

**Anomaly Detection in WAAM Deposition of Nickel Alloys
Single-Material and Cross-Material Analysis**

Rajesh, A.; Ya, Wei; Hermans, M.J.M.

DOI

[10.3390/met13111820](https://doi.org/10.3390/met13111820)

Publication date

2023

Document Version

Final published version

Published in

Metals

Citation (APA)

Rajesh, A., Ya, W., & Hermans, M. J. M. (2023). Anomaly Detection in WAAM Deposition of Nickel Alloys: Single-Material and Cross-Material Analysis . *Metals*, 13(11), Article 1820.
<https://doi.org/10.3390/met13111820>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Article

Anomaly Detection in WAAM Deposition of Nickel Alloys—Single-Material and Cross-Material Analysis

Aditya Rajesh ^{1,*}, Wei Ya ² and Marcel Hermans ¹

¹ Department of Materials Science and Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands

² RAMLAB BV, Scheepsbouwweg 8, 3089 JW Rotterdam, The Netherlands

* Correspondence: aditya@ramlab.com

Abstract: The current research work investigates the possibility of using machine learning models to deduce the relationship between WAAM (wire arc additive manufacturing) sensor responses and defect presence in the printed part. The work specifically focuses on three materials from the nickel alloy family (Inconel 718, Invar 36 and Inconel 625) and uses three sensor responses for data analysis, which are welding voltage, welding current and welding audio. Two different machine learning models are used—artificial neural networks (ANNs) and random forests (RF). The results for each of the materials, separately, indicate that the accuracies range from 60% to 90% and the correlation coefficient is less than 0.5 (indicating weak positive correlation), depending on the model and material. In addition to separate material analysis, a cross-material data analysis was formed to test the models' general prediction capabilities. This led to predictions that are significantly worse, with accuracies ranging from 20% to 27% and very weak correlation coefficients (less than 0.1), indicating that the choice of material is still important as a boundary condition. Analysis of the results indicates that the relative importance of audio sensor response depends on the nature of defect formation. Random forests are found to perform the best for single material analysis, with the comparatively inferior performance of ANNs possibly being due to lack of sufficient datapoints.

Keywords: WAAM; anomaly detection; machine learning; cross-material prediction; nickel alloys



Citation: Rajesh, A.; Ya, W.; Hermans, M. Anomaly Detection in WAAM Deposition of Nickel Alloys—Single-Material and Cross-Material Analysis. *Metals* **2023**, *13*, 1820. <https://doi.org/10.3390/met13111820>

Academic Editor: Chris Aldrich

Received: 22 August 2023

Revised: 24 October 2023

Accepted: 26 October 2023

Published: 28 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The development of different techniques within additive manufacturing, which is a manufacturing paradigm based around material addition of successive layers to form a complete product, has been the result of an increasing demand for complex geometries in various engineering fields such as aerospace and energy applications [1]. Additive manufacturing techniques feature a variety of advantages over more conventional “subtractive” techniques based on machining, which include the capability of depositing complex geometries, heterogenous composition structures and rapidly prototyping completely different components. While initially confined to prototyping research applications, additive manufacturing has since seen significant growth in mass manufacturing avenues in recent years [2].

Within the wide variety of AM technologies, wire and arc additive manufacturing (WAAM) makes use of an electric arc as a heat source, while a wire is added to deposit material in a layer-by-layer fashion. As is the case in multi-pass welding, different types of defects may occur, including inhomogeneous bead shapes, lack of fusion, pore formation, hot and cold cracks and inclusions. Each of these defects can be related to process parameter selection, disturbances in the process or the response of the specific material to the heat introduced in combination with the stresses induced. Defects can be detected by non-destructive testing techniques [3]. The question then arises of whether fingerprints of defect formation can be found in signals captured during the deposition process, allowing for real-time monitoring and quality control.

The concept of a multi-sensor monitoring system in welding is not a very recent approach, with Bhattacharya et al. [4] having combined, in 2011, acoustic and electric signals (specifically sound kurtosis, welding current and arc voltage) to monitor the pulsed GMAW welding of mild steel plates using ESAB S-6 filler wire. Kurtosis is a measure of signal peakedness (the tailedness of the distribution), with high values indicating a significant amount of outliers, and vice versa [5]. It can be a useful tool to extract the information of outliers from raw audio data, making computations more efficient. The study of Bhattacharya et al. [4] was focused on welding-deposition efficiency rather than specific defect detection, but there is a connection between deposition efficiency and certain classes of defects and imperfections, such as spatter, making this a valued reference for further research. Alfaro et al. [6] studied the combination of acoustic and optical infrared signals along with welding current and voltage in short-circuit GMAW (GMAW-S) performed on AISI 1020 steel plates using ER70S-6 filler wire. Their study showed that the usage of data fusion techniques (specifically Kalman filters) gave new parameters that could detect anomalies during the welding process better than single sensor parameters.

A review of machine learning usage in additive manufacturing conducted by Qin et al. [7] showed that deep learning models (which include all the variants of neural networks) were the most commonly employed models for processing a wide variety of data, including images, acoustic data and thermal data. Support vector machines were found to be fairly commonly used, whereas Gaussian processes were rarely used. It is interesting to note that the majority of studies referenced in the review used powder bed-based techniques as opposed to wire-based techniques. They also noted that deep learning models typically require large amounts of data to be useful, making them difficult to utilize in situations where data collection is difficult or time-consuming. Additionally, the paper pointed out that many of these models are “black boxes”, which are not physics-based. Ko et al. [8] constructed physics-based models using graph-based networks for laser powder bed fusion processes, indicating the possibility of a knowledge-based approach to investigate any potential accuracy improvements. He et al. [9] also corroborate the popularity of neural networks and support vector machines among the research works published in the field of wire arc additive manufacturing. They noted that support vector machines can be trained on small datasets to obtain reasonably good prediction accuracies, giving a solution for dataset availability problems. Denoising is especially important since there is the potential for loss of important patterns in the removed “noise”.

The authors of this work noted that the majority of the literature available on the monitoring and control of additive manufacturing processes focused on laser, powder-based additive manufacturing. While there have been publications based on TIG, wire-based WAAM processes, limited publications are available on the GMAW-based WAAM. Thus, the current work examines sensor responses (current and voltage waveforms, audio signals) from GMAW-WAAM deposition using neural networks and random forests to evaluate the quality of defect detection. In addition, it was interesting to consider the possibility of generalising defect feature detection in sensor signals across different materials. Three materials from the nickel alloy family were chosen for cross-material analysis. The sensor responses of all three materials were evaluated, and the models (neural networks and random forests) were tasked with finding defects on materials on which they had not been trained. This would give insight into the potential features that could be material independent, which could be isolated for building a more robust machine learning model.

2. Experimental Materials and Methods

All printing experiments were conducted at RAMLAB BV using a Panasonic TM1400 WGIII Robot (Panasonic, Japan, integrated by Valk Welding BV, NL). The robot has current and voltage sensors equipped (RAMLAB BV inhouse built, NL) that are capable of logging current and voltage signals at a sampling rate of up to 50 kHz. For this study, a sampling rate of 25 kHz was used. For the recording of audio signals, a Devine M-Mic USB BK microphone (Devine, Goes, NL) was attached to the welding arm. An attachment in

this fashion would ensure zero relative velocity between the microphone and weld torch, eliminating the Doppler effect, as well as maintaining constant distance from the welding arc. The microphone is capable of a sampling rate of up to 48 kHz. In this study, a sampling rate of 44.1 kHz is used. The attachment of the microphone is shown in Figure 1.



