



Development of an AI-based acoustic disturbance-detection method for robotic arc welding processes

Houichi Kitano, Masaki Kobayashi & Masahiko Demura

To cite this article: Houichi Kitano, Masaki Kobayashi & Masahiko Demura (2026) Development of an AI-based acoustic disturbance-detection method for robotic arc welding processes, Science and Technology of Advanced Materials: Methods, 6:1, 2688746, DOI: [10.1080/27660400.2026.2688746](https://doi.org/10.1080/27660400.2026.2688746)

To link to this article: <https://doi.org/10.1080/27660400.2026.2688746>



© 2026 The Author(s). Published by National Institute for Materials Science in partnership with Taylor & Francis Group



Published online: 23 Jun 2026.



Submit your article to this journal [↗](#)



Article views: 13



View related articles [↗](#)



View Crossmark data [↗](#)

Development of an AI-based acoustic disturbance-detection method for robotic arc welding processes

Houichi Kitano , Masaki Kobayashi and Masahiko Demura 

National Institute for Materials Science, Tsukuba, Ibaraki, Japan

ABSTRACT

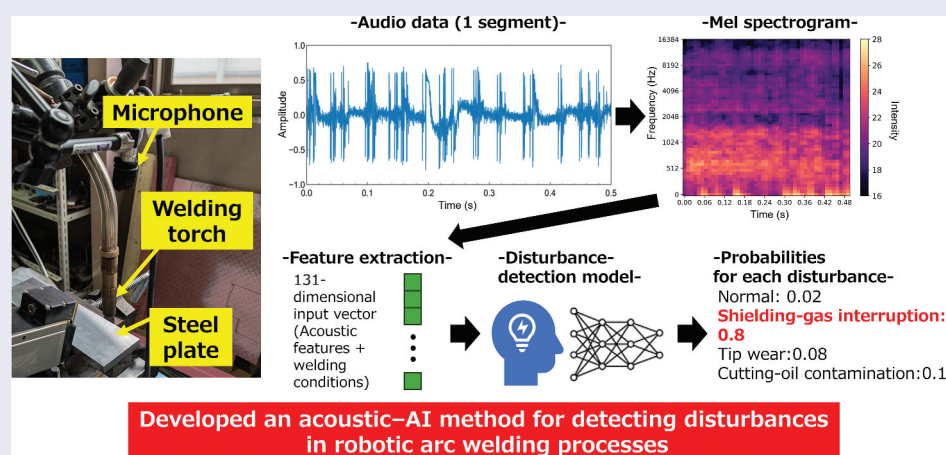
Disturbances generated during arc welding processes can be detrimental to the quality of welded structures, and existing automated disturbance-detection methods lack the capability for real-time deployment. This paper proposes an AI-based disturbance-detection framework for gas metal arc welding using microphone-recorded acoustic signals. Representative disturbances – shielding-gas interruption, tip wear, and cutting-oil contamination – were experimentally reproduced and acoustically recorded under three welding conditions. Acoustic features were then extracted from the Mel-spectrograms of the recorded welding sounds and combined with welding parameters (current, voltage, and travel speed) to train a multilayer perceptron classifier capable of identifying both the occurrence and type of disturbance. The trained model achieved an overall accuracy of 81.5% and a macro-F1 score of 82.9%, demonstrating reliable generalization performance. Time-series evaluation indicated that the model could maintain a stable classification performance from the early stage of welding and immediately after the onset of disturbance. Furthermore, a SHAP (Shapley Additive exPlanations) analysis revealed that the decision criteria of the model were physically interpretable: high-frequency attenuation was dominant in shielding-gas interruption, while low-frequency vibration components were characteristic of tip wear. Both the spectral intensity and its variance were identified as key features for accurate disturbance classification. The proposed approach provides a low-cost, noncontact, and real-time monitoring solution that can be easily integrated into robotic welding systems and adapted to various industrial environments, thereby contributing to the realization of autonomous and explainable in-process quality assurance in smart manufacturing.

ARTICLE HISTORY

Received 19 November 2025
Revised 10 March 2026
Accepted 9 June 2026

KEYWORDS

Acoustic sensing; in-process monitoring; gas metal arc welding; disturbance detection; machine learning; explainable AI; smart manufacturing



IMPACT STATEMENT

This study demonstrates a low-cost acoustic sensing and machine-learning framework enabling real-time disturbance detection in robotic arc welding, offering a practical foundation for autonomous in-process quality assurance.

1. Introduction

Arc welding is one of the most widely used joining technologies for metallic materials and is indispensable in a wide range of industrial fields including automobiles, construction, and energy plants. In recent

years, robotic arc welding using multi-joint industrial robots has been increasingly adopted to improve productivity and reproducibility in mass-production lines. Welding parameters, such as current, voltage, and travel speed, are generally optimized in

preliminary trials, allowing stable, high-quality welding during normal operations. However, in actual production, unexpected disturbances, such as a reduction in the shielding gas flow, wear of consumable parts, or contamination of the workpiece surface, may occur. These disturbances destabilize the arc process and can cause welding defects, such as porosity or lack of fusion, thereby reducing the reliability of the welded structures. Conventionally, such defects are detected through post-process visual or nondestructive inspection. However, these inspections require significant time and cost, and rework following defect detection markedly reduces manufacturing efficiency. Consequently, in-process quality assurance, in which anomalies are detected in real-time during welding, has become increasingly important.

Various in-process monitoring techniques have been investigated, including the evaluation of arc stability using current and voltage waveforms [1–5] and the observation of molten pool behavior using charge-coupled device (CCD) cameras or laser profiling [5–8]. Recent review studies have also summarized the development of sensing technologies and data-driven monitoring approaches for arc welding processes [9,10]. However, electrical signal monitoring is sensitive to factors, such as power-supply noise or torch-position variations, which do not affect weld quality, whereas optical measurements are expensive and require protective optics, making them difficult to apply in production environments with heavy fumes and spatter. Therefore, the development of a simpler, more versatile, and field-applicable disturbance-detection method is required.

To address this issue, this study developed a disturbance-detection method using a microphone that is inexpensive, easy to maintain, and does not require optical protection. Acoustic signals recorded by a microphone can reflect the welding process state in a noncontact and real-time manner. In other engineering fields, acoustic-based anomaly-detection and equipment-diagnosis techniques have been actively studied [11–17]. For example, Lee et al. [11] represented segmented acoustic recordings by aggregated statistical descriptors for equipment fault classification, and Ding et al. [13] extracted wavelet-based features and summarized them into statistical vectors before applying a convolutional neural network. In these and related studies, acoustic recordings of various durations – ranging from tens of milliseconds to several tens of seconds – have been analyzed depending on the target application and detection granularity. In arc welding, skilled welders intuitively judge the arc stability and molten pool behavior based on changes in the discharge sound. Hence, a quantitative analysis of welding sounds has great potential for detecting the onset of disturbances. However, welding sound signals contain broadband noise generated by discharge and

metal evaporation. They also exhibit highly nonstationary and nonlinear behaviors, as the acoustic characteristics vary dynamically with arc instability and molten pool fluctuations. Therefore, a simple frequency analysis or power spectrum evaluation is insufficient for accurately detecting or classifying such disturbances. Therefore, more advanced feature extraction and data-driven analytical approaches are required for complex signals.

In this study, a new AI-based disturbance-detection model was developed that utilizes acoustic features derived from welding sound signals transformed into the time – frequency domain using a Mel-spectrogram. In this model, the extracted features are inputted to a multilayer perceptron (MLP) for a multiclass classification of welding states encompassing both normal and disturbance conditions. A Mel spectrogram is a time – frequency representation of acoustic signals that helps visualize the signal intensity over time and frequency [18,19]. The linear frequency axis is transformed into a logarithmic Mel scale based on human auditory perception, and the neighboring frequency components are averaged within perceptually equivalent Mel bands. This process compresses the acoustic representation while reducing local noise and small fluctuations, making it effective for analyzing nonstationary broadband acoustic signals. The acoustic features derived from the Mel-spectrogram represent the statistical patterns of each frequency band and, when combined with welding parameters (current, voltage, and travel speed), allow the MLP to learn disturbance-specific spectral variations. Because a Mel spectrogram represents two-dimensional time – frequency information, convolutional neural networks (CNNs) can be used for sound classification tasks. However, in the present study the Mel-spectrogram information was summarized into statistical descriptors (mean and variance for each frequency band) before classification. Preliminary evaluations showed that this statistical representation combined with a MLP provided more stable classification performance for the present dataset.

