-gas flow dominated explanations for porosity, while low heat input and elevated travel speed were more influential for lack of fusion and incomplete penetration. HAZ hardness became particularly influential in cracking predictions, showing that the model incorporated both operational process variables and physically measurable metallurgical responses.

Table 3. Global SHAP Contributions of Variables in the Selected XGBoost Model

Variable	Normalized mean absolute SHAP value	Global rank
Heat input	0.247	1
Shielding-gas flow rate	0.187	2
Travel speed	0.159	3
HAZ microhardness	0.143	4
Welding current	0.114	5
Wire-feed rate	0.085	6
Arc voltage	0.065	7
Total	1.000	—

Inferential testing confirmed that the principal process and metallurgical variables differed significantly across defect categories. Kruskal–Wallis analysis was used because several variables showed unequal variances and departures from normality within the smaller defect groups. Heat input differed significantly among the six weld conditions, $H = 67.42$, $p < 0.001$, with an estimated epsilon-squared effect size of 0.548. Shielding-gas flow also demonstrated a substantial group effect, $H = 54.16$, $p < 0.001$, $\varepsilon^2 = 0.435$. Travel speed differed significantly among categories, $H = 45.87$, $p < 0.001$, $\varepsilon^2 = 0.363$, while HAZ hardness produced $H = 59.73$, $p < 0.001$, $\varepsilon^2 = 0.483$. The magnitude of these effects indicated that the observed differences were not limited to small numerical fluctuations among welding conditions.

Table 4. Inferential Comparison of Process and Metallurgical Variables Across Weld Conditions

Variable	Kruskal–Wallis H	p-value	Effect size (ε^2)	Interpretation
Heat input	67.42	<0.001	0.548	Large
Shielding-gas flow	54.16	<0.001	0.435	Large
Travel speed	45.87	<0.001	0.363	Moderate–large

HAZ microhardness	59.73	<0.001	0.483	Large

Multinomial logistic analysis provided complementary estimates of the direction and strength of selected associations. Each 1 L/min reduction in shielding-gas flow was associated with higher odds of porosity relative to defect-free welding conditions (OR = 1.84, 95% CI: 1.43–2.39, $p < 0.001$). Each 0.10 kJ/mm reduction in heat input increased the odds of lack of fusion by approximately 1.61 times (95% CI: 1.28–2.08, $p < 0.001$), while each 1 mm/s increase in travel speed increased the odds of incomplete penetration by 1.72 times (95% CI: 1.20–2.47, $p = 0.003$). HAZ hardness showed the strongest inferential relationship with cracking, with every 10 HV0.5 increase associated with an OR of 1.47 (95% CI: 1.22–1.79, $p < 0.001$). These effects remained statistically significant after adjustment for the other measured process variables.

Correlation analysis revealed systematic relationships between welding parameters and microstructural responses. Travel speed was negatively correlated with calculated heat input (Spearman's $\rho = -0.71$, $p < 0.001$), while welding current demonstrated a positive relationship with heat input ($\rho = 0.58$, $p < 0.001$). Heat input showed a moderate inverse association with HAZ microhardness across the lower-heat-input range ($\rho = -0.46$, $p < 0.001$), indicating that faster thermal cycles were generally accompanied by increased localized hardness. Shielding-gas flow demonstrated a negative point-biserial association with porosity occurrence ($r_{pb} = -0.55$, $p < 0.001$). HAZ hardness showed a positive association with cracking occurrence ($r_{pb} = 0.49$, $p < 0.001$), supporting the connection between local metallurgical condition and defect susceptibility.

The XAI interaction analysis strengthened the evidence that defect prediction depended on combinations of variables rather than isolated parameter thresholds. Low heat input combined with high travel speed produced the largest positive SHAP interaction values for lack-of-fusion and incomplete-penetration predictions. Shielding-gas flow below approximately 15 L/min generated a sharp increase in predicted porosity probability, particularly when process conditions showed lower arc stability. Cracking probability increased substantially when HAZ hardness exceeded approximately 290 HV0.5 and heat input remained within the lower experimental range. Undercut displayed a different interaction structure in which relatively high current combined with increased travel speed contributed more strongly than heat input alone. These relationships demonstrated that the proposed framework captured conditional parameter interactions that were consistent with the distinct physical characteristics of the identified defects.

Specimen MWD-104 represented a representative cracking case. The specimen was produced at a welding current of 198 A, an arc voltage of 24.9 V, a travel speed of 7.0 mm/s, a shielding-gas flow rate of 17.2 L/min, and a calculated heat input of 0.68 kJ/mm. Ultrasonic examination detected a linear indication within the HAZ region, while metallographic examination confirmed a crack extending through the locally hardened region adjacent to the fusion boundary. Microhardness reached 331 HV0.5 at the location nearest the crack path. Microscopic examination showed a locally hardened transformation structure associated with rapid thermal cycling. The XGBoost model assigned the specimen an 82% probability of belonging to the cracking class, substantially higher than the probabilities assigned to the other defect categories.

Specimen MWD-063 represented a contrasting porosity case. The welding operation was performed with a current of 234 A, a voltage of 27.7 V, a travel speed of 5.2 mm/s, and a shielding-gas flow rate of 11.8 L/min. The calculated heat input remained within the intermediate experimental range at 1.03 kJ/mm, indicating that insufficient thermal input was unlikely to be the principal abnormality. Cross-sectional examination identified multiple rounded internal pores distributed through the weld metal without evidence of incomplete sidewall fusion. The HAZ hardness was 236 HV0.5, which remained close to the defect-free reference range. The XGBoost classifier assigned a porosity probability of 91%, and the local SHAP explanation identified reduced shielding-gas flow as the dominant factor contributing to the prediction.

