Monitoring of Joint Gap Formation in Laser Beam Butt Welding Using Neural Network-Based Acoustic Emission Analysis

Saichand Gourishetti 1,*, Leander Schmidt 2, Florian Römer 3, Klaus Schricker 5, Sayako Kodera 3, David Böttger 3, Tanja Krüger 1, András Kátai 1, Joachim Bös 1,4, Benjamin Straß 3, Bernd Wolter 3 and Jean Pierre Bergmann 2

Abstract: This study aimed to explore the feasibility of using airborne acoustic emission in laser beam butt welding for the development of an automated classification system based on neural networks. The focus was on monitoring the formation of joint gaps during the welding process. To simulate various sizes of butt joint gaps, controlled welding experiments were conducted, and the emitted acoustic signals were captured using audible-to-ultrasonic microphones. To implement an automated monitoring system, a method based on short-time Fourier transformation was developed to extract audio features, and a convolutional neural network architecture with data augmentation was utilized. The results demonstrated that this non-destructive and non-invasive approach was highly effective in detecting joint gap formations, achieving an accuracy of 98%. Furthermore, the system exhibited promising potential for the low-latency monitoring of the welding process. The classification accuracy for various gap sizes reached up to 90%, providing valuable insights for characterizing and categorizing joint gaps accurately. Additionally, increasing the quantity of training data with quality annotations could potentially improve the classifier model's performance further. This suggests that there is room for future enhancements in the study.

Keywords: laser beam butt welding (LBW); joint gap formation; AE analysis; non-destructive testing (NDT); deep learning; audible-to-ultrasonic sensors

1. Introduction

The laser beam welding of butt-jointed sheets can be challenging due to the presence of a gap between the sheets to be joined (joint gap). The formation of the gap can be attributed to an improper pre-processing of the wrought material (e.g., production tolerances) and can be also result from a process-induced specimen displacement. The melting and resolidification of the material during the welding process results in the formation of strain, which can affect the size of the joint gap during the welding process by causing a specimen displacement [1,2]. Overall, it can be said that the joint gap formation depends on a variety of different factors that can interact during the welding process. In terms of weld seam quality and process stability, the gap size is of fundamental importance [3]. By affecting the energy absorption within the keyhole and the conditions...
of melt and vapor flow [4,5], the joint gap has a major impact on the pressure balance of the keyhole, which is crucial for keyhole stability [6]. As a result, the formation of weld imperfections (e.g., spatter, pores) can differ depending on the gap size [4,5]. It can therefore be concluded that the detection of the gap during the welding process is of fundamental importance in terms of quality assurance and environmental efficiency.

State-of-the-art methods for classifying and detecting joint gaps typically involve optical measuring techniques (e.g., photodiode sensors [7], spectrosopes, and cameras [8,9]) and tactile measuring techniques (e.g., inductive probes [2,10]). However, these methods often require expensive modifications to clamping and system technology, as well as direct access to the measurement site. In contrast, acoustic process monitoring has proven to be highly advantageous due to its ability to integrate sensors into fixtures (airborne and structure-borne acoustic emission sensors) and position them in versatile external locations, allowing for adaptable sensor placement. This approach offers the added benefit of detecting hidden seam defects. Previous studies have successfully utilized the acoustic analysis of process emissions to predict spatter-induced loss of mass [11], monitor penetration depth [12,13], differentiate between heat-conduction and deep-penetration welding processes [14,15], and establish connections between disparate sensor systems using feature-based machine learning [16]. Despite these advancements, the implementation of an acoustic process monitoring system specifically designed to detect and quantify joint gaps in laser beam welding has not yet been achieved.

Acoustic-based fault detection in welding poses several challenges. Firstly, there is a significant cost associated with collecting a substantial amount of industrial data, which must then be annotated by experts. Secondly, the diversity of data presents a challenge, as it is impractical to simulate all potential faults in the welding process. The third challenge revolves around selecting and placing appropriate sensors for process measurement, considering their sensitivity to background noise. Additionally, AI-based automated quality monitoring systems face significant challenges in industrial settings due to factors such as large volumes of data, data imbalance, and the need for expert annotations. To address these challenges, various strategies have been proposed in recent years. Data scarcity can be overcome using data augmentation methods, such as random rotations, spec augment, mixup, or adding noise to the actual training data, which create synthetic data points through transformations [17,18]. These techniques help tackle the diversity challenge by introducing more varied instances. Further, data imbalance can be addressed using the synthetic minority over-sampling technique (SMOTE) method [19], which generates synthetic examples for the minority class. These countermeasures enhance model effectiveness and fairness.