Figure 1. Attachment of microphone to welding arm.

Three materials are investigated in this study—Inconel 718, Inconel 625 and Invar 36. All consumables were 1.2-mm welding wires provided by Voestalpine Böhler Welding. The Alumaxx[®] Plus (30% He in Ar) shielding gas provided by Air Products[®] was used in all experiments. The material compositions are listed in Table 1 [10–12].

Table 1. Chemical compositions of Inconel 718, Inconel 625 and Invar 36 in weight percentages.

| Element | Fe | Ni | Cr | Nb | Mo | Ti | Co | Mn | Si |
|-------------|----|-------|-------|-----------|-----------|-----------|----|-------|-------|
| Inconel 718 | 17 | 50–55 | 17–21 | 4.75–5.50 | 2.80–3.30 | 0.65–1.15 | ≤1 | ≤0.35 | ≤0.53 |
| Inconel 625 | ≤5 | ≥58 | 20–23 | 3.15–4.15 | 8–10 | ≤0.4 | ≤1 | ≤0.5 | ≤0.5 |
| Invar 36 | 63 | 36 | - | - | - | - | - | 0.35 | 0.20 |

All the prints conducted in this research work were deposited using the super active weld process (S-AWP) mode, which is a process developed by Panasonic using short circuiting material transfer to stabilize the arc, resulting in minimal spatter [13]. The waveform, together with the back-and-forth wire movement, allows material to transfer through a stable short arc. A typical SAWP waveform of the voltage is as shown in Figure 2.

Experiments were performed with different geometries in order to ensure the inclusion of a variety of different deposition conditions in the training dataset. The different geometries have been categorized as shown in Figure 3.

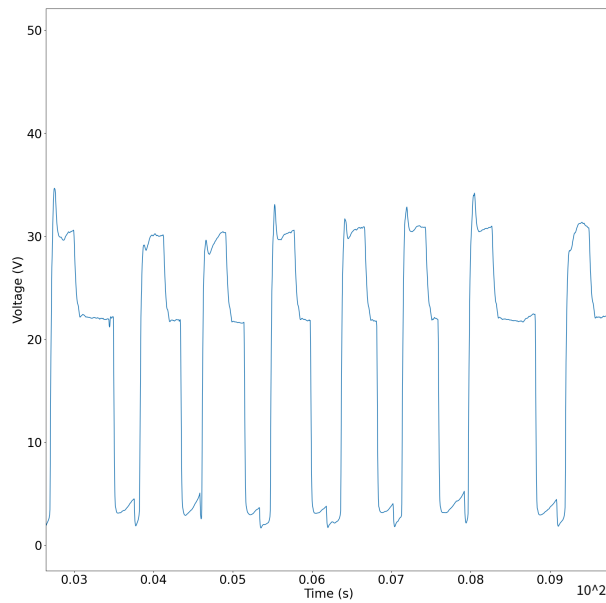
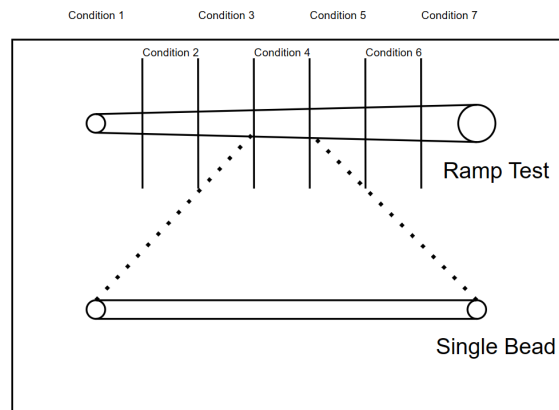
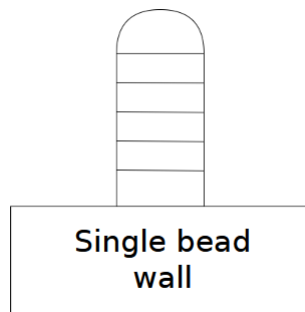


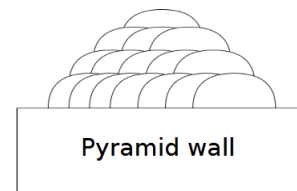
Figure 2. S-AWP voltage waveform during Invar 36 deposition.



(a) Tests of Ramp and Single Bead.



(b) Single Bead Wall



(c) Pyramid Block

Figure 3. Print geometries used in current research work—(a) Ramp tests, which are typically used for quick navigation of desired process parameters [14] and single beads deposited using selected parameters from ramp tests for extracting referenced clean data; (b) Walls made from successive deposition of single beads in alternating directions; (c) Pyramid blocks made from successive layers of reducing numbers of beads per layer.

The number of beads categorized by type of experiment performed for each material has been compiled in Table 2. The data for Inconel 625 are lower in quantity compared

to the other two materials since they were pulled from another experiment for which optimized process parameters had already been extracted and were used primarily for cross-material verification purposes.

Table 2. Bead types of Inconel 718, Invar 36 and Inconel 625.

| Sl. No. | Test Type | Inconel 718 | Invar 36 | Inconel 625 |
|---------|----------------------|-------------|----------|-------------|
| 1 | Single Beads | 15 | 9 | 6 |
| 2 | Single Bead Walls | 0 | 136 | 0 |
| 3 | Block Walls/Pyramids | 166 | 0 | 16 |
| | Total | 181 | 145 | 22 |

3. Feature Extraction, Data Labeling and Model Architecture

3.1. Feature Extraction and Data Labeling

3.1.1. High Frequency Voltage/Current Feature Extraction

To determine the features that can best describe each bead, the individual waveforms of each deposition process were examined. Figure 4 shows an example of a deposited Invar 36 bead with the deposition direction marked with a black arrow. The beads marked with red boxes were where material overflow occurred. Figure 5a shows the voltage waveform of the region where overflow occurred and Figure 5b shows the region where there is no overflow. The differences between Figure 5a and Figure 5b are the irregular secondary peaks that appeared on the voltage waveform, as highlighted within the dashed line box in Figure 5.

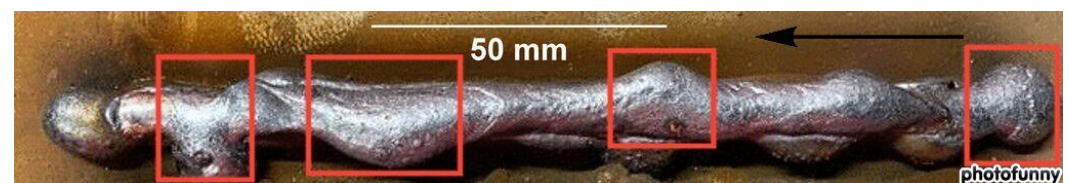


Figure 4. Invar 36 bead with visible overflow.