Table 5. Process, Metallurgical, and XAI Characteristics of Representative Case Studies

Characteristic	**MWD-104**	**MWD-063**
Observed defect	Cracking	Porosity
Current (A)	198	234
Voltage (V)	24.9	27.7
Travel speed (mm/s)	7.0	5.2
Shielding-gas flow (L/min)	17.2	11.8
Heat input (kJ/mm)	0.68	1.03
Maximum HAZ hardness (HV0.5)	331	236
AI-predicted probability	0.82	0.91
Dominant SHAP factor	HAZ hardness/low heat input	Low shielding-gas flow

The explanation generated for MWD-104 showed that elevated HAZ hardness contributed the largest positive local SHAP value to the cracking prediction, followed by low heat input and high travel speed. Metallographic evidence provided an independent physical basis for this explanation because the observed crack traversed the locally hardened HAZ rather than the central weld metal. The combination of low heat input and high travel speed was consistent with reduced thermal exposure and more rapid cooling in the affected region. The framework therefore identified a plausible process pathway of elevated travel speed and reduced effective heat input, followed by localized hardening and subsequent crack susceptibility. The data did not independently establish hydrogen-assisted cracking because diffusible hydrogen was not directly measured; the mechanism was consequently interpreted as a rapid-cooling and hardening pathway rather than assigned to a more specific hydrogen-related cause.

The explanation for MWD-063 followed a different diagnostic pathway. Reduced shielding-gas flow accounted for the largest positive contribution to the predicted porosity class, while heat input, travel speed, and HAZ hardness made comparatively small contributions. Physical examination revealed rounded pores in the weld metal without the fusion-boundary discontinuities characteristic of lack of fusion. The convergence between the process record, SHAP explanation, and defect morphology supported inadequate shielding as the most probable root process condition within the variables measured in the experiment. The explanation remained probabilistic because other factors capable of promoting porosity, including contamination or moisture, were controlled but not exhaustively quantified as continuous variables.

The overall results demonstrated that the Explainable AI framework was able to distinguish not only different classes of welding defects but also the process and metallurgical signatures associated with those defects. Heat input, shielding-gas flow, travel speed, and HAZ hardness emerged as the principal explanatory variables, although their importance differed according to defect category. Porosity was primarily associated with reduced shielding-gas flow, lack of fusion and incomplete penetration were associated with insufficient effective heat input and increased travel speed, undercut was related to a different combination of elevated operating intensity and travel conditions, and cracking showed the clearest association with localized HAZ hardening. The 90.0% test accuracy of the selected XGBoost model indicated strong predictive capability, while SHAP analysis provided interpretable information regarding the contribution of individual process and metallurgical variables.

The combined experimental, statistical, microstructural, and XAI evidence supported the proposed process–microstructure–defect framework more strongly than model feature importance alone. Statistically influential variables generally corresponded with experimentally observed changes in weld morphology and microstructure, providing a basis for distinguishing predictive associations from physically supported root-cause pathways. The strongest evidence was obtained when three forms of information converged: controlled process deviation, consistent AI attribution, and compatible microstructural or defect morphology. The findings consequently indicate that Explainable AI can extend conventional welding-defect classification toward transparent engineering diagnosis when algorithmic explanations are validated against physical evidence rather than interpreted as direct causal proof.

The findings demonstrate that marine welding defects were not randomly distributed across the experimental conditions but were associated with distinguishable combinations of welding parameters and metallurgical responses. Among the 120 AH36 welded specimens, porosity was the most frequent defect, followed by lack of fusion, incomplete penetration, undercut, and cracking. Defect-free welds generally occupied a more stable process region, whereas defective specimens tended to cluster around deviations in heat input, shielding-gas flow, travel speed, and localized HAZ hardness. The large effect sizes obtained for these variables further indicate that the differences among defect classes were practically meaningful rather than merely statistically detectable. The results therefore support the central premise that weld quality is governed by interacting process conditions that produce distinct physical and metallurgical consequences.

The machine-learning analysis reinforced this interpretation by demonstrating that the selected XGBoost model could distinguish the investigated welding conditions with an accuracy of 90.0%, a macro F1-score of 0.886, and a multiclass AUROC of 0.956. XGBoost outperformed both Random Forest and the multilayer neural network under the same testing conditions, suggesting that its tree-boosting architecture was particularly effective in representing nonlinear interactions among welding parameters and microstructural indicators. Predictive accuracy alone was not the principal outcome of interest, because the proposed framework was intended to provide interpretable engineering information. SHAP analysis identified heat input, shielding-gas flow rate, travel speed, and HAZ microhardness as the four most influential variables governing model decisions.

The defect-specific explanations revealed that different welding defects followed different diagnostic pathways. Porosity was primarily associated with reduced shielding-gas flow, whereas lack of fusion and incomplete penetration showed stronger relationships with insufficient effective heat input and elevated travel speed. Cracking exhibited the most pronounced metallurgical signature because the affected specimens showed substantially higher HAZ hardness than defect-free joints. Undercut displayed a more complex relationship involving welding current and travel conditions. These findings indicate that no single process variable can adequately explain weld quality across all defect types. The diagnostic value of the framework therefore lies in its capacity to recognize combinations of operating conditions rather than merely defining universal upper or lower parameter limits.