Acoustic-based fault detection methods typically involve the analysis of various attributes of sound signals including temporal patterns, frequency components, and signal intensity. However, these methods can be affected by factors such as variations in background noise, sensor positioning, and minor process deviations, which can impact the distinctive characteristics of the sound signals. To address these challenges, contemporary data processing techniques and robust training approaches for deep learning models can be employed to facilitate the detection and extraction of process-relevant signal components spanning the entire audible-to-ultrasonic frequency range [20,21].

Recent studies in industrial sound analysis have yielded promising results by employing Convolutional Neural Network (CNN)-based methods for audio event detection and classification [22–25]. These techniques have proven successful when applied to time–frequency representations, specifically spectrograms, as CNNs possess the capacity to capture local patterns and spatial dependencies within audio signals. Furthermore, the neural networks’ capability to learn complex patterns, ability to perform end-to-end learning, and scalability and adaptability have motivated their selection over traditional machine learning algorithms [24]. Additionally, industrial audio signals, such as those from welding processes, often contain substantial background noise,
necessitating laborious manual feature engineering to extract relevant signal characteristics [25]. However, CNNs can automatically learn meaningful audio features, including frequency patterns and spectral characteristics, without the need for additional manual feature engineering [23].

This study addresses a significant gap in the field of laser beam butt welding, which is the absence of open-source datasets and the underexplored use of airborne acoustic analysis spanning from audible to ultrasonic ranges. This study aims to leverage machine learning techniques to harness acoustic information for detecting gap formation, providing valuable insights into this specialized application of welding technology. The paper presents substantial contributions by systematically investigating joint gap formation through controlled experiments, incorporating acoustic emission measurements and high-speed cameras. It demonstrates the feasibility of monitoring joint gaps using airborne acoustic emission, introduces an accurate automated classification system, and improves defect detection in high-alloy steel welding. Additionally, the study offers insights into process parameters and outlines future enhancements to enhance the reliability and applicability of the acoustic monitoring system.

2. Materials and Methods

2.1. Specimen Configuration

AISI 304 (X5CrNi18-10, 1.4301) high-alloy austenitic steel with a sheet thickness of 1 mm was used. To analyze the effect of the joint gap on the welding process and correlated acoustic emissions under consistent experimental conditions, the joint gap was manually pre-set. Therefore, defined gaps were introduced by machining notches (0.1 mm, 0.2 mm, and 0.3 mm) into one of the two butt-jointed sheets at two different positions (Are Figure 1). This resulted in a total of 5 different patch sections (3 x zero gap, 2 x joint gap) for each butt joint configuration. An example of a welded specimen with 0.2 mm gap size (Figure 1a) and the corresponding specimen configuration (Figure 1b) is shown in the figure below. However, these gap sizes may be affected during the welding process due to process-induced strain (Section 1), which has been observed via high-speed recordings (Section 3.1.1). Therefore, the nominal gap sizes of 0.1 mm, 0.2 mm, and 0.3 mm are referenced prior to the start of welding. The direction of rolling of the sheets was the same as the welding direction.