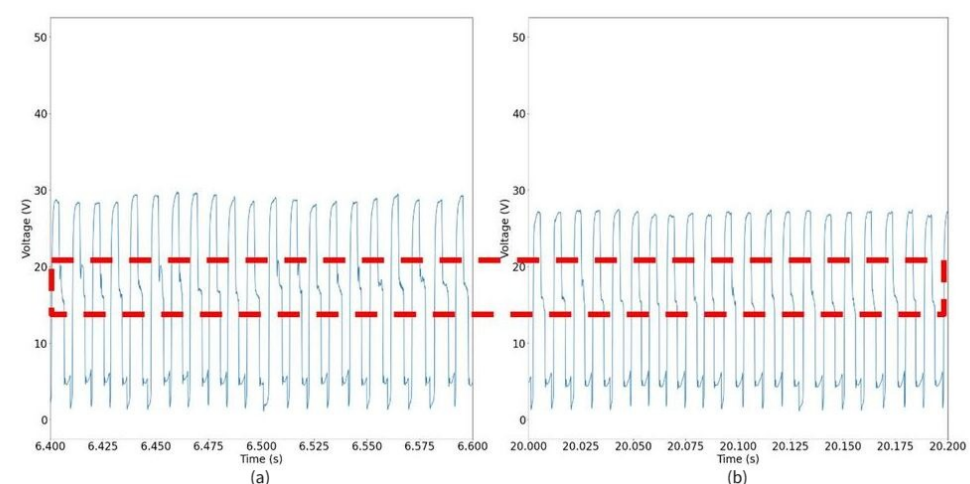


Figure 5. Voltage waveforms from two different sections of deposited Invar 36 bead—(a) Overflow region; (b) No overflow region.

Applying this approach to the full waveform of the deposited bead, as shown in Figure 3, the computed number of excessive ΔV (difference between successive voltage datapoints) instances ($n_{\Delta V}$) is plotted in Figure 6a, which corresponds to the visual observations of the overflow and non-overflow regions. The regions where the plot drops to 0

are the non-overflow regions, whereas the overflow regions show that the ΔV is high. To generalize this approach for potential applications where the process parameter changes or other materials are processed, the data plotted in Figure 6a was normalized by subtracting the mean of the plot from each point, resulting in Figure 6b.

In a 2D (xy coordinate) analysis, the usage of $n_{\Delta V}$ can provide height-based information of voltage. The width of each pulse may have a weak correlation with $n_{\Delta V}$. Hence, it is important to have a dedicated width parameter (feature) to ensure better correlation between the sensor response and anomalies. The calculation of waveform width is a fairly straightforward method, as each waveform represents a material transfer cycle. A threshold value is defined as the midpoint of a waveform crest and trough (akin to calculating FWHM, or full width at half maximum). Counting the number of points between consecutive crossings of such a threshold can provide the width of each waveform. It is possible to normalize the obtained point number using sample rate, thereby giving pulse width in time units. Once the plot of widths is obtained, the population variance is extracted to obtain a one-number width parameter. The same processing methodology is used for welding-current waveforms.

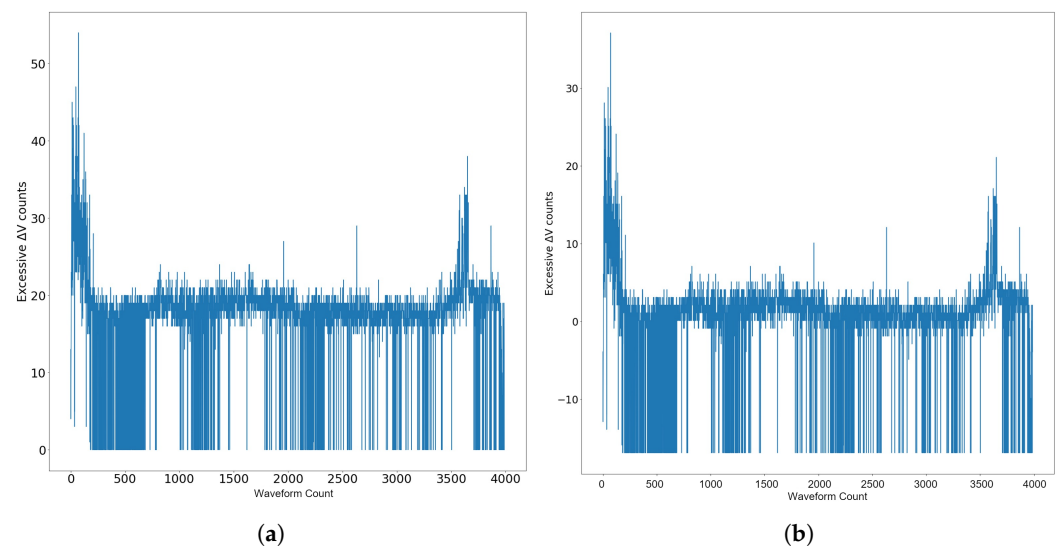


Figure 6. Calculated and normalized voltage parameters from deposition of Invar 36 bead in Figure 4—(a) Calculated $n_{\Delta V}$ of Invar 36 bead; (b) Normalized $n_{\Delta V}$ of Invar 36 bead.

3.1.2. Audio Feature Extraction

A variety of research works indicate the utility of acoustic signal analysis in anomaly detection during WAAM deposition [15–17]. Defects that are linked to arc instabilities, including porosity, spatter and lack-of-fusion, typically show distinct audio signatures. Arc sound is a reflection of the deposition process in the form of audio signals, including arc ignition, burning, extinguishing and others, among which anomalous sounds can be only distinguished by experienced welders. Therefore, the audio signal can be used for evaluating the arc stability, as many defects can be induced due to an unstable arc. However, the raw audio data, as shown in Figure 7a, does not normally provide any useful information because noise, such as surrounding sounds, machine sounds, etc. is also included.

Spectral gating is used to denoise the signal, which splits the spectrogram of the audio data into frequency bands and applies noise thresholds to each band [18]. Bands that are lower than the threshold are filtered, which essentially removes noise from the signal, as shown in Figure 7b, processed using Python 3.6 using a threshold value of 1 according to the scale used by the noisereducer library. The audio data processed in Figure 7 are the recording during the deposition of the Invar 36 bead, as shown in Figure 4.

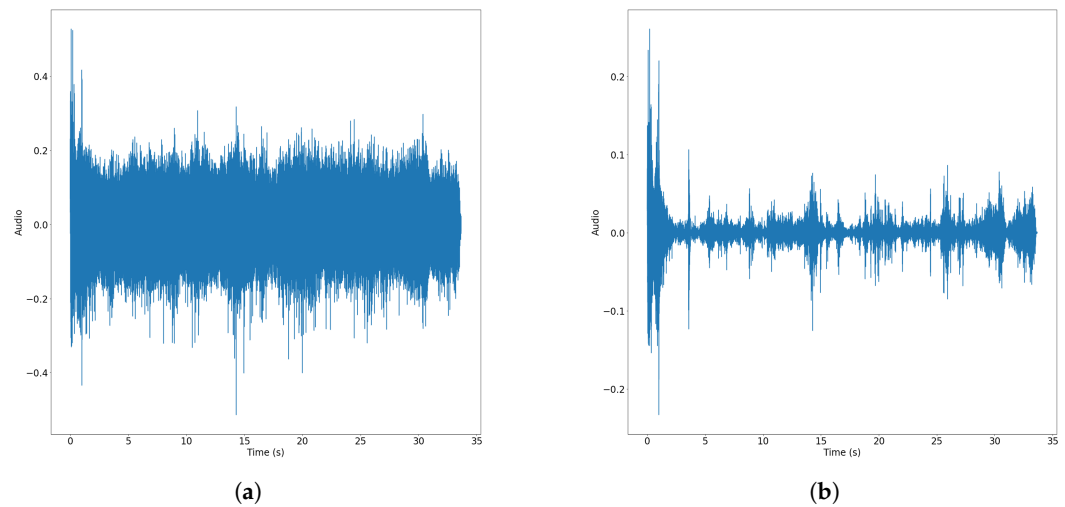


Figure 7. Raw and filtered audio signals from deposition of Invar 36 bead—(a) Raw audio data; (b) Spectral gating-denoised audio signal with threshold = 1.