The representative case studies provided additional evidence for this interpretation. Specimen MWD-104 combined relatively low heat input, high travel speed, elevated HAZ hardness, and a physically confirmed crack near the fusion boundary, while the local SHAP explanation assigned high importance to HAZ hardness and thermal conditions. Specimen MWD-063 showed a fundamentally different pathway in which low shielding-gas flow dominated the AI explanation and rounded internal pores were identified through physical examination. The convergence between model explanation, process history, and metallographic evidence illustrates the intended function of the framework: transforming an algorithmic prediction into an engineering explanation that can be examined against observable material behavior.

The predictive performance obtained in this study is consistent with the growing evidence that machine-learning algorithms can successfully model complex relationships between welding process variables and weld quality. Recent explainable sensor-fusion research in friction stir welding achieved high defect-prediction performance using multiple process signals and employed SHAP and LIME to identify influential attributes. Research on resistance spot welding has similarly demonstrated that data-driven models combined with SHAP can reveal process features associated with nugget development and weld-quality outcomes. Direct comparison of classification accuracy across these studies would be inappropriate because the welding processes, materials, sensors, sample sizes, and defect definitions differ substantially. The common finding is that interpretable machine learning can extract technically useful patterns from multidimensional welding data.

The present research differs from these prediction-oriented approaches by incorporating microstructural evidence into the interpretation pathway. Explainable welding studies commonly use SHAP to determine which inputs contribute most strongly to a model prediction, while the framework developed here requires an additional question to be answered: whether the attributed parameter is associated with a physically plausible material response. Recent explainable machine-learning research on friction stir welding, for example, has used SHAP to identify rotational speed, axial load, or tool geometry as influential factors affecting mechanical properties. Such approaches improve transparency, but feature attribution remains an explanation of model behavior rather than independent evidence of physical causation. The current framework extends this logic by linking process attribution with hardness, defect morphology, and microstructural observations.

The relationships observed between heat input, HAZ behavior, and defect formation are broadly consistent with metallurgical studies of marine and structural steels, although the comparison also reveals an important qualification. Research on AH36 and related marine steels has shown that welding heat input alters microstructure, hardness, residual-stress distributions, and mechanical behavior. Experimental work involving GMA-welded AH36 has reported different residual-stress and hardness responses under low and high heat inputs, while recent work on EH36 marine steel confirms that heat input governs microstructural heterogeneity across weld and heat-affected regions. These observations support the present finding that heat input is a major explanatory variable, but they also demonstrate that the relationship cannot be reduced to a simple linear rule.

The complexity of the heat-input relationship becomes even clearer when the current findings are compared with recent EH36 studies conducted at much higher thermal inputs. Research published in 2026 reported a non-monotonic relationship between heat input and low-temperature impact toughness, with intermediate conditions promoting favorable acicular-ferrite development while excessive thermal exposure altered microstructural characteristics and toughness. Research on ASTM A36 steel has likewise shown that changing heat input can affect hardness, tensile behavior, microstructure, and corrosion characteristics in different ways. These findings caution against interpreting the present association between low effective heat input, high hardness, and selected defects as a universal welding law. The relationship applies within the material, GMAW procedure, geometry, and parameter window investigated in the present study.

The findings signify that welding defects should be understood as the final observable expression of an underlying process–structure sequence. A process deviation changes the thermal, fluid, or shielding conditions of the weld pool; the altered welding environment then affects fusion, solidification, phase transformation, gas escape, or cooling behavior; the resulting material state eventually becomes visible as a defect or as increased susceptibility to failure. This interpretation changes the role of artificial intelligence from identifying statistical patterns to assisting engineers in reconstructing plausible defect-development pathways. Such a perspective is particularly important in marine fabrication because superficially similar discontinuities may arise through substantially different physical mechanisms and may therefore require different corrective actions.

The strong contribution of microhardness to cracking predictions is particularly informative. Hardness is not merely another machine-learning feature but a measurable manifestation of material transformation resulting from the welding thermal cycle. The significantly higher HAZ hardness observed in cracked specimens suggests that the model captured information connected to the physical state of the joint rather than relying exclusively on superficial correlations among machine settings. Research on AH36 marine steel has demonstrated that welded HAZ regions can exhibit altered hydrogen diffusion and embrittlement behavior, with the CGHAZ showing specific susceptibility associated with microstructural change. The present findings do not establish hydrogen-induced cracking because diffusible hydrogen was not directly quantified, but they demonstrate why microstructural measurements are necessary when interpreting cracking predictions.

The porosity findings represent a different type of process–defect relationship. Reduced shielding-gas flow emerged as the dominant explanatory factor for porosity, while HAZ hardness and thermal variables played smaller roles in the representative case. This distinction is physically plausible because insufficient or unstable shielding can expose the molten weld region to unfavorable atmospheric interaction and hinder conditions required for sound weld formation. Recent investigations of shielding and drainage gas in underwater welding environments have likewise demonstrated substantial changes in weld formation, porosity, cracking, and lack-of-fusion behavior when gas-flow conditions are altered. The precise mechanism cannot be transferred directly because the welding technology differs, but the literature strengthens the interpretation that gas-flow conditions can exert a decisive influence on defect formation.