![Figure 1. (a) Example of welded specimen with 0.2 mm joint gap; (b) specimen configuration including patch segmentation.](image-url)
2.2. Weld Setup

A disk laser (TruDisk 5000.75, Trumpf Laser- und Systemtechnik GmbH, Ditzingen, Germany) with a maximum power of \( P_{\text{max}} = 5 \text{ kW} \) and a wavelength of \( \lambda = 1030 \text{ nm} \) was coupled to a stationary arranged processing head (BEO D70, Trumpf Laser- und Systemtechnik GmbH, Ditzingen, Germany) providing a focal diameter of \( d_{\text{spot}} = 600 \mu \text{m} \) for laser beam butt welding (cf. Figure 2). The specimens were clamped to the front of a six-axis robot (Kuka KR 60 HA, Kuka AG, Augsburg, Germany), which was used for sample handling. All specimens were welded at a welding speed of \( v_{\text{w}} = 12 \text{ m/min} \) using a laser power of \( P_{\text{L}} = 3.5 \text{ kW} \). The laser power was set for a full penetration. Microphones were mounted on the stationary processing head and oriented to the position of the laser-induced keyhole to record acoustic process emissions. Technical specifications are given in Table 1. The AE sensor data were captured using a data acquisition system (Soundbook MK2, SINUS Messtechnik GmbH, Leipzig, Germany) operating with a sampling rate of 204.8 kHz per channel. A high-speed camera (SA-X2, Photron, Tokyo, Japan) equipped with a zoom lens (12X Zoom Lens System, Navitar, Ottawa, Canada) was used to record the process at 10,000 frames per second. The camera was positioned at an angle of approximately 85° to the specimen surface. A narrow-band optical filter with a center wavelength of 808 nm was used to eliminate process emissions. The welding process was illuminated using a laser illumination system (Cavilux HF, Cavitar Ltd., Tampere, Finland) operating at a wavelength of 808 nm. The cross jet was turned off during the experiments to provide an idealized data acquisition. Each gap configuration (0.1 mm, 0.2 mm, and 0.3 mm) was repeated 30 times to provide sufficient input for data processing.

Both microphones, the sE8 and the MK301, employed a sampling rate of 204.8 kHz for data acquisition. However, for the sE8 microphone, only frequencies up to 25.6 kHz were considered for further analysis since this is an audible range sensor. Additional information can be found in Table 1. Prior to finalizing the sensor positions for measurements, it is imperative to emphasize that the distances between sensors, which, in this case, were microphones, were carefully assessed. This assessment included a thorough examination of signal quality, involving measures such as checking for signal attenuation, background noise from reflections, and other relevant factors. This comprehensive evaluation was conducted to ensure practicality, maintain signal integrity, facilitate comparability, and enhance statistical reliability. This approach guaranteed that our microphone placement was not only strategically determined but also optimized for precise and reliable data collection.
Table 1. Technical specifications of the used microphones.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Frequency Range</th>
<th>Sensor Distance $d_{mic}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sE electronics sE8</td>
<td>Broadcast microphone, directional characteristics</td>
<td>20–20,000 Hz</td>
<td>262 mm</td>
</tr>
<tr>
<td>Microtech Gefell GmbH MK301</td>
<td>Free field microphone, encapsulated</td>
<td>5–100,000 Hz</td>
<td>282 mm</td>
</tr>
</tbody>
</table>

3. Results

This section focuses on examining the impact of joint gap formation in laser beam butt welding. It will explore the data analysis of the acoustic signals produced during the welding process, specifically considering the relationship between the varying gap formations between metal sheets. Furthermore, an experimental design will be presented to investigate a classification system based on airborne acoustics. Finally, the obtained results will be discussed, providing plausible explanations for the observed phenomena.

3.1. Effect of Butt Joint Gap on Welding Process

3.1.1. Keyhole and Melt Pool Dynamics

In order to correlate the acoustic signal with gap-related events in the welding process, the keyhole and melt pool dynamics were analyzed as functions of the gap using high-speed video recordings. Since the effect of a 0.1 mm gap size was negligible and the effects of 0.2 mm and 0.3 mm gaps were nearly identical, Figure 3 shows only the effect of a 0.2 mm gap size as an example.

![Figure 3](image-url)  

- **Patch 1** (zero gap, cf. Figure 3a): During zero-gap welding, a stable energy input within the keyhole and a symmetrical weld seam formation were observed. The upper keyhole aperture was almost cylindrically shaped, while the lower keyhole aperture was slightly elongated in the welding direction. The resulting weld seam could be classified as defect-free.