The audio data can be further processed in the Mel spectrogram, as shown in Figure 8 [19]. Mel bands delineate frequencies based on the ease of distinguishing them using the human ear, thereby giving a more accurate picture of how a human ear would evaluate the audio signal. As shown in Figure 8, the variations are best seen in the bands between 3072 Hz and 8192 Hz, which corresponds to the overflow regions of the deposited Invar 36 bead (Figure 4). This indicates that this particular frequency band should be isolated from the spectrogram and analyzed on its own. The average of this frequency band is taken and the variance of the resulting set of values is taken to be the first audio parameter in this study (termed as spectral variance). In addition, audio kurtosis was used for audio analysis. Kurtosis is a measure of the “peakedness” of data and can be used for identifying anomalies [5]. Bhattacharya et al. [4] used audio kurtosis in combination with current and voltage data to predict weld deposition efficiency. In Python 3.6, audio kurtosis can be calculated using the scipy module formulated as shown in Equation (1), where n is the population count, μ is the population mean and σ is the population standard deviation.

$$k = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \mu)^4}{\sigma^4} \quad (1)$$

All the features obtained from the above data analysis are summarized in Table 3, which will be used in the ML models in subsequent sections.

3.1.3. Data Labeling

Dataset labeling for supervised machine learning is performed on a binary basis, using 0 to indicate clean beads (free from anomaly) and 1 to indicate defective beads. Inconel 718 is expected to show hot cracking presence, defined according to ASME standards [20]. An example of hot cracking in Inconel 718 is shown in Figure 9. The sample was prepared using standard metallographic procedures and etched using Adler’s reagent. Micrographs were obtained using a Leica LMD7 Optical Microscope (Germany) at 100× magnification. This research work primarily focused on hot cracking and bead overflow, since these were the most prominent defects observed in the analyzed materials.

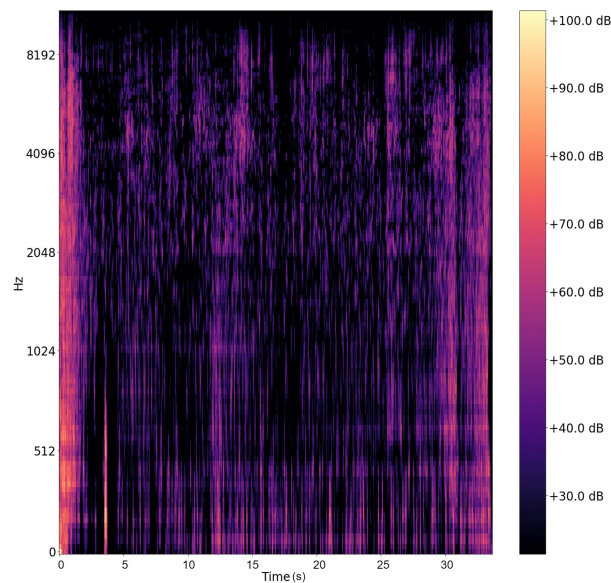


Figure 8. Mel spectrogram of filtered audio signal of Invar 36 bead, as shown in Figure 4 with threshold = 1.

Table 3. Features extracted from collected data for use in ML modeling.

| Parameter | Symbol | Description |
|-----------------------------|--------|---|
| Voltage peak count variance | V_1 | The variance of the number of peaks per voltage pulse |
| Voltage peak width variance | V_2 | The variance of the voltage pulse widths |
| Current peak count variance | I_1 | The variance of the number of peaks per current pulse |
| Current peak width variance | I_2 | The variance of the current pulse widths |
| Audio kurtosis | A_1 | Kurtosis of filtered audio signal |
| Audio spectral variance | A_2 | Variance of 3072 Hz–8192 Hz frequency band of audio Mel spectrogram |

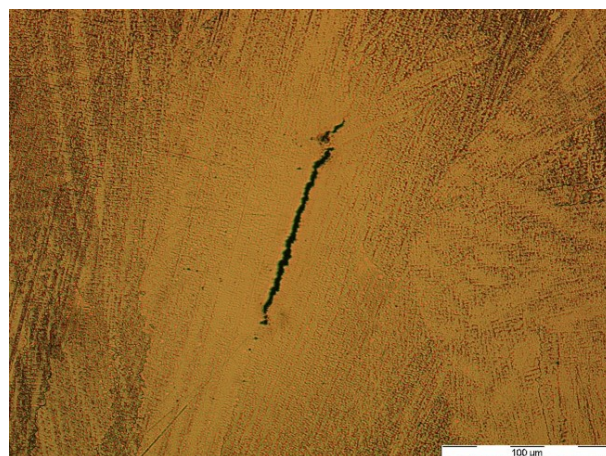


Figure 9. A micrograph of Inconel 718 showing hot cracking (liquation cracking).

In the case of Invar 36, the primary anomaly was bead overflow, which is not well-defined in existing standards. Based on experimental observations and bead shape that could be favorable for additive manufacturing, we have defined the following:

Definition 1. An instance of bead overflow is considered to be unacceptable once the bead width deviates from the expected value by more than 40%, or the bead runs off due to excessive local overheating.

An example of Invar 36 bead-overflow labeling is shown in Figure 10, where the normal expected bead width is 8.72 mm deposited using welding voltage = 15.6 V, wire feed speed = 8 m/min and travel speed = 0.25 m/min. The overflow occurred when the bead width became larger the 40% (marked in Figure 10 as 2 and 3), which corresponds to 56% (4.88/8.72) and 42% (3.63/8.72), respectively.

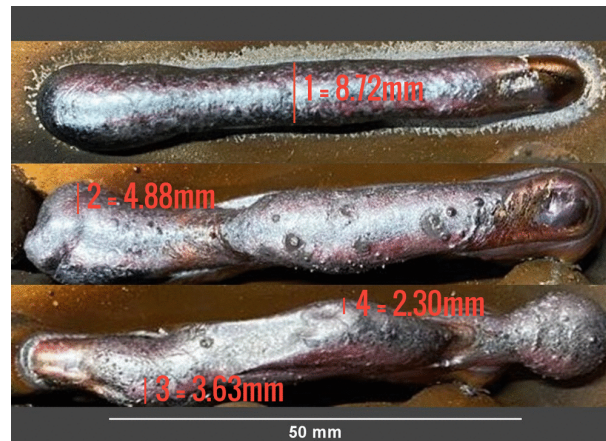


Figure 10. Unacceptable (2 and 3) and acceptable (4) overflow seen alongside normal bead width (1) of deposited Invar 36 beads.