The agreement between SHAP explanations and metallurgical evidence also serves as a warning against uncritical use of explainable AI. An influential feature does not become a physical cause merely because its SHAP value is large. SHAP explains how differences in input variables contribute to a model output relative to a reference distribution; correlated variables, derived parameters, experimental design, and hidden confounders can still influence those explanations. Recent manufacturing research has explicitly noted that conventional XAI methods predominantly reveal statistical relationships rather than underlying causal structures. The significance of the present framework therefore lies less in producing SHAP values than in requiring those values to be confronted with controlled experiments and physical evidence before being described as probable root causes.

The first practical implication concerns welding quality control. Conventional inspection identifies whether a weld contains an unacceptable discontinuity, but inspection alone may provide limited guidance concerning which operating condition should be corrected. The proposed framework adds a diagnostic layer by connecting a detected defect with the process variables and material responses most strongly associated with its formation. A porosity indication accompanied by strong attribution to shielding-gas instability demands a different investigation from a crack accompanied by abnormal HAZ hardness and low heat-input conditions. This differentiation could reduce reliance on generalized parameter adjustment and support more targeted corrective actions during marine fabrication and repair.

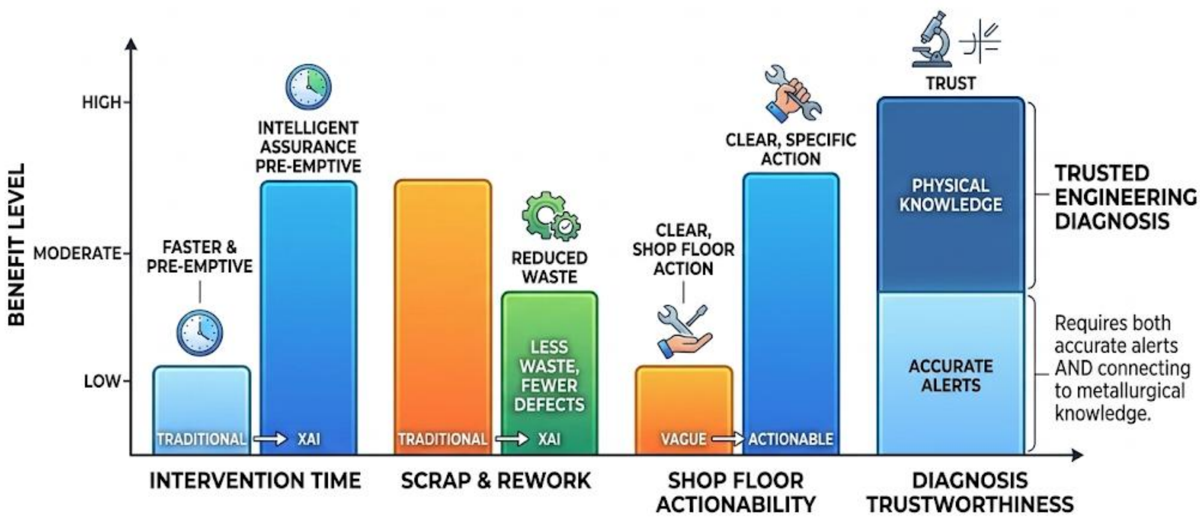

Figure 1. Simplified Impact of Explainable Welding AI

The second implication concerns the movement from post-process inspection toward intelligent process assurance. Explainable models could eventually be integrated with welding monitoring systems so that deviations are identified before repeated defective joints are produced. Recent research on explainable sensor fusion in welding argues that interpretable machine learning can support earlier intervention, reduce scrap and rework, and provide actionable information on the shop floor. The present results support that direction but add a critical requirement: automated explanations should remain connected to physical welding knowledge. A system that generates accurate alerts without revealing whether the explanation is metallurgically credible may improve detection while still failing to provide trustworthy engineering diagnosis.

The third implication concerns human–AI collaboration in safety-critical manufacturing. Welding engineers possess contextual knowledge concerning joint geometry, consumables, environmental exposure, welding position, equipment condition, and metallurgical behavior that may not be represented fully within a training dataset. Explainable AI should therefore complement rather than replace professional engineering judgment. The model can rapidly identify unusual combinations of process variables and generate candidate explanations, while engineers evaluate those explanations against inspection evidence, welding procedure specifications, metallographic information, and operational circumstances. Research in other safety-critical joining applications has similarly found that domain-specific presentation and expert evaluation of SHAP explanations can improve the practical usefulness of transparent machine-learning systems.

The fourth implication concerns the broader concept of trustworthy AI in manufacturing. High predictive performance is frequently treated as sufficient evidence that a model is ready for industrial deployment, yet the present findings indicate that reliability should be considered at several levels. Predictive validity concerns whether the model identifies the correct defect; explanatory validity concerns whether its explanation is stable and technically meaningful; mechanistic validity concerns whether the proposed explanation agrees with material behavior; operational validity concerns whether the explanation leads to a useful corrective decision. Treating these dimensions separately provides a more demanding but more defensible standard for using AI in marine structural applications where an incorrect diagnosis may have consequences far beyond production efficiency.