- **Patch 2** (0.2 mm gap, $t_{frame} = \text{gap start}$, cf. Figure 3b): At the start of the 0.2 mm gap patch, an unstable welding process with an asymmetric weld seam formation was observed, which resulted in a lack of fusion and shape deviations (e.g., weld seam undercuts) in the two sheets. A significant decrease in melt pool length was also observed.
• Patch 2 (0.2 mm gap, \( t_{\text{frame}} > \text{gap start} \), cf. Figure 3c): By continuing the 0.2 mm gap patch, changes in the gap size and the corresponding keyhole and melt pool behavior were observed. The increase in the weld lengths resulted in both an increase and an atypical decrease in gap size. A joint gap is formed during the welding process primarily due to process-induced strain (e.g., solidification-induced contraction of the weld), typically grows with an increasing weld length, and depends, among other things, on the sheet thickness, the welding speed, and the specimen fixture [2,3,26]. The variation in gap size resulted in both a deviating energy input as well as the formation of weld asymmetries.

Since the joint gap was found to have a significant effect on the keyhole and melt pool dynamics, it is conceivable that the gap also influences the acoustic emissions. Therefore, the following section gives an overview of the affected airborne sound.

3.2. Dataset Preparation

The welding experiments involved four distinct data classes: Zero Gap (zero gap), Gap 0.1 (0.1 mm gap), Gap 0.2 (0.2 mm gap), and Gap 0.3 (0.3 mm gap). Each experiment consisted of three patches of Zero Gap and two patches of Gap. The welding process for each patch lasted 160 ms without any delay between the transition from Zero Gap to Gap or vice versa. To remove the welding transition between consecutive patches, each patch recording was trimmed into segments with an onset and offset of 5 ms, resulting in a 150 ms recording per patch. The total number of files per class in the dataset used for the work is given in Figure 4.

Utilizing the high-speed video recordings of the welding process (cf. Figure 2), the estimation of the real gap sizes between all Zero Gap patches was made possible. These inspections unveiled the presence of unintended non-uniform gaps between the sheets, which can be attributed to the process-induced thermal expansion of molten metal [27,28] and the solidification-induced contraction of the weld seam after the welding process [27,28]. Leveraging the available additional gap information, interpolation techniques were employed to approximate the additional gaps for the Gap 0.1, Gap 0.2, and Gap 0.3 regions.

In the study, the presence of non-uniform gaps between the metal sheets in both Zero Gap and Gap patches led to the mislabeling of the pre-labeled data. To address this, a threshold of 0.06 mm was established for the additional gap in the Gap 0.1 and Gap 0.2 classes. The researchers carefully considered this decision, aiming to balance precision and practicality. Setting a smaller threshold might have led to overly fine-grained gap classification, resulting in sparse data and hindering the model’s learning capacity. Conversely, a larger threshold (0.06 mm) was chosen to accommodate the natural variations in gap sizes often present in real-world scenarios while ensuring sufficient samples for each class.

The 0.06 mm threshold was established based on a thorough dataset analysis, including the distribution of gap sizes and potential variations. The threshold was also chosen based on the reached production tolerances during the sample preparation. Since the milling machine used (Spinner U-620) has a machining accuracy in practice of about 0.06 mm, the threshold was derived from the achievable production tolerances. Due to a periodical measurement of the milling machine tool offset, it can be assumed that the milling accuracy is kept constant. The researchers also conducted experiments with different threshold values to assess their impact on dataset size and model performance. By incorporating these insights into the manuscript, the researchers aim to provide a comprehensive understanding of the rationale behind the chosen threshold and its suitability in handling varying gap sizes within each class.

Table 2 presents the minimum and maximum range of actual gap sizes for each class.
Figure 4. Classes and number of files per class in the dataset.

Table 2. The actual gap sizes between the metal sheets specified in the table, attributed to the given labels.