Furthermore, to clarify the decision process of the trained model, a SHAP (SHapley Additive exPlanations) analysis [20] was applied to visualize the contribution of each frequency band and feature to the disturbance classification. SHAP is an explainable AI framework based on cooperative game theory that can quantify the contribution of each input feature to the prediction outcome, enabling a quantitative and intuitive interpretation of model behavior.

The study objectives were: (1) to develop a new framework for real-time disturbance detection in arc welding processes using low-cost and low-maintenance acoustic sensing and (2) to analyze and visualize the reliability and decision-making behavior

of an AI-based disturbance classification model in relation to the physical phenomena encountered in the welding process. This study experimentally demonstrates the effectiveness of an AI-driven, acoustic-based, in-process monitoring method and provides a technological basis for smart manufacturing and material digital twins in the context of material integration [21–24].

2. Methods

2.1. Welding setup and disturbance conditions

Bead-on-plate welding tests were conducted using an industrial multi-joint robot (DAIHEN Corporation, Japan) to acquire acoustic signals during arc welding. Bead-on-plate welding is a fundamental test in which a single bead is deposited on a flat plate without groove preparation or joint assembly and is widely used for evaluating welding phenomena and comparing process parameters.

Figure 1 shows an overview of the experimental setup. The test plate was fixed to a jig, and the welding torch mounted on the robot traversed along the plate surface. The details of the microphone and the CCD camera are described in Section 2.2. The base material was a rolled structural steel plate, SM490 (JIS G 3106 SM490A equivalent), and the filler wire was a 50 kg-grade solid wire (JIS Z 3312 YGW11 equivalent, 1.2 mm diameter). Gas metal arc welding (GMAW) was performed using an Ar–20% CO₂ shielding gas, and three conditions were set for the welding current, voltage, and travel speed as summarized in Table 1. The weld length was 200 mm.

Table 2 presents the reproduced disturbance conditions and implementation methods. Shielding-gas interruption was simulated by manually closing the pressure-regulating valve during welding to reproduce a severe shielding-gas supply failure. Tip wear was

emulated by enlarging the torch tip orifice using electrical discharge machining, as illustrated in Figure 2. Three levels of enlargement—0.25, 0.5, and 1.0 mm – were tested; the 0.25 mm case was considered slight wear and treated as a non-disturbance (normal) condition in subsequent analyses, as previous studies have reported that the effect of tip wear on process stability becomes significant when the orifice diameter exceeds roughly 1.4 times the wire diameter [25]. Cutting-oil contamination was simulated by locally applying oil onto the plate surface. These disturbances were selected because they occur frequently in actual production environments. All the combinations of the three welding conditions and five disturbance types, along with the normal condition (18 cases in total), were tested.

2.2. Acoustic measurement and data acquisition

A high-sensitivity microphone (Xiris Automation Inc, Canada) operating at frequencies ranging from 50 Hz to 20 kHz and with a dynamic range of +120 dB was used to record the welding sounds. The sampling frequency was 48 kHz. As shown in Figure 1, The microphone was mounted on a fixture attached to the welding torch and moved together with the torch during welding. Consequently, the relative distance between the microphone tip and the plate surface (approximately 200 mm) was maintained constant throughout the experiments. The arc behavior was simultaneously observed using a CCD camera (Xiris Automation Inc., Canada). The CCD images were not used as inputs to the AI-based disturbance-detection system but were utilized to determine the welding start and end times as well as the start time of each disturbance. The start time of the shielding-gas interruption was defined as the moment when a visible change in arc behavior was observed after the shielding gas supply was stopped. Specifically, this included changes

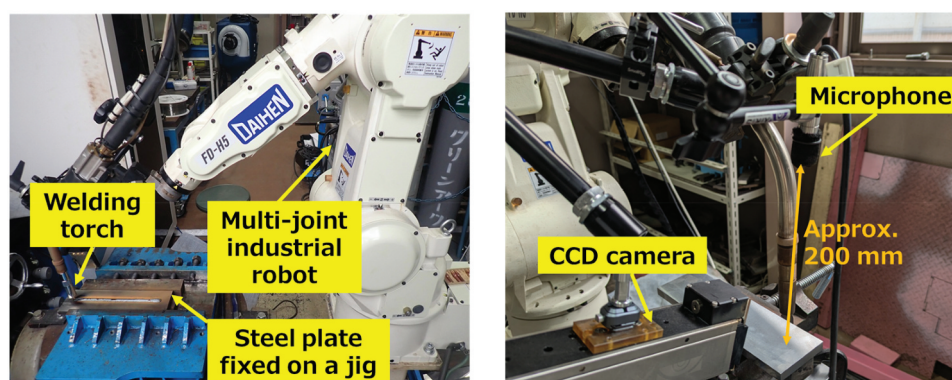


Figure 1. Experimental setup for bead-on-plate welding and acoustic signal acquisition. The steel plate was fixed on a jig, and the welding torch mounted on a multi-joint industrial robot traversed along the plate surface. The microphone and the CCD camera were mounted on fixtures attached to the welding torch and moved together with the torch during welding. The distance between the microphone tip and the plate surface was approximately 200 mm. The CCD camera was installed inside a protective shielding enclosure to prevent damage from arc light, spatter, and fumes during welding.

Table 1. Welding conditions for gas metal arc welding (GMAW). The welding current values represent average steady-state currents under constant-voltage operation. The filler wire was a 50 kg-grade solid wire (JIS Z 3312 YGW11 equivalent, 1.2 mm diameter), the shielding gas composition was Ar–20% CO₂, and the contact tip-to-work distance (CTWD) was 13 mm.

Condition No.	Welding current (A)	Welding voltage (V)	Welding speed (cm/min)
1	200	23	30
2	250	26	45
3	300	28	60

in arc stability, increased fluctuation of arc brightness, and the occurrence of irregular spatter behavior. The end time of the shielding-gas interruption was identical to that of the welding. For cutting-oil contamination, the start time was defined as the moment when the torch center reached the beginning of the contaminated region on the plate. The end time for the

cutting-oil contamination condition was identical to the welding end time. For the tip-wear condition, because the worn tip was mounted before welding, the start and end times of the disturbance coincided with those of the welding.

The audio data clipping ranges were defined for each condition as follows: For normal welding and tip wear, the recorded sound was clipped between the start and end times of the welding. Therefore, these datasets include the initial arc ignition period. For shielding-gas interruption and cutting-oil contamination, the range extended from the disturbance onset (0 s) to the end of welding. In these cases, the analyzed data do not include the initial arc ignition period because the disturbance occurs after stable welding has already been established. After clipping, each clipped data point was divided into 0.5 s segments, and every segment was used as one analysis sample.

Table 2. Reproduced disturbance conditions and their implementation methods.