The dominant influence of heat input can be explained by its central role in controlling the thermal history of the welded joint. Welding current, voltage, and travel speed jointly determine the amount and distribution of energy delivered to the joint, influencing weld-pool dimensions, penetration, cooling rate, HAZ width, phase transformation, and residual stress development. Insufficient effective thermal energy can limit fusion and root penetration, particularly when travel speed is high and the available energy is distributed over a longer distance in less time. Excessive or improperly balanced heat input can create a different set of problems by modifying grain growth, thermal distortion, bead geometry, and local mechanical properties. The model therefore ranked heat input highly because it condenses several physically coupled aspects of the welding process into a variable strongly connected with both geometry and metallurgy.

The association between high travel speed and lack of fusion or incomplete penetration is also physically reasonable. Increased travel speed reduces the residence time of the heat source at a given location and can limit energy transfer into the joint faces and root region. The resulting molten zone may be insufficient to establish complete metallurgical bonding even when current and voltage individually remain within nominal operating ranges. The SHAP interaction effects observed between travel speed and heat input reinforce this interpretation because the influence of speed changed according to the corresponding thermal conditions. This interaction illustrates why parameter-by-parameter troubleshooting can be inadequate: the same travel speed may be acceptable under one energy regime and problematic under another.

The association between reduced shielding-gas flow and porosity can be explained through the protective function of the shielding atmosphere around the arc and molten metal. Stable shielding restricts detrimental interaction between the weld pool and the surrounding atmosphere, whereas inadequate coverage can facilitate contamination and gas absorption or interfere with stable weld-pool behavior. Entrapped gas that cannot escape before solidification may remain as rounded or clustered pores. Experimental research using other welding processes has also demonstrated that changes in shielding-gas conditions modify pore formation and weld stability, although the detailed fluid and arc mechanisms depend on the specific technology employed. The strong local SHAP contribution observed for shielding-gas flow therefore has a plausible physical basis rather than appearing as an isolated statistical artifact.

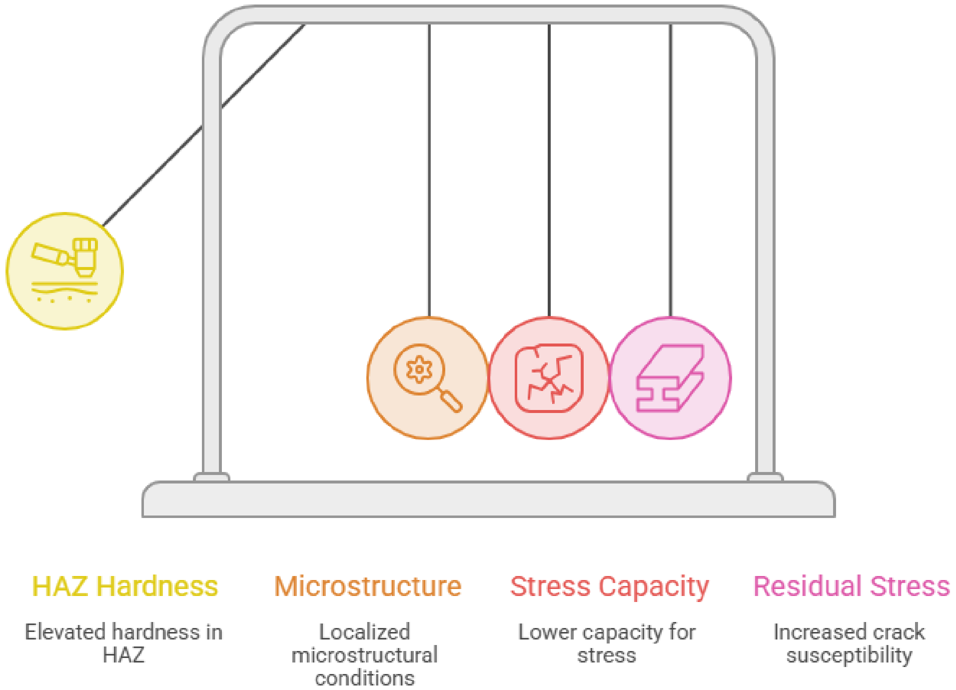

Figure 2. HAZ Hardness Impacts Marine Steel Cracking

The association between elevated HAZ hardness and cracking reflects the material sensitivity of marine steel to the welding thermal cycle. Rapid heating followed by rapid cooling can produce localized microstructural conditions with higher hardness and lower capacity for accommodating stress, depending on steel composition, peak temperature, cooling rate, constraint, and hydrogen availability. Residual stresses created during nonuniform heating and contraction may further increase crack susceptibility. Previous AH36 studies have demonstrated meaningful links among heat input, microstructure, hardness, residual stress, hydrogen transport, and embrittlement behavior. The present model appears to capture part of this coupled response, but the absence of direct diffusible-hydrogen and residual-stress measurements means that the cracking pathway should remain described as physically supported rather than exhaustively proven.

The immediate next step is to validate the framework using a larger and more heterogeneous experimental population. The current results were obtained from a controlled set of AH36 specimens within a defined GMAW parameter window, which is appropriate for mechanism-oriented development but insufficient for claiming universal applicability. Future validation should include different plate thicknesses, groove geometries, welding positions, consumable batches, operators, and environmental conditions. External datasets from separate fabrication facilities would provide an especially demanding test because model performance obtained from random partitions of one experimental campaign can overestimate the generalization achievable under real production variability.