<table>
<thead>
<tr>
<th>Gap Size (Label)</th>
<th>Range in mm (Min–Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero Gap</td>
<td>0.008–0.06</td>
</tr>
<tr>
<td>Gap 0.1</td>
<td>0.1–0.16</td>
</tr>
<tr>
<td>Gap 0.2</td>
<td>0.2–0.26</td>
</tr>
<tr>
<td>Gap 0.3</td>
<td>≥0.3</td>
</tr>
</tbody>
</table>

3.3. Acoustic Emission

In this section, the data analysis of the airborne acoustic signals is discussed. Figure 5 shows the amplitude spectrogram of a recorded acoustic signal from the weld process with and without gap formation, which was calculated using the Short-Time Fourier Transform (STFT). The sound pressure of the acoustic signals emitted during the weld process may vary in terms of amplitude and bandwidth, depending on whether there is a zero gap or there are different gap sizes.

Figure 5. A spectrogram of a Gap 0.3 mm weld image showing the amplitude variation from the Zero Gap to the Gap region, highlighted by white dashes.

An exploratory data analysis was performed on the entire dataset to gain a better understanding of the Zero Gap and Gap classes. The analysis focused on average signal strength and the distribution of data within each class. The Root Mean Square (RMS) values were calculated from the audio samples to measure the differences between the sound classes. This involved squaring each sample, averaging the squared samples, and taking
the square root of the result to obtain each RMS value. Additionally, the RMS values were converted to the decibel scale to make them more interpretable in physical terms.

The results showed that the RMS level of the signals varied based on the size of the gap between the metal sheets, as shown in Figure 6. Specifically, as the gap size increased, there was a decrease in the RMS level. One possible explanation for this trend is that the reduced contact area between the metal sheet and the laser beam in the gap regions led to a lower signal strength. However, accurately classifying the signals based on these RMS values posed a challenge due to the overlapping distributions observed between the Zero Gap and Gap classes.

Furthermore, the data distribution in terms of RMS level within each class can be observed on the y-axis. The Zero Gap class exhibited a distribution that was slightly skewed towards positive values. Gap 0.1 and Gap 0.2 followed a Gaussian distribution pattern. In contrast, Gap 0.3 showed a positive skew because a few data samples had lower RMS levels, likely due to larger gap sizes compared to the rest of the samples in that class. The median values of all Gap classes were similar to each other, but the Zero Gap class deviated from this pattern, as expected. This can be seen as a white dot in the center of the whisker in a violin plot.

![Figure 6. Violin plots of averaged RMS levels (dB): comparison between classes and data distribution within each class.](image)

3.4. Detection of Butt Joint Gaps

3.4.1. Experiment

Detecting and classifying the formation of gaps between metal sheets during the welding process using acoustics, specifically based on average signal strength, presents a considerable challenge due to the overlapping RMS levels between gap classes. To accurately identify and categorize the Zero Gap and Gap classes, it is necessary to employ neural networks in an automated detection and classification system. This system utilizes relevant time-frequency representations obtained through the STFT of the acoustic signals.
The experiments can be divided into two tiers. The first tier concentrates on detecting the Zero Gap and Gap. The second tier aims to classify Gap sizes to determine the level of weld degradation based on the gap formation size. These assessments allow for the effective evaluation of the informative nature of the segments without the need for the direct measurement of their joint gaps. Additionally, experiments were carried out to detect and classify data in both audible and ultrasonic frequency ranges. The following section offers a detailed explanation of the detection algorithm pipeline.

### 3.4.2. Detection Algorithm

This section discusses the details of the input data processing, neural network model design, and training scheme used to develop a classification model for the Zero Gap and Gap parameters. In the initial step of the welding-process audio data classification, we perform feature extraction by computing the magnitude spectrogram. This is achieved using the Short-Time Fourier Transform (STFT) with specific parameters: an FFT size of 512, a Hann window size of 512, and a hop size of 256 samples. The audio signal’s duration is 150 ms, sampled at 51,200 Hz, resulting in a spectrogram size of 29 × 257, representing 29 time frames and 257 frequency bins, covering frequencies up to 25.6 kHz in the audible range. For ultrasonic frequencies (up to 102.4 kHz), the spectrogram computation is based on a sampling rate of 204,800 Hz, FFT size of 2048, and hop size of 256 samples. The resulting spectrogram size is 113 × 1025, representing 113 time frames and 1025 frequency bins, maintaining a trade-off between frequency and time resolution. To reduce the dynamic range, a logarithmic magnitude scaling is applied to the spectrograms, and input features are then normalized to have zero mean and unit variance per frequency band. Additionally, the input features are normalized to have zero mean and unit variance per frequency band since the input signal’s amplitude and energy may vary significantly across different frequency ranges. By normalizing the features in each frequency band, we ensure that the model can focus on the relative variations in the spectrogram patterns rather than being influenced by the absolute magnitude of the signal at different frequencies. This normalization step enhances the model’s ability to discern meaningful patterns and improves its overall performance in classifying welding-process audio data [21]. These steps contribute to enhanced feature representation and prepare the data for further classification analysis.