Disturbance type	Reproduction method	Description/Remarks
Shielding-gas interruption	Gas regulator manually closed during welding to completely stop the shielding-gas flow	Simulated a total loss of shielding atmosphere, representing gas-supply failure
Tip wear (0.25 mm)	Tip orifice enlarged by 0.25 mm using electrical-discharge machining (EDM)	Mild wear condition. This case was treated as a non-disturbance (normal) condition in the subsequent analysis.
Tip wear (0.5 mm)	Same as above (0.5 mm enlargement)	Moderate wear condition.
Tip wear (1.0 mm)	Same as above (1.0 mm enlargement)	Severe wear condition
Cutting-oil contamination	Applied cutting oil locally on the plate surface	Simulated residual oil film on base metal

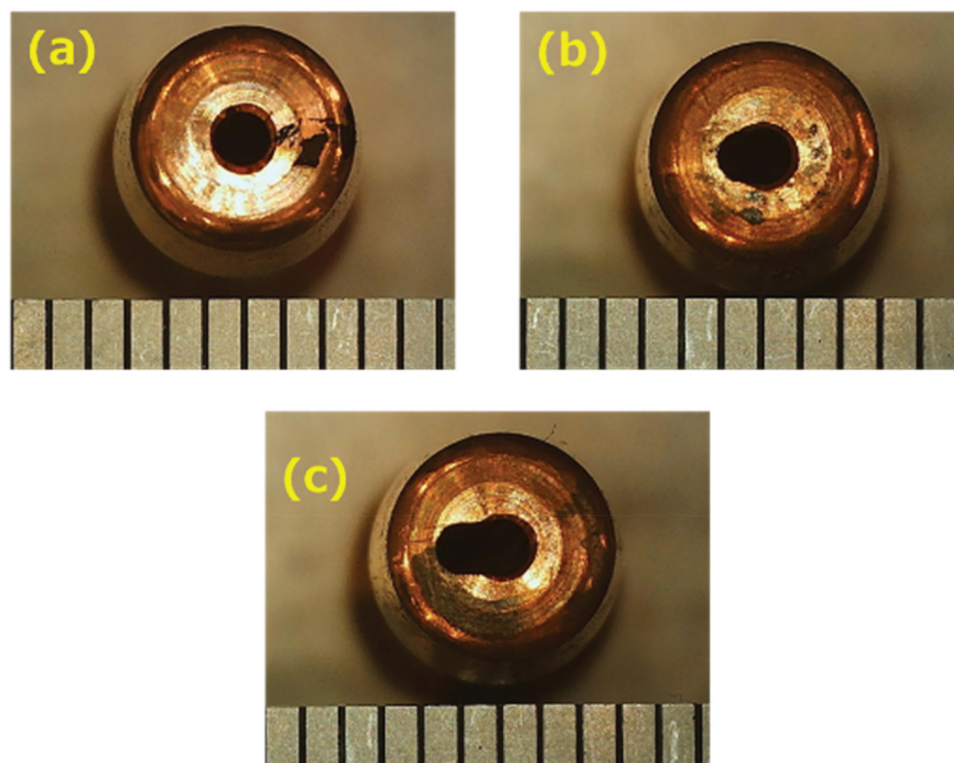


Figure 2. Representative appearances of the emulated welding-tip wear conditions. (a) Tip orifice enlarged by 0.25 mm, (b) 0.5 mm, and (c) 1.0 mm by electrical-discharge machining.

2.3. Feature extraction and preprocessing

The recorded sound data were divided into 0.5 s segments and converted into Mel-spectrograms to extract acoustic features. As illustrated in Figure 3, each segment was transformed into a Mel-spectrogram with 64 mel bands, a window size of 2048 samples (≈ 42.7 ms at 48 kHz), and a hop length of 512 samples (≈ 10.7 ms at 48 kHz). The resulting time – frequency map comprised 64 frequency components and 47 timeframes. For each frequency component, the mean and variance along the time axis (47 frames) were calculated and used as the acoustic features. Consequently, 128-dimensional acoustic features (64 means +64 variances) were obtained from one audio segment. This statistical representation provides a compact description of the spectral distribution of each frequency band while reducing noise and dimensionality compared with using the full time – frequency map. The corresponding welding parameters (current, voltage, and travel speed) were appended to form a 131-dimensional input vector. In this study, these parameters correspond to the set values of the welding system rather than instantaneous measurements. Mel spectrograms were computed using the Librosa Python library [26]. In the Mel-spectrogram computation using Librosa, centered framing with zero-padding was applied. Consequently, the signal was padded by half of the window size on both sides before the short-time Fourier transform was computed, resulting in 47 timeframes for each 0.5 s segment. The maximum frequency of the Mel filter bank was set to 16 kHz

($f_{\max} = 16,000$ Hz), although the sampling frequency was 48 kHz (Nyquist frequency: 24 kHz). Frequencies above 16 kHz were not considered because the dominant acoustic energy of the welding arc was observed within this range. Therefore, the upper limit of the frequency axis in Figure 3 reflects the specified Mel filter range rather than the Nyquist frequency.

2.4. Machine learning framework

An MLP classifier was developed to identify various disturbance conditions, including normal conditions. Figure 4 shows the structure of the MLP used in this study. The input comprised the standardized 131-dimensional vector constructed as described in Section 2.3, comprising 128 acoustic feature dimensions and three welding parameter dimensions. The vector was passed through two fully connected hidden layers with 2048 units, and the output layer produced four logits (unnormalized scores) corresponding to four classes: normal (no disturbance), shielding-gas interruption, tip wear, and cutting-oil contamination. After the Softmax transformation, the probabilities for each class were obtained, and the class with the highest probability was selected as the final prediction. The sum of all class probabilities is always one. The calculation method for the parameters used for the input vector standardization (mean and variance) is described later. The training procedure for the MLP is described below:

The dataset containing the input vectors and corresponding class labels was first divided and stratified by disturbance labels into an initial training subset and

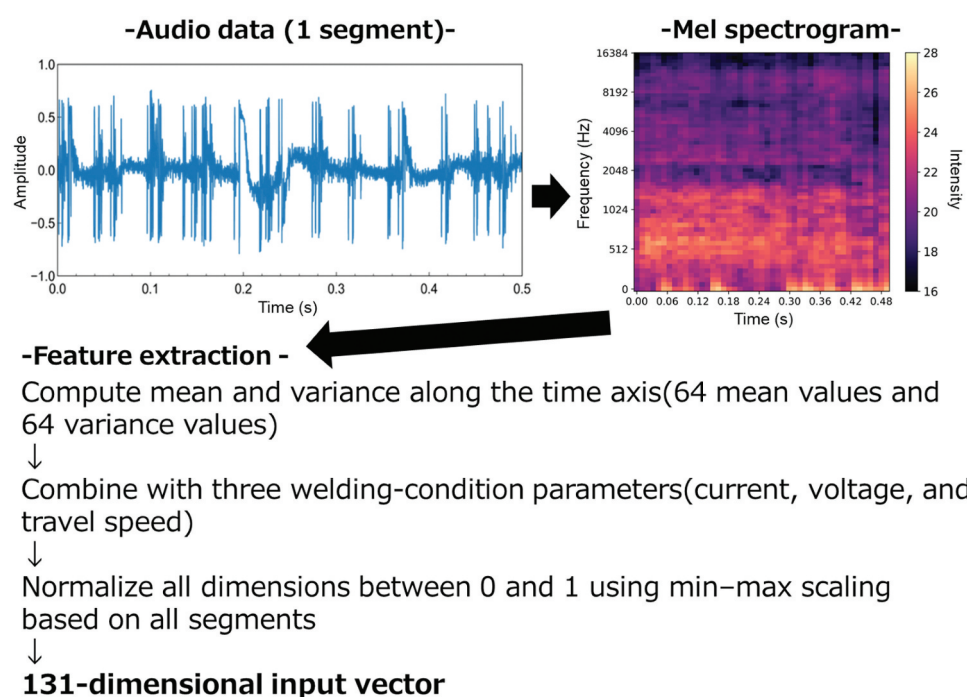


Figure 3. Procedure for acoustic feature extraction and input vector construction.