The framework should subsequently incorporate additional physical measurements that can strengthen causal interpretation. Thermal histories obtained from thermocouples or infrared sensing could provide direct information on cooling behavior, while residual-stress measurements could clarify cracking mechanisms. Diffusible-hydrogen analysis would be necessary before attributing specific cracks to hydrogen-assisted mechanisms. EBSD could provide quantitative grain-orientation and boundary information, while chemical or inclusion analysis could improve interpretation of crack initiation and porosity-related phenomena. Recent investigations of EH36 welding have demonstrated the value of combining optical microscopy, SEM, EBSD, hardness, and mechanical characterization when explaining the effects of thermal input on marine-steel performance.

The computational architecture should also progress from post hoc explainability toward stronger causal and physics-informed reasoning. SHAP can continue to serve as a useful diagnostic tool, but future models could incorporate causal graphs, physically constrained machine learning, process knowledge, counterfactual analysis, or hybrid mechanistic–data-driven architectures. Such approaches would allow researchers to distinguish questions such as "Which variable influenced the model prediction?" from the more demanding question "Which intervention would most likely prevent this defect?" Recent causal-XAI research in manufacturing has begun addressing precisely this limitation by integrating data-driven prediction with causal structure learning rather than relying entirely on statistical feature attribution.

The longer-term development of this research should focus on an explainable decision-support system capable of translating welding data into preventive engineering actions. Such a system could continuously collect process signals, estimate defect risk, identify the most probable process–microstructure pathway, quantify explanation confidence, and recommend which parameter or physical condition should be inspected before corrective action is taken. Human verification should remain part of the diagnostic cycle, particularly for high-severity defects such as cracking. The present findings provide an initial foundation for that transition by demonstrating that machine-learning predictions become substantially more informative when process parameters, metallurgical measurements, defect morphology, and explainability are interpreted together. The broader contribution is therefore not merely a more accurate welding classifier but a pathway toward physically grounded, transparent, and accountable artificial intelligence for marine manufacturing.

CONCLUSION

The most important finding of this study is that marine welding defects can be interpreted more effectively when process parameters, microstructural responses, and Explainable AI outputs are analyzed within a single diagnostic framework. Porosity was mainly associated with inadequate shielding-gas conditions, whereas lack of fusion and incomplete penetration were more strongly related to insufficient effective heat input and increased travel speed. Cracking exhibited a distinct metallurgical pathway characterized by elevated heat-affected-zone hardness and unfavorable thermal conditions. The XGBoost model achieved strong predictive performance, while SHAP analysis revealed that heat input, shielding-gas flow rate, travel speed, and HAZ microhardness were the most influential variables. The convergence between process records, model explanations, defect morphology, and microstructural evidence demonstrates that welding defects are better understood as outcomes of specific process–structure–failure pathways rather than as isolated quality abnormalities.

The principal contribution of this research lies in the development of a mechanism-informed Explainable AI approach that extends conventional defect prediction toward interpretable root-cause diagnosis. The conceptual contribution is the process–microstructure–defect framework, which positions microstructural failure modes as an explanatory layer between manufacturing parameters and observed welding defects. The methodological contribution is the integration of controlled welding experiments, non-destructive testing, metallographic characterization, microhardness measurements, machine-learning classification, and SHAP-based interpretation within a unified validation procedure. This approach distinguishes predictive feature attribution from physically supported diagnostic evidence and therefore provides a more defensible basis for applying artificial intelligence in safety-critical marine manufacturing. The framework also offers practical value by enabling engineers to identify probable sources of weld-quality deterioration and to prioritize corrective actions based on technically interpretable evidence.

The main limitation of this study is that the framework was developed using a controlled experimental dataset involving AH36 marine steel, a defined GMAW process window, and a limited number of welding parameters and microstructural indicators. The findings should therefore not be generalized directly to different marine alloys, welding processes, joint geometries, thicknesses, environmental conditions, or production settings without further validation. SHAP explanations also represent model-based associations rather than independent proof of causality, while potentially important variables such as diffusible hydrogen, residual stress, real-time thermal history, arc instability, and detailed crystallographic characteristics were not comprehensively measured. Future research should validate the framework across multiple marine-grade steels and industrial welding environments, incorporate real-time sensing and advanced microstructural characterization, and integrate causal inference, physics-informed machine learning, and counterfactual analysis. Such developments could transform the proposed framework into a reliable decision-support system for real-time defect prevention, transparent process optimization, and trustworthy AI-assisted welding quality assurance.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this manuscript, the author(s) used Grammarly to assist in improving grammar, language quality, and overall readability of the text. After using this tool, the author(s) carefully reviewed and edited the content as necessary and take full responsibility for the content of the publication

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

DECLARATION OF COMPETING INTEREST

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

REFERENCES

Adeleke, O., & Jen, T.-C. (2025). Explainable AI and machine learning-based analysis of municipal solid waste generation rate: A South African case study. *Waste Management*, *206*, 115036. https://doi.org/10.1016/j.wasman.2025.115036

Alkhamis, S. A., Alhalangy, A., Eltayeb, G. E. A., & Abouelkheir, E. (2026). Fed-CGIDS-UAV: Federated Causal Graph Learning for Cross-Domain Intrusion Detection in Cyber-Physical Drone Networks. *Symmetry*, *18*(8), 1292. https://doi.org/10.3390/sym18081292