In the subsequent step, the computed log-magnitude spectrograms are fed into a CNN architecture previously explored in our research. Although CNNs are commonly used for computer vision tasks like image classification, they also prove valuable in audio analysis due to their hierarchical structure. CNNs can learn complex representations of the input, capturing both low-level features that encompass simple and localized patterns and high-level abstractions that encompass complex and global patterns, including shapes, objects, or specific attributes associated with the input data. Moreover, CNNs exhibit translation invariance, allowing them to identify patterns in an audio signal regardless of their precise locations [29].

The CNN architecture used in this work consists of two convolutional blocks, each with two 3 × 3 convolutional layers. Following each convolutional layer, a Rectified Linear Unit (ReLU) activation layer, a max-pooling layer, and a dropout layer are applied. A flatten layer is used before the fully connected layer to provide a one-dimensional representation of the data. Finally, a Softmax layer serves as the classification layer to obtain the classification results, as shown in Figure 7.

The dataset was randomly shuffled to ensure a diverse representation of the data. It was then divided into three subsets: a training set, which contains 60% of the data and is used to train the model; a validation set, which consists of 20% of the data and is used to fine-tune the model and make decisions about its performance; and finally a test set, which also contains 20% of the data and is used to evaluate the trained model performance. To address the issue of class imbalance in the dataset, the SMOTE was employed [22]. This involved increasing the number of data samples in the training and validation
sets by interpolating feature vectors between existing minority class samples to match the count of the majority class file. Additionally, to ensure a robust evaluation of the model’s performance, a five-fold cross-validation strategy was employed. In this strategy, the dataset is divided into five subsets to ensure the robust evaluation of the model’s performance. Each subset is used as validation set while the remaining four subsets are combined to form the training set. Further, to improve the robustness of the classification model, mixup data augmentation is applied to the log spectrograms [17]. Mixup data augmentation creates synthetic training samples by blending pairs of training data points, which helps regularize the model and improve its ability to generalize to unseen data by fostering smooth interpolation between different classes. We train the CNN model using the Adam Optimizer with categorical cross-entropy loss for 300 epochs, with a learning rate of $10^{-3}$. Due to the limited size of the dataset, the model is susceptible to overfitting, which occurs when the model excessively tailors itself to the training data and fails to generalize well to new, unseen data. To mitigate this issue, the training process incorporates the early stopping technique with a patience parameter set to 50 epochs [30]. This implies that the training will terminate if the model’s performance on the validation dataset fails to exhibit improvement for 50 consecutive epochs. By employing this approach, the model is safeguarded against overfitting and gains the ability to generalize effectively to unseen data. Additionally, the choice of multiple dropout layers in the CNN architecture ensures robust regularization against overfitting, enhancing generalization by preventing neuron co-adaptation. The varied introduction of dropouts in limited labeled data scenarios improves noise resilience and real-world reliability. This aligns with the goal of balanced regularization and model complexity.

![CNN architecture diagram](image)

**Figure 7.** CNN architecture with two convolutional layers, rectified linear unit (ReLU) activation, dropout (D), flatten layer, fully connected (FC) layer, and final Softmax classification layer.

### 3.4.3. Detection Results

Initially, a binary classification task was conducted to differentiate between those data samples with gaps and those without gaps. This involved merging all gap classes (such as Gap 0.1, Gap 0.2, Gap 0.3) into a single category while considering the Zero Gap class as a separate category. When applied to the audible range sensor dataset, the resulting accuracy on the test set averaged at 98.9% per file for binary classification. Subsequently, a multiclass classification task was performed to achieve a more detailed
classification; see Figure 8. In this case, each Gap class was treated as a distinct category. The model achieved an average file-wise accuracy of 86.8%.