3.2. Model Architecture

Two kinds of machine learning models were used in this study, namely, random forests and fully-connected artificial neural networks. Features extracted from the collected data, as detailed in the previous subsection, were used as inputs for the models. The used models and their parameters are detailed below.

3.2.1. Random Forests

A random forest is a collection of tree-based decision makers that classifies the input dataset based on a series of decisions made at each node [21]. Each tree runs its own analysis, and the collection of trees eventually formulate a decision. Random forests can be extremely useful for identifying the relative importance of features, since the decisions made at each node in each tree depends on the input features. Python 3.6 offers the functionality of calculating the relative importance of each feature based on the mean decrease in impurity (or Gini impurity). Impurity describes the extent of division of classes at each node based on the “question” asked by the node. Low impurity means there is a good extent of division of classes (i.e., the parameter that determines the question asked by the node is of high importance in classifying the input data). The mathematical formulation of Gini impurity at a particular node τ is shown in Equation (2) [22].

$$GI(\tau) = 1 - \sum_{i=1}^k p_i^2 \quad (2)$$

$$p_i = \frac{n_i}{n} \quad (3)$$

In Equation (3), n_i represents the number of datapoints belonging to class i , whereas n represents the total number of datapoints at the node. Perfect classification into one class would result in zero impurity. The hyperparameters used for the random forest analysis are shown in Table 4.

Table 4. Random Forest model hyperparameters.

| Hyperparameter | Value |
|--------------------------------|---------|
| Number of trees (n_estimators) | 1000 |
| Maximum tree depth (max_depth) | 8 |
| Criterion (criterion) | Entropy |
| Number of features at node | 2 |

The implementation of random forests in Python 3.6 is performed using the scikit-learn package (specifically, the RandomForestClassifier object). In this package, the default number of features considered at a node for splitting is the square root of the number of features, i.e., the square root of 6, which rounds down to 2. The number of trees was increased to 1000 in order to get better statistical results for the classification, and the tree depth is constrained to limit the complexity of the fitting criteria. Both of these lead to reduction in overfitting.

3.2.2. Artificial Neural Networks

Artificial neural networks are designed on the basis of the human brain and use nodes (or neurons) to construct a function that best defines the input data distribution [23]. Based on the literature, support vector machines and neural networks are the most commonly used models. In this research work, neural networks are chosen due to their high customizability (in the form of a large number of hyperparameters) and suitability for big data [24]. Multilayer perceptrons (MLPs) were chosen for this research work since the nature of the data did not necessitate the usage of RNNs or CNNs. In this paper, the terms MLP and ANN will be used interchangeably. The implementation of ANNs in Python 3.6 is performed using the Keras package (specifically, the KerasRegressor object). The model architecture used in the current work is shown in Figure 11. The hyperparameters used for the neural network analysis are shown in Table 5.

The number of hidden layers was chosen based on different trial runs. Above four layers, the model was found to drop in performance due to overfitting, whereas below four layers, the objective function had insufficient complexity to fit the input data. Binary cross entropy loss was chosen as the loss criterion and the sigmoid function was used at the output layer due to the binary nature of classification being performed. The optimizer learning rate was also chosen based on multiple trials. Higher learning rates were found to cause reductions in performance due to the weight corrections being large and skipping over possible minimas in the objective function, whereas lower learning rates caused the convergence to be too slow.

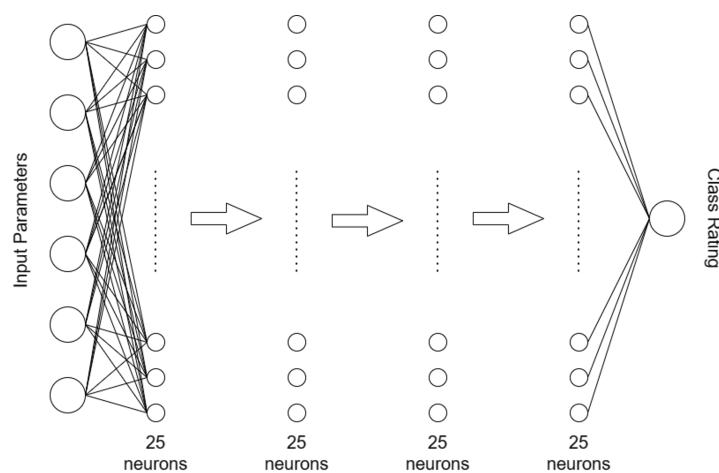
**Figure 11.** MLP architecture used for bead data analysis.

Table 5. MLP model hyperparameters.

| Hyperparameter | Value |
|-----------------------------|--|
| Number of hidden layers | 4 |
| Number of neurons per layer | 25 |
| Activation function | ReLU (hidden layers), Sigmoid (output layer) |
| Loss | Binary cross entropy |
| Optimizer | Adam |
| Optimizer learning rate | 0.0004 |
| Number of epochs | 450 |

4. Results and Discussion

4.1. Class Balance Evaluation

It is useful to evaluate the class composition, i.e., the number of samples which are clean versus defective. Class balance (or imbalance, thereof) can have an impact on the performance of machine learning models, and thus it is important to conduct this evaluation before the actual learning process. An imbalance ratio has been calculated for each material, defined as the ratio of the number of samples in the majority class (class with higher number of samples) to the number of samples in the minority class (class with lower number of samples). This ratio, along with the number of samples in each class for all the materials in this study, are presented in Table 6.

Table 6. Class composition of all materials.

| Material | Inconel 718 | Inconel 625 | Invar 36 |
|-----------------|-------------|-------------|----------|
| Clean | 30 | 22 | 121 |
| Defective | 151 | 0 | 24 |
| Imbalance Ratio | 5.033 | ∞ | 5.042 |

The highest extent of imbalance is in Inconel 625, since the defective class is completely absent. In order to evaluate the extent of imbalance in the remaining two materials, it is necessary to consult the literature to identify the limit of imbalance ratio beyond which prediction performance is affected. Buda [25] conducted a study on different types of convolutional neural networks, and the results indicated that at an imbalance ratio close to 15, the performance of some of the CNNs would drop by approximately 5% only. The others required much higher imbalance ratios to show similar performance drops. This indicates that the extent of imbalance in Inconel 718 and Invar 36 is small enough to have minimal effects on the performance of ML models. Another literature survey conducted by Leevy et al. [26] suggested that high imbalance ratios generally lie between 100:1 and 10,000:1, indicating that the ratios in Table 6 do not qualify as extreme in nature (except Inconel 625). Johnson et al. [27] surveyed the usage of deep neural networks to analyze class imbalanced data and found that the actual number of samples in the minority class is an important factor in determining the quality of prediction. To take class imbalance into account, tests were conducted with the training and testing sets for different ML models, being chosen such that the minority class would be sufficiently represented in the testing set. The idea behind this approach was to ensure that the model correctly learns the differences between clean and defective samples in spite of the imbalance.