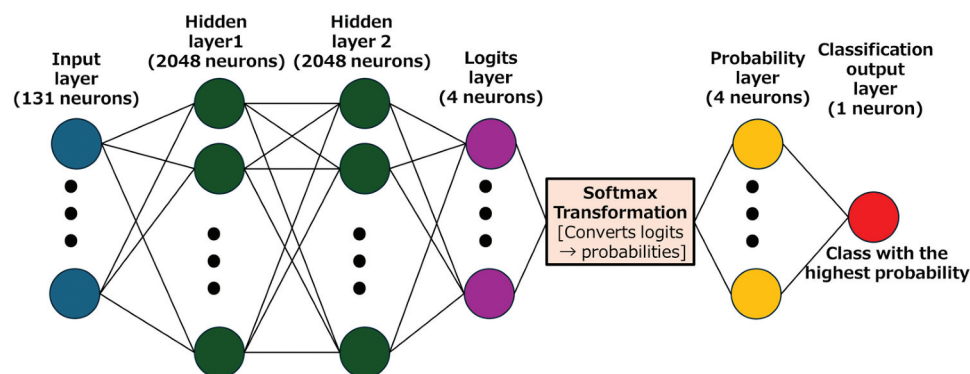


Figure 4. Architecture of a multilayer perceptron (MLP) model for disturbance classification. A normalized input vector comprising 131 dimensions (128 acoustic features +3 welding-condition parameters) is processed through five hidden layers with ReLU activation. The output layer produces seven logits corresponding to each disturbance category, including the normal (no-disturbance) condition, which are transformed into class-probability values using a softmax function. The class with the highest probability is selected as the final output of the model.

a test subset at a ratio of 7:3 to ensure that the class distribution was preserved. To mitigate class imbalance, a synthetic minority oversampling technique (SMOTE) [27] was applied to the initial training subset. An implementation in the imbalanced-learn (imblearn) Python library [28] was used, with the number of nearest neighbors set to 30. Before oversampling, the numbers of samples for the four classes – normal welding, shielding-gas interruption, tip wear, and cutting-oil contamination – were 238, 238, 101, and 113, respectively. SMOTE was used to generate synthetic samples for the minority classes (tip wear and cutting-oil contamination) such that the number of samples in all classes was 238, matching the majority class size. The resulting balanced dataset was used as the augmented training subset for subsequent model training. After this augmentation, the resulting subset (hereinafter referred to as the augmented training subset) was further divided again (7:3) into final training and validation subsets. The validation subset was used exclusively to monitor the loss and determine early stopping. Feature standardization was performed using the mean and variance computed solely from the initial training subset, and the same parameters were applied to the test subset. The MLP was implemented using the *MLPClassifier* in Scikit-learn [29]. A rectified linear unit (ReLU) was adopted as the activation function and multiclass cross-entropy as the loss function. Model parameters were optimized using the Adam algorithm with an initial learning rate of 1×10^{-4} and an L_2 regularization coefficient of 1×10^{-3} . The model was trained using the final training subset. Throughout the training, the loss values were recorded for the validation subset to monitor convergence and prevent overfitting. The training was terminated early when the validation loss showed no improvement for 200 consecutive epochs, and the model that achieved the lowest validation loss up to that point was retained as the final model. The model performance was evaluated on an independent test

subset in terms of the confusion matrix, overall accuracy, and macro-F1 score.

The resulting trained model was applied to identify the type of welding disturbance for each 0.5-s audio segment.

2.5. SHAP analysis

To interpret the decision criteria of the trained MLP model, a SHAP analysis was performed. In this study, the SHAP values were calculated using the SHAP Python library developed by Lundberg et al. based on their original paper [20]. The Kernel Explainer method implemented in this library was applied to the trained MLP model to evaluate the contribution of each input feature to the output probabilities of the four neurons in the probability layer (Figure 4), corresponding to four disturbance classes: normal, shielding-gas interruption, tip wear, and cutting-oil contamination. For the KernelExplainer, the baseline (background) distribution was constructed from all original samples before SMOTE augmentation, comprising the entire initial training subset and the test subset. The SHAP values were computed using the mean Input vector for each combination of welding conditions and disturbance types. Specifically, for each welding condition and disturbance type, all the 131-dimensional input vectors (Section 2.3) belonging to that category were averaged to obtain a representative feature vector. The SHAP values were then calculated for these representative vectors using the final MLP classifier described in Section 2.4. This approach allows the contribution of each frequency-band feature to be interpreted at the level of the overall disturbance characteristics, rather than at the level of individual 0.5-s segments. The resulting SHAP values were subsequently analyzed, as described in Section 3.2, to identify which acoustic features and frequency bands most strongly contributed to distinguishing each disturbance type.

3. Results and discussion

3.1. Model performance

Figures 5–7 show the time-series variations in the predicted disturbance conditions for each welding case.

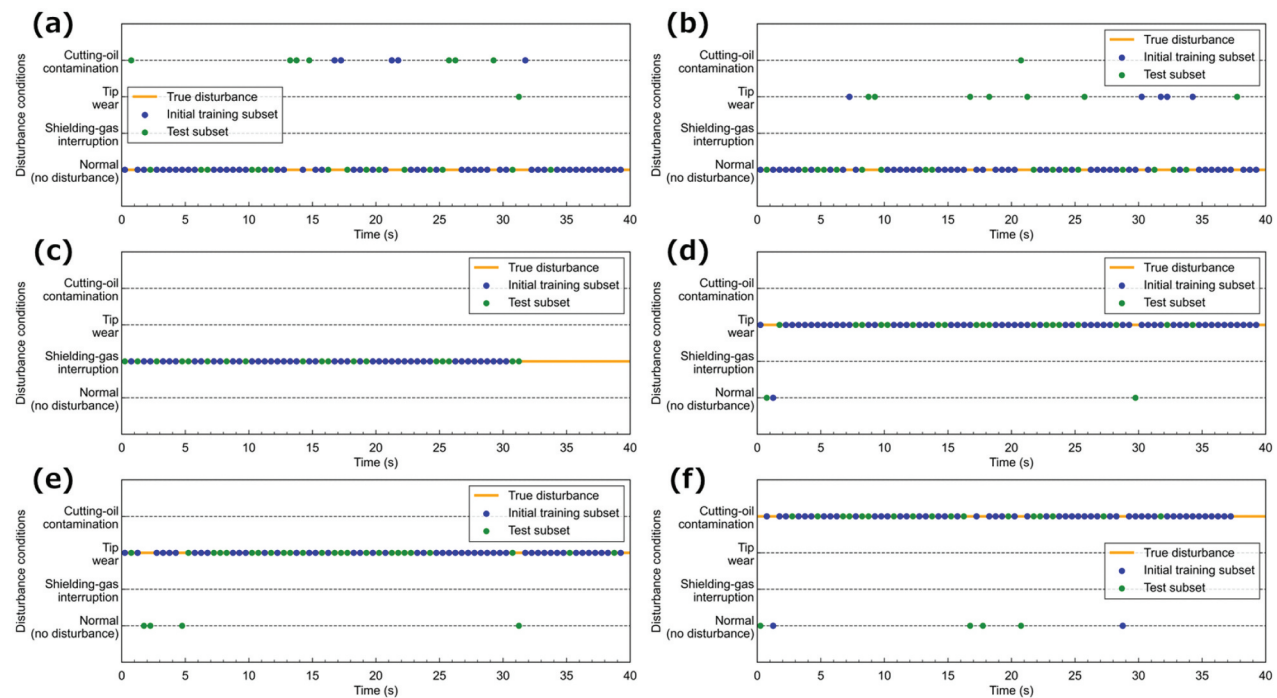


Figure 5. Temporal variation in predicted disturbance classes (Welding condition No. 1): (a) Normal welding, (b) Tip wear (0.25 mm, treated as non-disturbance), (c) Shielding-gas interruption, (d) Tip wear (0.5 mm), (e) Tip wear (1.0 mm), and (f) Cutting-oil contamination. The time origin (0 s) is defined as the welding start time for (a), (b), (d), and (e), and as the disturbance onset time for (c) and (f).