For the ultrasonic range sensor dataset, the resulting accuracy on the test set averaged 98.6% per file for binary classification, which was similar to the performance observed in the audible range dataset. Subsequently, a multiclass classification task was carried out to achieve a more granular classification. The model achieved an average file-wise accuracy of 91.4%, indicating a significant improvement (4.6% accuracy) compared to the audible range dataset; see Figure 9. It is worth noting that the observed improvement in performance in the multiclass classification of the ultrasonic range dataset could potentially be attributed to factors such as the broader frequency band and the various settings explored for feature extraction, including FFT size, overlapping window, or the positioning of the microphone during measurements. The confusion matrices depicted in Figures 8 and 9 illustrate noticeable misclassifications between Gap 0.2 and Gap 0.3 during the multiclass classification process. This misclassification could potentially be attributed to both the minimal disparity in the acoustic emission characteristics between the Gap 0.2 and Gap 0.3 classes as well as the limited number of data samples available in the test set. It is highly probable that by augmenting the training data with additional samples, providing accurate annotations, and further optimizing the model, significant improvements can be achieved in the classification results. These improvements would be evident through the analysis of a substantial amount of test data.

Figure 8. Confusion matrices for binary (left) and multiclass (right) classifications (results of audible range data (sE8)).
Figure 9. Confusion matrices for binary (left) and multiclass (right) classification (results of ultrasonic range data (MK301)).

While the results presented in this paper appear promising, it is important to note that the classification outcomes are confined to the specific dataset utilized. As a result, this model may not be readily applicable or generalize well to real-world applications. However, by fine-tuning the feature and model parameters, it is possible to obtain more reliable and accurate results.

4. Conclusions

The laser beam butt welding of metal sheets made of high-alloy steel is a technically challenging task, mainly due to the formation of joint gaps that can cause weld defects such as a lack of fusion, undercuts, or seam sagging. In this study, we delved into the potential of using airborne acoustic emission in laser beam butt welding, specifically focusing on automating the classification of joint gap formations through neural networks. Controlled welding experiments were conducted to simulate various butt joint gap sizes, with captured acoustic signals ranging from audible to ultrasonic frequencies. Employing a method based on short-time Fourier transformation, we successfully extracted relevant audio features, which were then applied in a convolutional neural network architecture, augmented with data augmentation techniques. The results showcased the high effectiveness of this non-destructive and non-invasive approach in identifying joint gap formations, achieving an impressive accuracy rate of 98%.

This research addresses a significant gap in the field of laser beam butt welding, particularly in the application of airborne acoustic analysis across audible-to-ultrasonic ranges. By applying machine learning techniques, we harnessed acoustic information for gap formation detection, advancing this specialized welding application. Our study stands out for its meticulous exploration of joint gap formation, seamlessly integrating acoustic emission measurements with high-speed cameras. The introduced automated classification system represents a significant step towards enhancing the reliability of monitoring methods. It not only aids in tracking gap sizes but also enables the timely detection of potential seam defects. Additionally, the classification accuracy for various gap sizes reached up to 90%, providing valuable insights for characterizing and categorizing joint gaps accurately.

While our findings are promising, it is crucial to acknowledge the limitations of our dataset, which may restrict the generalizability of the results. Further extensive
experimental and analytical investigations are warranted. These efforts will expand the dataset and pave the way for the practical application of our findings in real-world industrial settings, accounting for acoustic interactions with common disturbances encountered, such as cross jet phenomena.


**Funding:** This research was part of “Leistungszentrum InSignA”, funded by the Thuringian Ministry of Economics, Science and Digital Society (TMWWDG), grant number 2021 FGI 0010 and by the Fraunhofer-Gesellschaft, grant numbers 10-05014 and 40-04115.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to acknowledge the support received from the Fraunhofer Internal Programs under Grant No. Attract 025-601128. Additionally, we extend our gratitude to the Fraunhofer IZFP and IDMT projects for their cross-support in facilitating the completion of this research.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.