Another matter of importance is the metrics that are used for evaluation of the predictions. Elmrabbit et al. [28] analyzed multiple ML classification models to evaluate various instances of network behavior to identify cyberattacks. They used seven metrics to analyze the models, including accuracy, precision, recall, false positive rate, F1 score and area under the receiver operating characteristic curve (ROC-AUC). A confusion matrix was also used

to evaluate the true and false positives and negatives. Earlier research regarding time series anomaly detection by Aminikhanghahi et al. [29] detailed the aforementioned parameters and mentioned that they can be useful in classification-related problems. Chicco et al. [30] interestingly noted that another metric called the Matthew’s correlation coefficient (MCC) could be more effective than F1 score and accuracy when evaluating binary classification problems. They conducted tests on highly class-imbalanced datasets and found that the MCC score could better assess the poor quality of prediction compared to accuracy and F1 score (both of which were erroneously high in some cases). Considering that the current research work also involves situations with class imbalance (specifically concerning Inconel 625 due to the large amount of clean data versus no defective data), it is useful to consider MCC as an evaluation metric.

In addition, scaling was performed on all input parameter vectors to ensure an equal weightage of every parameter in the model. Scaling was performed using the Standard-Scaler function from scikit-learn in Python 3.6, which scales the data such that the mean is 0 and the variance is 1 (zero mean unit variance). For the generation of training and testing sets from the total dataset, a test set fraction of 0.25 was used (i.e., 75% of the data would be used in training and 25% of the data would be unseen by the model and used for testing). In addition, for the neural network, the training process involved the formation of a validation set with a fraction of 0.1 (i.e., 10% of the training data would be used for validation). Using the information from Table 6, it was decided to include 66% (or two thirds) of the minority class dataset in the testing set, and the remaining in the training set. The split was performed using the `train_test_split` function from scikit-learn (Python 3.6).

4.2. Random Forest Classification

The error metrics obtained when running this model on the data of Inconel 718 and Invar 36 are shown in Table 7. Inconel 625 is omitted due to the lack of defective beads.

Table 7. Random Forest results on Inconel 718 and Invar 36.

| Evaluation Metric | Inconel 718 | Invar 36 |
|---|-------------|----------|
| Accuracy | 0.911 | 0.611 |
| Precision | 0.862 | 1.000 |
| Recall | 1.000 | 0.125 |
| False Positive Rate (FPR) | 0.200 | 0.000 |
| F1 | 0.926 | 0.222 |
| Matthew’s Correlation Coefficient (MCC) | 0.830 | 0.271 |

The results of Inconel 718 show a high MCC, indicating very strong positive correlation between the prediction and ground truth (the assigned data labels). The accuracy, precision, recall and F1 are also very high. The false positive rate is 20%. Ideally, an FPR of less than 10% should be targeted at the very least. The most important result, however, is that even with a low number of minority class datapoints to train on, the model was able to correctly classify minority datapoints for the most part. The results of Invar 36 clearly indicate the lack of ability to detect true positives (defects) correctly, since the recall score and F1 score are very low. The MCC is also less than 0.5, indicating a weak correlation between the ground truth and predictions.

The feature importances for the random forest model in case of both Inconel 718 and Invar 36 are shown in Figure 12a,b.

4.3. Artificial Neural Network Regression and Classification

While the random forest approach was of classification nature, the ANN was configured to run as a regression. The reasoning behind this was that modeling the differences

between sensor responses and defect presence as a continuous function introduced the possibility of deducing defect severity or extent of presence using the value of class rating predicted by the model. The caveat to such an approach, however, would be that more datapoints which sufficiently represent varying levels of defect severity would be needed for finetuning the model. Apart from the classical regression approach, another possibility would be to convert the predictions into a classification result by defining prediction windows for each class. The easiest way to do this would be to apply a threshold of 0.5 (midway between 0 and 1) for the regression class labels. Both regression and classification results were examined.

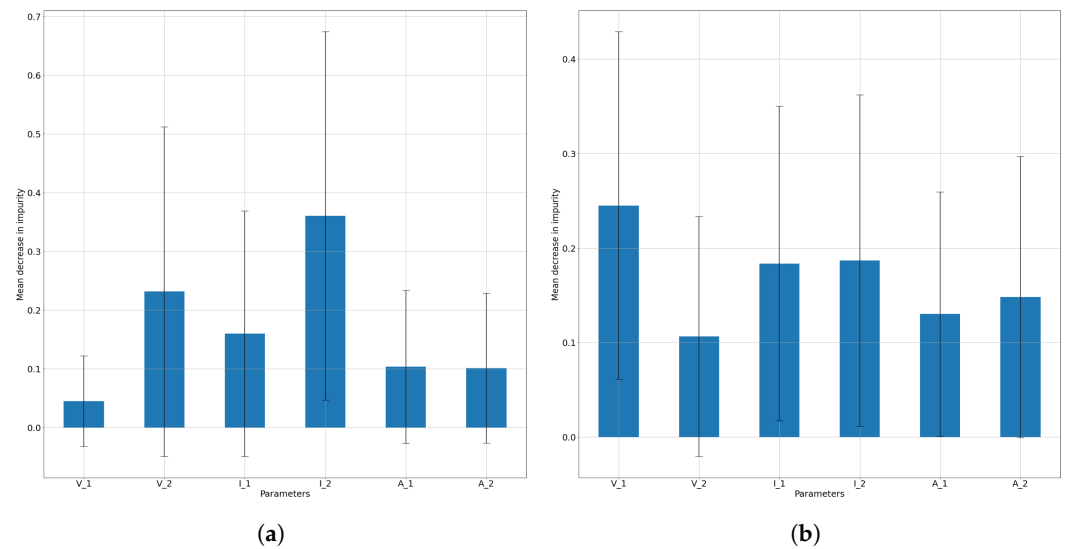


Figure 12. Relative feature importances of RF model for (a) Inconel 718 and (b) Invar 36.

Obtaining confidence intervals for predictions with neural networks is not straightforward due to its “black-box” nature. The usual methodology for constructing such intervals is bootstrapping, where multiple training sets are constructed from the total available dataset with replacements (i.e., some elements in the training set could be repeated), and the model is run on all these individual training sets to obtain multiple predictions [31]. The mean and variance of these predictions are used to make confidence intervals (akin to a brute-force approach) [32]. In this work, bootstrapping was used to generate confidence intervals for statistical analysis (1000 sets).

The regression error metrics obtained when running this model on the data of Inconel 718 and Invar 36 are shown in Table 8. Inconel 625 is omitted due to the lack of defective beads. Applying thresholding on the regression results gives “classification” results from which metrics can be calculated for both materials. These metrics are presented in Table 9.