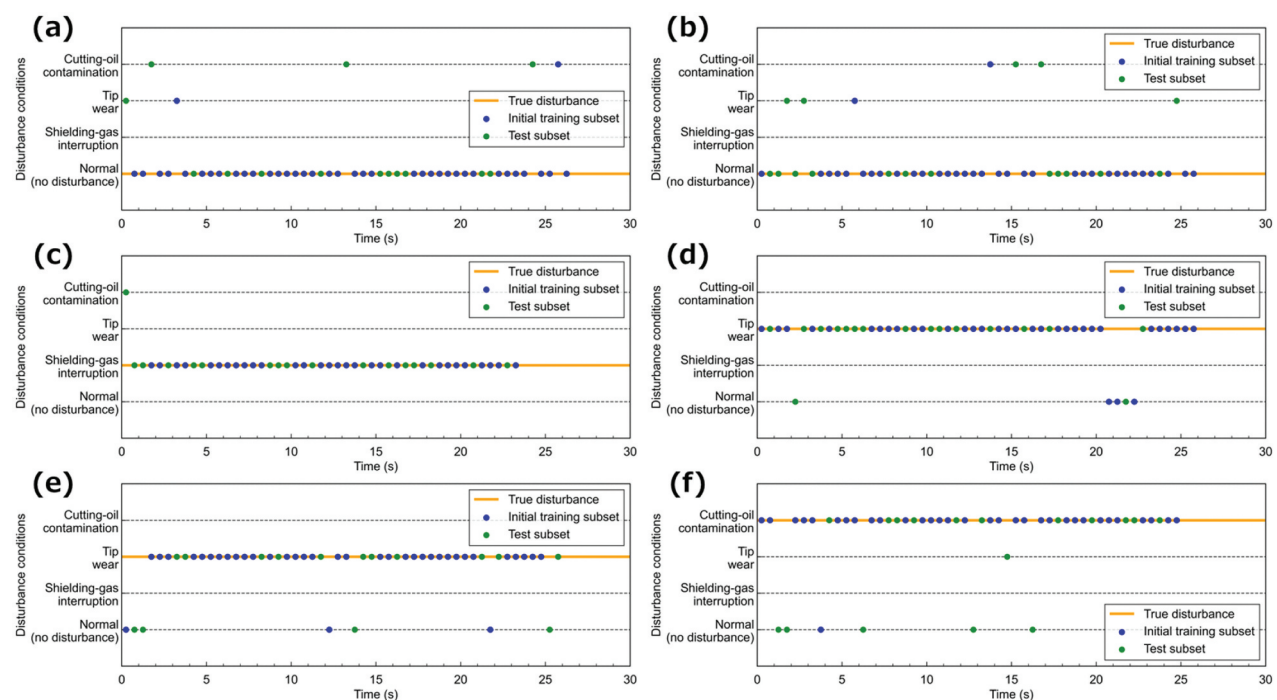


Figure 6. Temporal variation in predicted disturbance classes (Welding condition No. 2): (a) Normal welding, (b) Tip wear (0.25 mm, treated as non-disturbance), (c) Shielding-gas interruption, (d) Tip wear (0.5 mm), (e) Tip wear (1.0 mm), and (f) Cutting-oil contamination. The time origin (0 s) is defined as the welding start time for (a), (b), (d), and (e), and as the disturbance onset time for (c) and (f).

The predicted results were obtained by applying the MLP model, trained according to the procedure described in Section 2.4, to the 131-dimensional feature vectors extracted from each 0.5-s audio segment. For normal welding and tip wear conditions, the welding

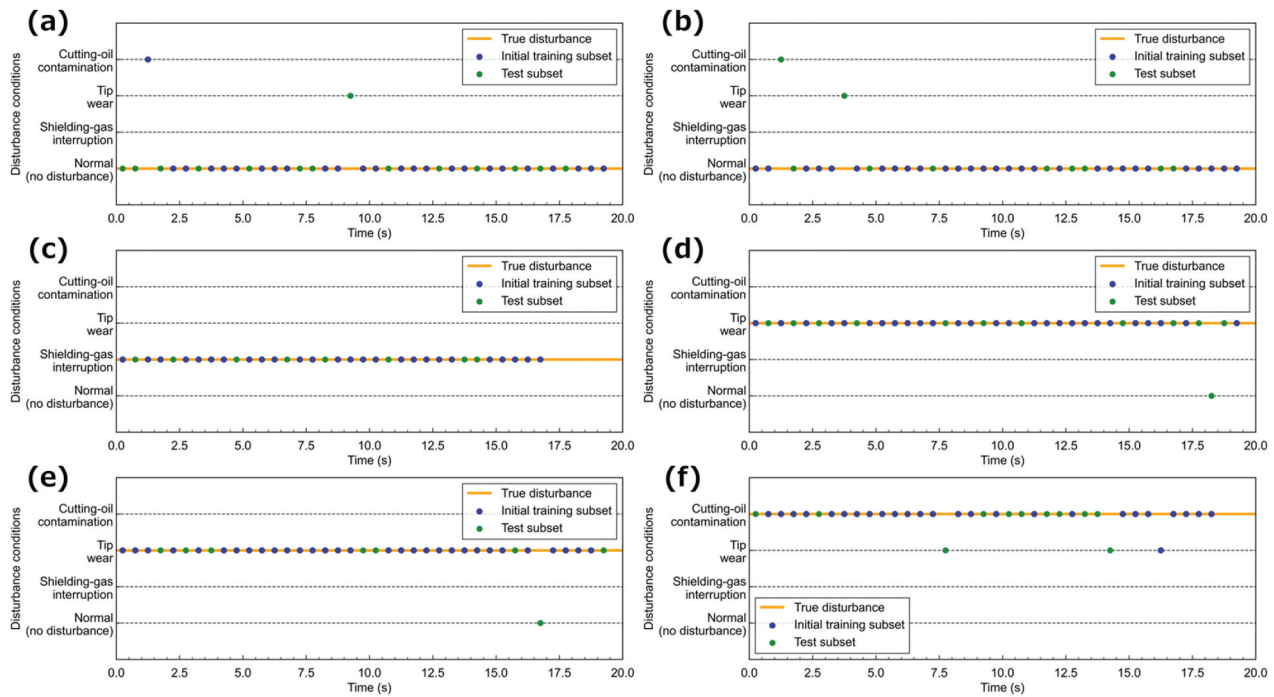


Figure 7. Temporal variation in predicted disturbance classes (Welding condition No. 3): (a) Normal welding, (b) Tip wear (0.25 mm, treated as non-disturbance), (c) Shielding-gas interruption, (d) Tip wear (0.5 mm), (e) Tip wear (1.0 mm), and (f) Cutting-oil contamination. The time origin (0 s) is defined as the welding start time for (a), (b), (d), and (e), and as the disturbance onset time for (c) and (f).

start time was defined as 0 s, whereas for the shielding-gas interruption and cutting-oil contamination conditions, the disturbance onset was defined as 0 s. The obtained results indicate that the model correctly classified the disturbance conditions from the very beginning of welding or immediately after disturbance onset, even during transient periods when the arc behavior tends to be unstable. These transient periods include the arc ignition phase in Figures 5–7(a,b,d,e), as well as the disturbance-onset phase in Figures 5–7(c,f). The classification outputs remained stable throughout the entire

welding process, with no notable fluctuations near the end of welding. Except for a single localized concentration in the range of approximately 20–23 s, as shown in Figure 6(d), the misclassifications were temporally scattered and not biased toward specific periods or welding conditions. These findings demonstrate that the trained model exhibited high temporal stability across the entire welding sequence.

Figure 8 shows the confusion matrices obtained by applying the trained model to both the augmented training and independent test subsets. Diagonal

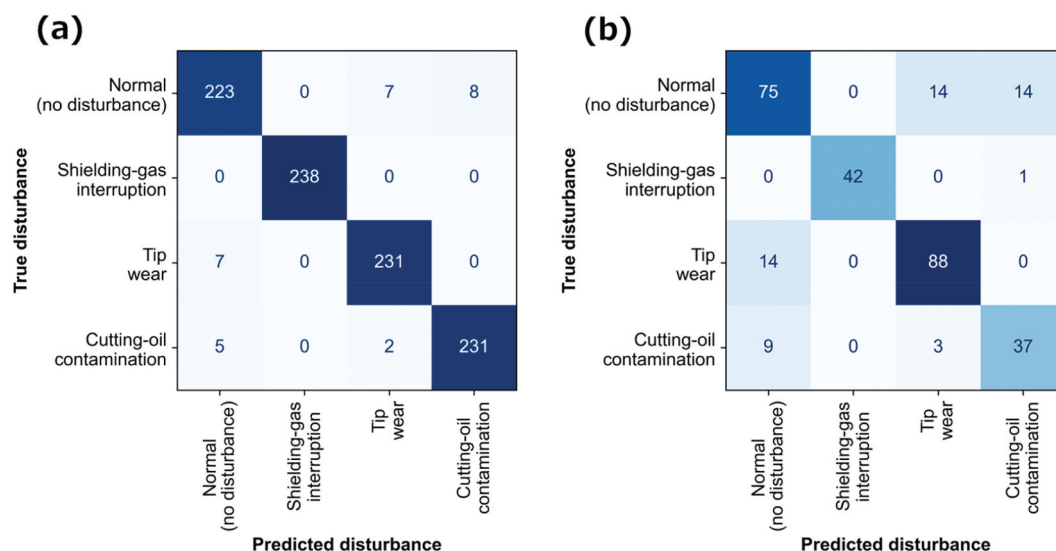


Figure 8. Confusion matrices for (a) augmented training subset before data augmentation and (b) independent test subset.

elements represent correctly classified samples, whereas off-diagonal elements indicate misclassifications. For the augmented training subset, the model achieved high accuracy, with only a small number of misclassifications, indicating that the training process converged successfully and that the model could fit the disturbance classification task under the given feature representation. The trained model also achieved a high classification performance for the independent test subset, with an overall accuracy of 81.5% and a macro-F1 score of 82.9%, confirming that it generalized well to previously unseen data. These results confirm that the trained model could accurately identify both the presence and type of disturbances and effectively capture the acoustic signatures associated with different disturbance mechanisms. However, a small number of mutual misclassifications were observed among the normal welding, tip wear, and cutting-oil contamination classes. This can be attributed to the fact that unlike shielding-gas interruptions, tip wear and

cutting-oil contamination do not always produce strong or consistent acoustic changes, and their spectral characteristics may occasionally resemble those of normal conditions depending on the process state.