Ammar, S., Ben Fraj, B., Hentati, H., Saouab, A., Ben Amar, M., & Haddar, M. (2024). Mechanical performances of printed carbon fiber-reinforced PLA and PETG composites. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, *238*(8), 1488–1499. https://doi.org/10.1177/14644207231225761

Awad, I. G. A. (2026). Failure Analysis and Simultaneous Optimization of Flexural Strength and Elastic Modulus in Friction Stir Welded PP / HDPE – TiO_2 Nanocomposite Joints. *Polymers for Advanced Technologies*, *37*(7), e70660. https://doi.org/10.1002/pat.70660

Das, S., & Ganesh Narayanan, R. (2025). Lap shear fracture prediction of friction stir spot welds between AA6063-T6 and CRCA/IS-513 sheets. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, 09544089251335077. https://doi.org/10.1177/09544089251335077

Gangwal, A., & Lavecchia, A. (2026). IMPACT Framework: Establishing Global Standards for Artificial Intelligence Implementation, Methodology, and Translation in Drug Discovery. *WIREs Computational Molecular Science*, *16*(2), e70072. https://doi.org/10.1002/wcms.70072

Ghazi, Y., Tabaa, M., Ennaji, M., & Zaz, G. (2025). An Explainable Markov Chain–Machine Learning Sequential-Aware Anomaly Detection Framework for Industrial IoT Systems Based on OPC UA. *Sensors*, *25*(19), 6122. https://doi.org/10.3390/s25196122

Huang, Z., Jiang, M., Cao, S., & Yilmaz, E. (2026). Engineering strength through microstructure: Gradation and content effects in cemented high-sulfur tailings-crushed rock backfill. *Results in Engineering*, *32*, 111841. https://doi.org/10.1016/j.rineng.2026.111841

Kewalramani, M., Mahmoodzadeh, A., Hassan, R., El Ouni, M. H., Alqahtani, A., Alanazi, A., Alsubai, S., & Palani, S. (2025). Predicting critical crack propagation length in sustainable additive-enhanced concrete using explainable machine learning. *Scientific Reports*, *16*(1), 26. https://doi.org/10.1038/s41598-025-31900-7

Khadka, P., Varglund, S., Akinwamide, S., & Vilaça, P. (2024). Characterization of friction stir-based linear continuous joining of aluminium alloy to structural polymer. *Welding in the World*, *68*(1), 71–90. https://doi.org/10.1007/s40194-023-01622-3

Khorshidi, M., Nejatipour, P., Teimortashlu, A., Oliveto, G., & Afaridegan, E. (2026). Explainable AI-based optimization framework for sustainable fly ash concrete design. *Scientific Reports*, *16*(1), 22397. https://doi.org/10.1038/s41598-026-59294-0

Li, Y., Cheng, K., Wang, L., & Yang, W. (2026). Fault Analysis and Optimization of PT100 Temperature Sensor for Marine Exhaust Duct. In J. P. T. Mo (Ed.), *Advances in Transdisciplinary Engineering*. IOS Press. https://doi.org/10.3233/ATDE260562

Mallipudi, S. R., & Marampudi, R. (2025). Mechanical and Metallurgical Characterization of Friction Stir Welded Al–Mg–Mn Alloy with Minor Additions of Sc and Zr.

Transactions of the Indian Institute of Metals, *78*(6), 148. https://doi.org/10.1007/s12666-025-03624-9

Msomi, V., Mabuwa, S., Ngonda, T., Mehdi, H., & Saxena, K. K. (2025). Microstructure and mechanical properties of tungsten inert gas welded dissimilar aluminum alloys joint subjected to normal multi-pass friction stir process. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, *239*(4), 1925–1935. https://doi.org/10.1177/09544089231207424

Naik, R., & Majhi, B. (2025). Explainable AI reverse verification approach for monthly rainfall prediction in Chhattisgarh, India. *Theoretical and Applied Climatology*, *156*(7), 412. https://doi.org/10.1007/s00704-025-05645-2

Onyelowe, K. C., Hanandeh, S., Kamchoom, V., Ebid, A. M., Imran, H., Duque Vaca, M. A., Herrera Morales, G. C., Ulloa, N., & Arunachalam, K. P. (2025). Data-driven framework for prediction of mechanical properties of waste glass aggregates concrete. *Scientific Reports*, *15*(1), 20902. https://doi.org/10.1038/s41598-025-05229-0

Özüpak, Y. (2026). A hybrid physics-guided ensemble learning framework for predicting aviation co_2 emissions with explainable artificial intelligence. *Theoretical and Applied Climatology*, *157*(7), 426. https://doi.org/10.1007/s00704-026-06356-y

Patel, R., Tiwari, S. K., Rakesh, D. K., Kailasanathan, N., & Selavarajan, S. (2025). An automated framework for traffic noise level analysis using explainable artificial intelligence techniques. *Scientific Reports*, *15*(1), 36657. https://doi.org/10.1038/s41598-025-20377-z

Peng, J., Wang, S., Ding, W., Huang, X., Slamani, M., & Zhao, B. (2026). Machinability and surface formation mechanisms of SiCp/Al composites at different cutting speeds. *Precision Engineering*, *101*, 310–322. https://doi.org/10.1016/j.precisioneng.2026.05.001

Piyar S, N., & S, D. B. (2026). An explainable spatio-temporal deep learning framework for crop yield prediction and recommendation. *Scientific Reports*, *16*(1), 24662. https://doi.org/10.1038/s41598-026-54439-7