Table 8. ANN regression results on Inconel 718 and Invar 36.

| Evaluation Metric | Inconel 718 | Invar 36 |
|--------------------------------|-------------------------------|-------------------------------|
| Mean Squared Error (MSE) | 0.086 (95% CI {0.084, 0.089}) | 0.358 (95% CI {0.355, 0.362}) |
| Root Mean Squared Error (RMSE) | 0.286 (95% CI {0.282, 0.290}) | 0.597 (95% CI {0.595, 0.600}) |
| Mean Absolute Error (MAE) | 0.092 (95% CI {0.089, 0.095}) | 0.368 (95% CI {0.365, 0.371}) |

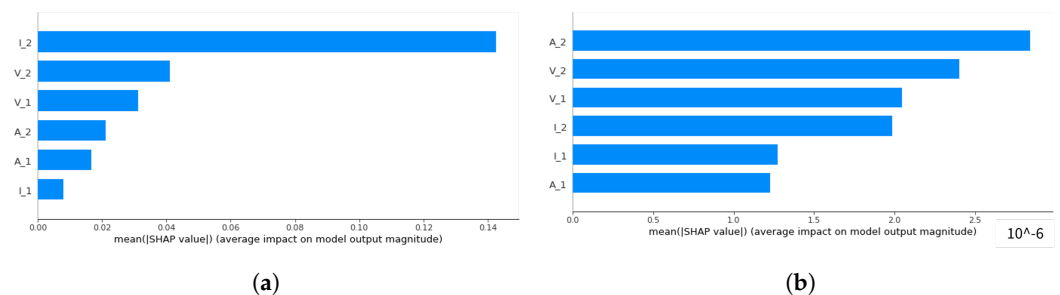
Table 9. Thresholded ANN classification results on Inconel 718 and Invar 36.

| Evaluation Metric | Inconel 718 | Invar 36 |
|---|-------------------------------|-------------------------------|
| Accuracy | 0.894 (95% CI {0.887, 0.901}) | 0.611 (95% CI {0.605, 0.617}) |
| Precision | 0.844 (95% CI {0.835, 0.853}) | 0.846 (95% CI {0.831, 0.860}) |
| Recall | 0.972 (95% CI {0.961, 0.982}) | 0.271 (95% CI {0.259, 0.282}) |
| False Positive Rate (FPR) | 0.202 (95% CI {0.197, 0.208}) | 0.117 (95% CI {0.099, 0.135}) |
| F1 | 0.903 (95% CI {0.893, 0.912}) | 0.354 (95% CI {0.346, 0.363}) |
| Matthew's Correlation Coefficient (MCC) | 0.796 (95% CI {0.781, 0.810}) | 0.245 (95% CI {0.230, 0.259}) |

The results of Inconel 718 are good, with a high MCC indicating an extremely strong correlation between prediction and original data. The false positive rate is approximately 20%. The high precision and recall indicates a good ability to predict high-quality positives (defects), which is expected since it is the majority class.

The results of Invar 36 are seen to be average–poor. The high precision indicates that the positive (defect) detection performed by the model is of high quality (i.e., low false negative predictions). The recall score is poor, indicating a poor ability to detect true positives. This indicates that the model is very strict with the criteria for classifying a datapoint as positive, leading to a low number of high-quality positive predictions. The false positive rate is also very low. The MCC is closer to 0 than to 1, indicating a weak positive correlation between ground truth and predictions.

The feature importances in neural networks are calculated differently from random forests. Shapley additive explanations (SHAP) is an approach developed by Lundberg et al. which utilizes game theory to determine the contribution of each feature to the model prediction [33]. The SHAP feature importances for the ANN model in the case of both Inconel 718 and Invar 36 are shown in Figure 13a,b.

**Figure 13.** SHAP feature importances of ANN model for (a) Inconel 718 and (b) Invar 36.

4.4. Cross-Material Analysis

The objective of this section is to examine the possibility of training a model on the defective/clean samples of one material (say, Material 1) and testing the model on a different material (say, Material 2). Training on Inconel 625 was omitted due to the lack of a sufficient number of audio datapoints.

ANN models separately trained on Inconel 718 and Invar 36 using the hyperparameters in Table 5 were tested on the other two materials (for example, trained on Inconel 718 and tested on Invar 36 and Inconel 625). Bootstrapping was performed (250 samples) in order to construct confidence intervals. The obtained classification results are shown in Tables 10 and 11.

Table 10. Thresholded ANN classification results on Invar 36 and Inconel 625 using training on Inconel 718.

| Evaluation Metric | Score on Invar 36 | Score on Inconel 625 |
|---|-------------------------------|-------------------------------|
| Accuracy | 0.258 (95% CI {0.245, 0.272}) | 0.078 (95% CI {0.069, 0.087}) |
| Precision | 0.181 (95% CI {0.178, 0.184}) | 0.000 (95% CI {0.000, 0.000}) |
| Recall | 0.962 (95% CI {0.954, 0.969}) | N/A |
| False Positive Rate (FPR) | 0.881 (95% CI {0.865, 0.898}) | 0.922 (95% CI {0.913, 0.931}) |
| F1 | 0.303 (95% CI {0.299, 0.307}) | 0.000 (95% CI {0.000, 0.000}) |
| Matthew's Correlation Coefficient (MCC) | 0.086 (95% CI {0.076, 0.097}) | N/A |

When looking at the prediction on Invar 36, it is clear that the metrics are very poor. The MCC is very close to 0, indicating a weak positive correlation between prediction and ground truth. The recall and false positive rates are both high, indicating that the model considers most examples to be defective. The prediction on Inconel 718 is also poor. The false positive rate is less than 10%. The precision is high and false positive rate is low, indicating that the model is good at selecting positives of high quality. The recall is low, indicating that only a low fraction of the actual positives are detected by the model. The MCC is close to 0, indicating nearly no positive correlation between predictions and ground truth. For Inconel 625, the lack of defective datapoints leads to two undefined metrics and other poor results.

Table 11. Thresholded ANN classification results on Inconel 718 and Inconel 625 using training on Invar 36.

| Evaluation Metric | Score on Inconel 718 | Score on Inconel 625 |
|---|-------------------------------|-------------------------------|
| Accuracy | 0.271 (95% CI {0.260, 0.281}) | 0.681 (95% CI {0.643, 0.720}) |
| Precision | 0.906 (95% CI {0.884, 0.928}) | 0.000 (95% CI {0.000, 0.000}) |
| Recall | 0.144 (95% CI {0.131, 0.158}) | N/A |
| False Positive Rate (FPR) | 0.093 (95% CI {0.071, 0.116}) | 0.319 (95% CI {0.280, 0.357}) |
| F1 | 0.234 (95% CI {0.216, 0.252}) | 0.000 (95% CI {0.000, 0.000}) |
| Matthew's Correlation Coefficient (MCC) | 0.064 (95% CI {0.043, 0.084}) | N/A |

Similar to earlier, RF models trained separately on Inconel 718 and Invar 36 using the hyperparameters in Table 4 were tested on the other two materials. The classification results are shown in Tables 12 and 13.