3.2. SHAP-based interpretation of learned features

The SHAP analysis was applied to the MLP model trained according to the procedure described in Section 2.4, to quantify the contribution of each input feature to the disturbance classification. Figures 9 and 10 present the relationships between the acoustic feature patterns and corresponding SHAP values for each disturbance type under the three welding conditions at each frequency band. Each figure contains two panels corresponding to each disturbance label. The upper panel shows the acoustic features, which were obtained by taking the mean input vector for each welding condition –

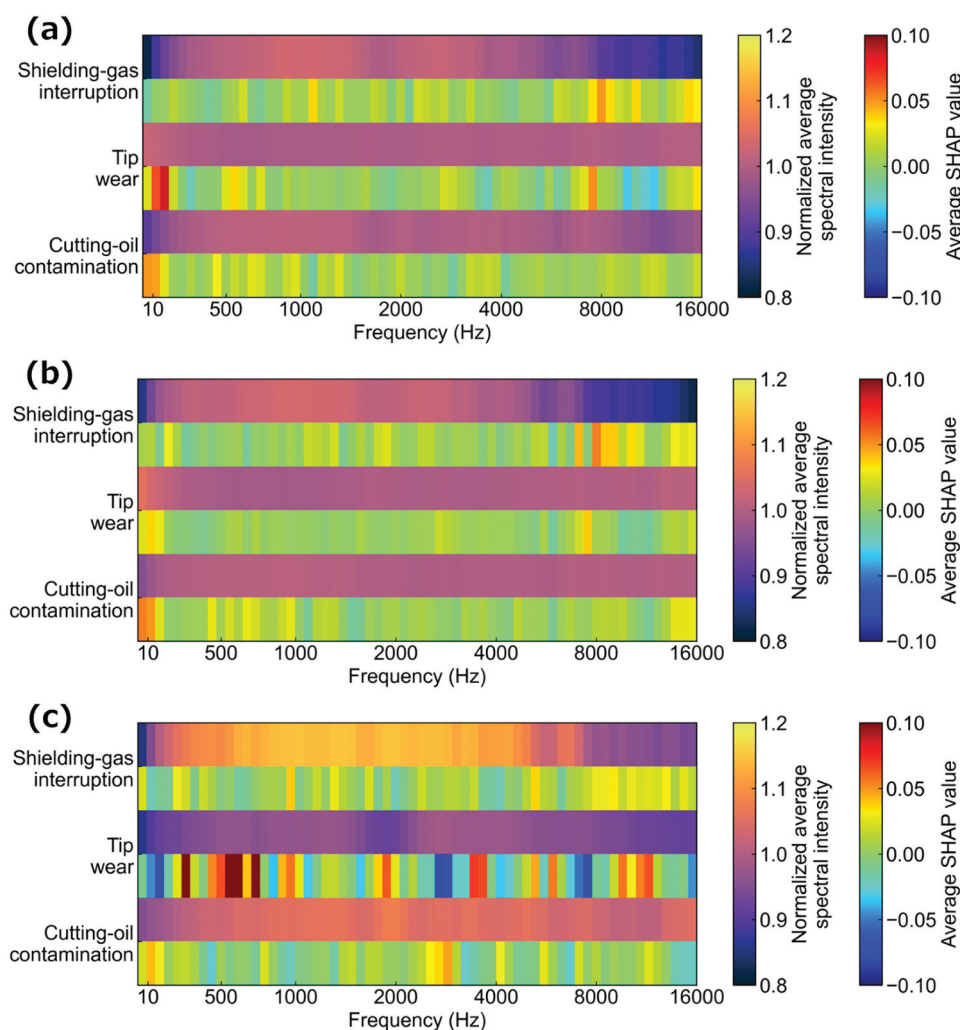


Figure 9. Normalized average spectral intensity and corresponding SHAP values for each disturbance type under the three welding conditions. Panels (a), (b), and (c) correspond to welding conditions 1, 2, and 3, respectively. For each disturbance type, the upper panel displays the normalized average spectral intensity derived from the mean input vector of that condition (Section 2.5), while the lower panel shows the SHAP values computed for the same input vector. The horizontal axis is annotated with selected frequencies; however, the tick positions are not evenly spaced because they are mapped onto the nonlinearly distributed center frequencies of the Mel filterbank.

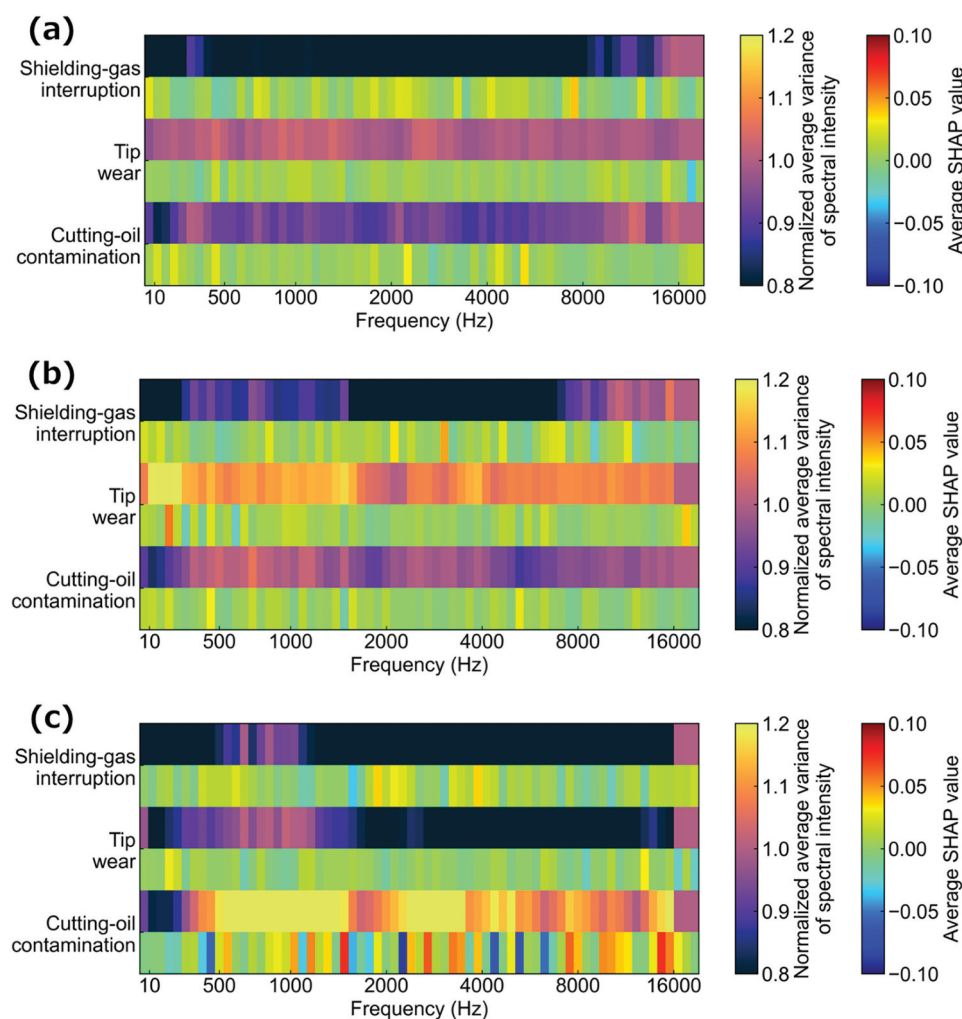


Figure 10. Normalized average variance in the spectral intensity and corresponding SHAP values for each disturbance type under the three welding conditions. Panels (a), (b), and (c) correspond to welding conditions 1, 2, and 3, respectively. The upper panel presents the normalized average variance in the spectral intensity derived from the mean input vector, and the lower panel shows the SHAP values computed for the same input vector. The horizontal axis is annotated with selected frequencies; however, the tick positions are not evenly spaced because they are mapped onto the nonlinearly distributed center frequencies of the Mel filterbank.