Rajagopal, V., Rajendran, D. K., & Muthukumaran, S. (2024). Multi-objective optimization and characterization of AA2219-T87 and AA6061-T6 dissimilar friction stir welded joints under the effect of copper backing using AUGMECON. *Sādhanā*, *49*(1), 32. https://doi.org/10.1007/s12046-023-02360-4

Saroj, S., Sahoo, P., & Kore, S. D. (2026). Influence of induction preheating and post-weld heat treatment on the mechanical, microstructural and corrosion properties of friction stir welded marine-grade aluminium alloy. *Welding International*, *40*(3), 211–228. https://doi.org/10.1080/09507116.2025.2525155

Shawon, S. M., Zaman, M., Maniam, S., Al Qwaid, M., Sarker, M. T., Kheng, T. Y., & Wong, H. Y. (2026). Explainable Quantile CNN-LSTM model for uncertainty-aware multilayer soil moisture prediction in tropical cocoa plantations. *Scientific Reports*, *16*(1), 16994. https://doi.org/10.1038/s41598-026-48517-z

Shen, L., Shojaeikhah, S., & Soliman, M. (2025). Experimental analysis of crack propagation in stiffened box girders under variable-amplitude loading. *Journal of Constructional Steel Research*, *235*, 109812. https://doi.org/10.1016/j.jcsr.2025.109812

Shivaprasad, S., Chadaga, K., Dias, C. C., Sampathila, N., & Prabhu, S. (2024a). An interpretable schizophrenia diagnosis framework using machine learning and explainable artificial intelligence. *Systems Science & Control Engineering*, *12*(1), 2364033. https://doi.org/10.1080/21642583.2024.2364033

Shivaprasad, S., Chadaga, K., Dias, C. C., Sampathila, N., & Prabhu, S. (2024b). An interpretable schizophrenia diagnosis framework using machine learning and explainable artificial intelligence. *Systems Science & Control Engineering*, *12*(1), 2364033. https://doi.org/10.1080/21642583.2024.2364033

Singh, A., Sharma, S. K., & Batish, A. (2026). Achieving high-performance SS304 joints through dynamic recrystallization in friction stir welding. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, *240*(9), 1567–1580. https://doi.org/10.1177/14644207251384748

Singh, A., Singh, S., Kamboj, R. K., & Bala, A. (2025). Explainable OptiCNN-SLSTM hybrid model for enhanced lithium-ion battery state of health prediction. *Scientific Reports*, *15*(1), 44364. https://doi.org/10.1038/s41598-025-28091-6

Varma, S. O., Jadhav, P. P., Vishwakarma, A. L., Sonawane, M. R., & Chaudhari, A. (2026). Explainable neural network models for rapid prediction of dielectric constant and acid value in edible oils using FTIR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *363*, 128376. https://doi.org/10.1016/j.saa.2026.128376

Wang, B., Zhao, S., Ren, Q.-X., & Xie, R.-Q. (2025). Axial compressive behavior of circular hollow steel tube-reinforced UHTCC column. *Structures*, *71*, 107943. https://doi.org/10.1016/j.istruc.2024.107943

Wei, F., & Frederick, A. (2025). Comprehensive Die Strengths Comparisions for Glass-Core Advanced Packaging Substrates using Different Singulation Methods—Dicing-Induced SeWaRe Failures Revisited Ten Years Later-. *2025 IEEE 75th Electronic Components and Technology Conference (ECTC)*, 146–153. https://doi.org/10.1109/ECTC51687.2025.00031

Wu, D., Wang, C., Liu, H., Zhang, Z., Liang, J., & Liu, Y. (2026). Correlation of breakage characteristics between single-particle and uniformly-graded volcanic soil: Insights from compression tests. *Powder Technology*, *468*, 121611. https://doi.org/10.1016/j.powtec.2025.121611

Xu, Y., Saleem, M. F., Mingyu, G., Xie, W. L., Zhang, S.-H., Zhu, H. L., Khina, B. B., & Pokrovsky, A. I. (2026). Formability and failure mechanisms of ultra-thin TA2 titanium sheets in microchannel hydroforming: Effect of thickness-to-grain-size ratio. *Thin-Walled Structures*, *218*, 114042. https://doi.org/10.1016/j.tws.2025.114042

Zhang, Y., He, C., Yu, Q., Li, X., Wang, X., Zhang, Y., Wang, J., Jiang, C., Jia, Y., Zhang, X.-C., Sun, B., Ritchie, R. O., & Tu, S.-T. (2024). Nacre-like surface nanolaminates enhance fatigue resistance of pure titanium. *Nature Communications*, *15*(1), 6917. https://doi.org/10.1038/s41467-024-51423-5

Zhou, B., Li, Y., Wu, Y., Zhang, E., Chen, X., Tao, Z., & Xing, N. (2026). Investigation of the failure characteristics of goaf roof in a metal mine using the acoustic emission technique. *Nondestructive Testing and Evaluation*, 1–30. https://doi.org/10.1080/10589759.2026.2627620

Zhou, Q., Liu, H., Tu, W., Xue, T., & Xue, H. (2026). Multi-scale decoupling analysis of geometric imperfection effects on stress concentration in butt-welded joints of IMO type B tanks. *Ocean Engineering*, *355*, 125078. https://doi.org/10.1016/j.oceaneng.2026.125078