disturbance type combination (as defined in Section 2.5) and then normalizing it by dividing each component by the corresponding mean vector of the normal welding condition under the same welding parameters. In Figures 9 and 10, only the acoustic components of the normalized vectors are shown. Figure 9 shows the first 64 dimensions corresponding to the spectral intensity, and this set of values is referred to as the normalized average spectral intensity. Figure 10 shows the next 64 dimensions corresponding to the variance in the spectral intensity, and this set of values is referred to as the normalized average variance in the spectral intensity. The remaining three dimensions corresponding to the welding parameters (current, voltage, and travel speed) were omitted. This normalization highlights the relative deviations in the spectral intensity or variance caused by each disturbance. The lower panel shows the SHAP values computed for the same mean input vector used

to derive the acoustic features in the upper panel. The SHAP values represent the contribution of each frequency band feature to the classification output. A positive SHAP value implies that, relative to the background distribution, the actual value of the corresponding normalized feature contributes to increasing the predicted probability of that class in the probability layer (see Figure 4), whereas a negative value implies that it contributes to decreasing the probability. The absolute SHAP value reflects the relative importance of the frequency component in determining the final classification.

As shown in Figure 9, for the shielding-gas interruption, all welding conditions resulted in a decrease in the normalized average spectral intensity above 8 kHz, accompanied by high positive SHAP values in the same frequency range. In addition, a relatively high normalized average spectral intensity is observed around 1 kHz, and the corresponding frequency region

also exhibits relatively high positive SHAP values. This implies that both the attenuation of high-frequency components and the relative enhancement of low-frequency components contribute to disturbance discrimination during gas interruption. These spectral changes are likely associated with modifications in arc plasma discharge behavior caused by atmospheric intrusion after the shielding gas is removed.

For tip wear, under Welding Conditions 1 and 2, both the normalized average spectral intensity and SHAP values were high in the low-frequency range (tens of Hertz) (Figures 9(a,b)). These low-frequency components are likely related to changes in the mechanical and electrical interaction between the welding wire and the worn tip, which can introduce fluctuations in arc length and wire-feeding behavior. The MLP model effectively recognized this vibration-induced acoustic pattern as a characteristic feature of tip wear. In contrast, under Welding Condition 3, an increase in the low-frequency intensity was not observed (Figure 9(c)), probably because the higher current and voltage strengthened the arc force and increased the wire-feeding speed, thereby suppressing the mechanical vibration. Nevertheless, the relatively high SHAP values distributed over a broader frequency range indicate that the model used other spectral features to identify the worn-tip condition.

For cutting-oil contamination, as shown in Figure 9, all the welding conditions resulted in a lower normalized average spectral intensity in the low-frequency range (tens of Hertz), along with higher SHAP values. This behavior may be associated with the destabilization of the cathode attachment point caused by surface oil contamination, which can modify the arc discharge behavior. The evaporation of cutting oil and its possible mixing with the arc plasma may also contribute to the observed spectral changes. Under Welding Condition 3, the absolute SHAP values corresponding to the normalized average variance in the spectral intensity were high over a wide frequency range (Figure 10(c)). Because the low-frequency intensity reduction is common to multiple disturbances under this condition, the model appears to have utilized variance-related information as an auxiliary feature for accurate discrimination. In addition, under other welding conditions and for other disturbance types, as shown in Figure 10, certain frequency bands exhibited high absolute SHAP values, indicating that the normalized average variance in the spectral intensity also contributed to the classification.

These results demonstrate that the developed AI-based disturbance-detection model learns and identifies acoustic features consistent with the physical phenomena that occur during disturbance generation in arc welding. Changes in both spectral intensity and its variance were found to be important factors for accurate disturbance classification.

3.3. Practical implications and limitations

The analyses presented in the previous sections demonstrate that the AI-based disturbance-detection model developed in this study is an effective method for detecting disturbances in arc welding processes using a low-cost noncontact microphone. Because this approach requires neither optical protection nor expensive imaging equipment, it can be readily integrated into existing robotic welding systems, thereby significantly reducing implementation costs in industrial environments. In particular, its capability to identify both the occurrence and type of disturbance in real-time through a 0.5-s segment analysis offers a practical advantage directly linked to in-process quality assurance, thereby helping reduce the inspection burden associated with post-process visual or nondestructive testing. Another notable finding is that the SHAP analysis makes it possible to identify the frequency bands and features that the model relies on for disturbance classification, thereby confirming that its decision criteria are physically consistent with the actual welding behavior. The SHAP analysis also provides information on the contribution of the welding parameters (current, voltage, and travel speed). However, because only three welding conditions were examined in the present experiments, the present dataset does not allow a reliable discussion of the relative influence of these parameters.

However, several challenges need to be overcome. First, the influence of environmental factors, such as the position and orientation of the microphone and the surrounding noise, on the model output should be quantitatively evaluated. Second, the present experiments were limited to the bead-on-plate welding of SM490 steel. Therefore, future work should examine the applicability of the model to different materials, groove geometries, and multilayer welding. Third, the disturbance categories should be expanded to include complex anomalies that may occur during production, such as power supply fluctuations, torch-position deviations, and gas flow variations.

Another limitation concerns the manner in which the test subset was constructed. Ideally, generalization should be tested by holding one entire welding condition as the test set and evaluating whether the model can classify disturbances under unseen welding parameters. However, when this approach was used in this study, the classification accuracy decreased substantially. This is likely because only three welding conditions were available, which were insufficient for the model to learn acoustic patterns that were invariant to changes in the current, voltage, and travel speed settings. In practical industrial environments, robotic welding is typically preceded by preliminary welding trials conducted under the same welding parameters as those used in production. From this perspective, the

evaluation approach adopted in this study, in which the test subset was sampled from the same welding conditions but at different time segments, remains meaningful because it reflects how the model would actually be deployed. To apply the proposed method to a production line, it is necessary to collect disturbance sound data under specific welding parameters used onsite and train the model accordingly. This requirement does not impose a substantial burden, given that such preliminary welding trials are routinely conducted as part of the standard welding procedure setup in industrial practice.

Future research will focus on extending the system to multimodal AI monitoring, combining microphones with other sensors (e.g. current and voltage sensors), and implementing real-time inference in edge-computing environments. By integrating multimodal signals, the framework can evolve from detecting disturbance occurrences to predicting their precursors depending on the disturbance type. Through these developments, the proposed AI-based disturbance-detection model has the potential to serve as a foundation for autonomous welding quality assurance systems in future smart factory applications.

4. Conclusions

We developed an AI-based disturbance-detection model that can automatically identify the occurrence and type of welding disturbances from microphone-recorded acoustic signals. The main findings are summarized as follows:

- Representative disturbances in the GMAW process – shielding gas interruption, tip wear, and cutting-oil contamination – were determined experimentally. The recorded welding sounds were converted into Mel spectrograms, and the acoustic features were extracted from each frequency band.
- An MLP model was constructed using the extracted acoustic features along with welding parameters (current, voltage, and travel speed). The trained model achieved a high classification performance, attaining an overall accuracy of 81.5% and a macro-F1 score of 82.9%, indicating robust generalization to unseen data.
- The time-series evaluation demonstrated that the model could continuously and stably classify disturbance conditions throughout the welding process, with only minor fluctuations during transient periods, such as at the start or end of welding.
- The SHAP analysis revealed that the model referred to frequency bands that were physically consistent with the mechanisms of each disturbance. For example, the attenuation of high-frequency components was dominant in

shielding-gas interruptions, whereas low-frequency vibration components were characteristic of tip wear. Moreover, both the spectral intensity and its variance were found to play essential roles in disturbance classification, indicating that the model simultaneously utilizes amplitude- and fluctuation-related acoustic information to achieve accurate discrimination among different disturbance types.

- The proposed method provides a low-cost, non-contact, and real-time monitoring solution that can be easily integrated into existing robotic welding systems, offering practical benefits for in-process quality assurance in arc-welding operations.

Future work will aim to integrate additional sensing modalities, such as current and voltage sensors, to establish multimodal AI monitoring and implement real-time inference in edge-computing environments. These developments are expected to help advance the proposed framework from disturbance detection to predictive diagnosis, thereby contributing to the realization of autonomous welding quality assurance systems in smart manufacturing.

Acknowledgements

The authors would like to express their sincere gratitude to Mr. Keisuke Torigata of the IHI Corporation for his valuable advice on the experimental conditions and assumed disturbances during the course of this study.

Disclosure statement

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

ORCID

Houichi Kitano  <http://orcid.org/0000-0002-0778-574X>
Masahiko Demura  <http://orcid.org/0000-0002-7308-3041>

References

- [1] Barrera G, Velez M, Espinosa MA, et al. Monitoring the submerged-arc welding (SAW) process using current and voltage transducers. *Weld Int.* 1997;11(10):795–801. doi: [10.1080/09507119709454427](https://doi.org/10.1080/09507119709454427)
- [2] Kah P, Edigbe GO, Ndiwe B, et al. Assessment of arc stability features for selected gas metal arc welding conditions. *SN Appl Sci.* 2022;4(10):268. doi: [10.1007/s42452-022-05150-5](https://doi.org/10.1007/s42452-022-05150-5)
- [3] Suban M, Tušek J. Methods for the determination of arc stability. *J Mater Process Technol.* 2003;143–144:430–437.
- [4] He K, Si Y, Lu W, et al. Time frequency feature extraction of the arc energy for quality detection of the aluminum alloy double pulse MIG welding. *J Adv Mech Des*

- Syst Manuf. 2020;14(6):JAMDSM0080–JAMDSM0080. doi: [10.1299/jamdsm.2020jamdsm0080](https://doi.org/10.1299/jamdsm.2020jamdsm0080)
- [5] Kim YM, Cheon J. Recent research trend of arc welding quality monitoring technology in Korea. *J Weld Join*. 2023;41(2):81–89. doi: [10.5781/JWJ.2023.41.2.1](https://doi.org/10.5781/JWJ.2023.41.2.1)
 - [6] Kim GG, Kim DY, Yu J. Vision-based acquisition model for molten pool and weld-bead profile in gas metal arc welding. *Metals*. 2024;14(12):1413. doi: [10.3390/met14121413](https://doi.org/10.3390/met14121413)
 - [7] Kim GG, Kim YM, Kim DY, et al. Vision-based online molten pool image acquisition and assessment for quality monitoring in gas–metal arc welding. *Appl Sci*. 2024;14(14):1–5. doi: [10.3390/app14145998](https://doi.org/10.3390/app14145998)
 - [8] Zhu Y, Wu Z, Liu C. Research on mapping recognition of arc welding molten pool characterisation and penetration state based on embedded system. *IET Comput Digit Tech*. 2023;17(3–4):100–110. doi: [10.1049/cdt2.12055](https://doi.org/10.1049/cdt2.12055)
 - [9] Cheng Y, Yu R, Zhou Q, et al. Real-time sensing of gas metal arc welding process – a literature review and analysis. *J Manuf Process*. 2021;70:452–469. doi: [10.1016/j.jmapro.2021.08.058](https://doi.org/10.1016/j.jmapro.2021.08.058)
 - [10] Yu R, Cao Y, Chen H, et al. Deep learning based real-time and in-situ monitoring of weld penetration: where we are and what are needed revolutionary solutions? *J Manuf Process*. 2023;92:15–46.
 - [11] Lee JG, Kim KS, Lee JH. Sound-based unsupervised fault diagnosis of industrial equipment considering environmental noise. *Sensors*. 2024;24(22):7319. doi: [10.3390/s24227319](https://doi.org/10.3390/s24227319)
 - [12] AlShorman O, Alkhatni F, Masadeh M, et al. Sounds and acoustic emission-based early fault diagnosis of induction motor: a review study. *Adv Mech Eng*. 2021;13(2):1687814021996915. doi: [10.1177/1687814021996915](https://doi.org/10.1177/1687814021996915)
 - [13] Ding S, Zhang S, Yang C. Machine tool fault classification diagnosis based on audio parameters. *Results Eng*. 2023;19:101308. doi: [10.1016/j.rineng.2023.101308](https://doi.org/10.1016/j.rineng.2023.101308)
 - [14] Jombo G, Zhang Y. Acoustic-based machine condition monitoring—methods and challenges. *Eng*. 2023;4(1):47–79. doi: [10.3390/eng4010004](https://doi.org/10.3390/eng4010004)
 - [15] Hu S, Guo G, Jiang Z. Research on anomaly sound detection methods based on large models. In: *Proceedings of the 2025 9th International Conference on Control Engineering and Artificial Intelligence*, Association for Computing Machinery; 2025 Jan 16–19; Ho Chi Minh City, Vietnam: Association for Computing Machinery, New York, NY, United States; 2025. p. 15–22.
 - [16] Appadoo R, Xu Y, Gu F, et al. Performance monitoring and fault diagnosis of vacuum pumps based on airborne sounds. In: *2018 24th International Conference on Automation and Computing (ICAC)*; 2018 Sep 6–7; Newcastle Upon Tyne, (UK): IEEE; 2019. p. 1–6.
 - [17] Salm T, Tatar K, Chilo J. Real-time acoustic measurement system for cutting-tool analysis during stainless steel machining. *Machines*. 2024;12(12):892. doi: [10.3390/machines12120892](https://doi.org/10.3390/machines12120892)
 - [18] Dhonde DSB, Chaudhari AA, Gajare MP. Performance evaluation of Mel and Bark scale based features for text-independent speaker identification. *Int J Innov Technol Explor Eng*. 2019;8(11):3734–3738. doi: [10.35940/ijitee.K1999.0981119](https://doi.org/10.35940/ijitee.K1999.0981119)
 - [19] Li H, Li J, Liu H, et al. MelTrans: mel-spectrogram relationship-learning for speech emotion recognition via transformers. *Sensors*. 2024;24(17):5506. doi: [10.3390/s24175506](https://doi.org/10.3390/s24175506)
 - [20] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*; Long Beach, (CA), USA: Curran Associates Inc; 2017. p. 4768–4777.
 - [21] Demura M. Challenges in materials integration. *ISIJ Int*. 2024;64(3):503–512. doi: [10.2355/isijinternational.ISIJINT-2023-399](https://doi.org/10.2355/isijinternational.ISIJINT-2023-399)
 - [22] Koseki T. Materials data and materials integration system. *J Inf Process Manag*. 2016;59(3):165–171.
 - [23] Demura M, Koseki T. Sip-materials integration projects. *Mater Trans*. 2020;61(11):2041–2046. doi: [10.2320/matertrans.MT-MA2020003](https://doi.org/10.2320/matertrans.MT-MA2020003)
 - [24] Demura M. Materials integration for accelerating research and development of structural materials. *Mater Trans*. 2021;62(11):1669–1672. doi: [10.2320/matertrans.MT-M2021135](https://doi.org/10.2320/matertrans.MT-M2021135)
 - [25] Morris M, Quinn TP, Siewert T, et al. Sensing of contact tube wear in gas metal arc welding. Gaithersburg (MD): US Department of Commerce, National Institute of Standards and Technology; 1992.
 - [26] McFee B, McVicar M, Faronbi D, et al. Librosa/librosa: 0.11.0, Zenodo 0.11.0. 2025.
 - [27] Chawla NV, Bowyer KW, Hall LO, et al. Smote: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–357. doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)
 - [28] Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*. 2017;18(1):559–563.
 - [29] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.