Table 12. Random forest results on Invar 36 and Inconel 625 using training on Inconel 718.

| Evaluation Metric | Score on Invar 36 | Score on Inconel 625 |
|---|-------------------|----------------------|
| Accuracy | 0.172 | 0.000 |
| Precision | 0.162 | 0.000 |
| Recall | 0.958 | N/A |
| False Positive Rate (FPR) | 0.983 | 1.000 |
| F1 | 0.277 | 0.000 |
| Matthew's Correlation Coefficient (MCC) | -0.066 | N/A |

Table 13. Random forest results on Inconel 718 and Inconel 625 using training on Invar 36.

| Evaluation Metric | Score on Inconel 718 | Score on Inconel 625 |
|---|----------------------|----------------------|
| Accuracy | 0.199 | 0.545 |
| Precision | 0.636 | 0.000 |
| Recall | 0.093 | N/A |
| False Positive Rate (FPR) | 0.267 | 0.455 |
| F1 | 0.162 | 0.000 |
| Matthew's Correlation Coefficient (MCC) | -0.198 | N/A |

Tables 12 and 13 show that even with the random forests approach, the results are poor. For the prediction on Invar 36, the accuracy is found to be low, and the MCC is negative and close to 0, indicating a very weak negative correlation. The recall score is high, indicating a high fraction of true positives (defects) being detected, but the precision is very low, indicating a poor quality of positive prediction. The false positive rate is high, indicating the tendency of the model to label most samples it sees as positive. Even with switching to Invar 36 for training the model, the results remain poor. For the prediction on Inconel 718, the accuracy is still low and the precision is higher than with ANNs. The recall score and false positive rate have become worse, indicating that the model has become too stringent with finding positives. The MCC is negative and higher in magnitude, indicating a mild negative correlation between predictions and ground truth. For Inconel 625, the lack of defective data causes poor results.

4.5. Discussion

Using the welding process parameters used in this study, the average arc power used when depositing clean and defective beads can be calculated for each material. The experiments conducted involved a relatively small range of deposition speeds (0.50 m/min–0.66 m/min for Inconel 718), and thus a population average of traverse speeds was taken for the defective and clean bead sets of each material. Using the average arc power and average traverse speeds, the average value of arc energy per unit length (AEL) can be determined for each of the bead sets. The formulae for both quantities are shown in Equations (4) and (5), where V is welding voltage, I is welding current and v is welding speed. In addition to these quantities, it is useful to examine the ideal energy [34], which is the energy that melts a unit volume of the material (volumetric melting energy—VME). This can be calculated using Equation (6), where ρ is mass density, C_p is specific heat, T_m is melting temperature, T_i is initial temperature and L_m is latent heat of fusion [34]. The calculated ideal energy for each material is listed in Table 14 together with the average power and AEL values. The material thermophysical properties used for computations are taken from references [10–12,35–37].

$$P_{avg} = VI \quad (4)$$

$$AEL = \frac{VI}{v} \quad (5)$$

$$E_i = \rho(C_p(T_m - T_i) + L_m) \quad (6)$$

Comparing the AEL values of clean and defective beads listed in Table 14, both Inconel 718 and Invar 36 have correspondence between defects and increased AEL. This makes logical sense when looking at Invar 36 since the reason for bead overflow is the lack of sufficient heat dissipation from the melt pool due to low thermal conductivity (10.49 W/mK [11]). Thus, increased heat input would exacerbate the issue and lead to an increased probability of defects. For Inconel 718, the literature suggests that increased

heat input can lead to cracking due to increased segregation of elements, leading to the formation of brittle Laves phases [38,39]. Volumetric melting energy (VEM) is commonly used in selective laser melting (SLM) for the evaluation of dense material deposition [40]. By taking into account the wire feeding rate and wire diameter (1.2 mm), together with the AEL, the calculated VEMs in our cases are 5–10 times more than the ideal energy, which also indirectly supports our observation that we do not find any lack of fusion or porosity in our current study. From the section on feature extraction, it is clear that the two kinds of defects in question are hot cracking and bead overflow. The most significant similarity between these defects is the dependence on AEL, since that controls the amount of heat supplied for bead formation. Thus, the average values of AEL in Table 14 will be used to try to explain the trends in error metrics shown in the results.

Table 14. Average arc power, AEL and VME values of all materials.

| Material | Volumetric Melting Energy (J/mm ³) | Clean | | Dirty | |
|-------------|--|-----------|--------------------------|-----------|--------------------------|
| | | Power (W) | Arc Energy/Length (J/mm) | Power (W) | Arc Energy/Length (J/mm) |
| Inconel 718 | 5.92 | 3201.88 | 291.08 | 2938.61 | 322.57 |
| Inconel 625 | 6.86 | 3769.24 | 339.62 | - | - |
| Invar 36 | 7.99 | 2216.95 | 308.84 | 2492.30 | 536.46 |

4.6. Predictions between Inconel 718 and Invar 36

Consider an ML model trained on Inconel 718 data. This model would likely learn the fact that high arc power typically corresponds to clean beads, whereas reduced arc power corresponds to defective beads (the exact threshold is difficult to identify, but the average powers of both bead sets are known from Table 14). It is important to note that since traverse speed was not passed as an input parameter, the model should not have information regarding AEL. If the power range of Inconel 718 is compared to the power range of Invar 36, it is immediately clear that the entire power range of Invar 36 is smaller in magnitude than the range of Inconel 718. This effectively means that if input power was the only factor being considered, any system that learns the defect–sensor correlations for Inconel 718 would assume nearly the entire set of Invar 36 to be defective (due to the power being even lower than the average defective bead power of Inconel 718). This is clearly reflected in the cross-material analysis results shown in Tables 10 and 12, where the precision score is low and the recall score is high. This combination indicates the tendency of random forest and neural network models to consider an excessively large fraction of the testing set to be positive (defective). At this juncture, one may argue about the influence of audio parameters since they do not directly correspond to heat input. To explain this, one can examine the feature importances for the single-material analysis performed on Inconel 718. Audio parameters are expected to have low importance relative to voltage and current for Inconel 718, which can be logically understood when looking at the kind of defect that is formed in this material, i.e., hot cracks. Hot cracking involves the formation of low-melting phases and thermal strains, none of which are likely to have a significant impact on the stability of the arc since they occur during material solidification. Since audio signals primarily capture the stability of the electric arc during the WAAM process, it stands to reason that a defect which does not directly influence or depend on arc instabilities would have a smaller signature in audio sensor responses compared to welding current and welding voltage. Due to this, it is unlikely that a model trained on Inconel 718 would consider the existence of a significant relation between audio signals and defects, due to which there is minimal effect of audio discrepancies on the prediction capabilities of a model trained on Inconel 718 and tested on Invar 36.

The next step would be to consider the inverse situation, i.e., an ML model trained on Invar 36 data attempting to make predictions on Inconel 718 data. From Table 14, it

is clear that such a model would likely learn the fact that increased welding voltage and current correspond to defective beads and reduced input power corresponds to clean beads. The exact line of delineation between “high” and “low” is difficult to identify, but the mean power values are available (in Table 14). Comparing the ranges of power values of Inconel 718 and Invar 36, the entire range of Invar 36 is seen to be of lower magnitude than the range of Inconel 718. Thus, a model trained on Invar 36 would be likely to consider most beads of Inconel 718 to be defective (if the input power is the primary feature being considered). This, however, is nearly the exact opposite of what is portrayed in the cross-material analysis results in Tables 11 and 13. The precision score is high and the recall score is low, indicating that both random forest and neural network models are inclined to predict a low number of high-quality positives (defects), which is strange.

At this point, it is useful to return to the AEL. Even though it was stated earlier that the model is unlikely to have information regarding the AEL due to the lack of a welding speed parameter in the input, if the AEL is taken as a reference instead of input power, all the abovementioned results suddenly make sense. Contrary to what was earlier mentioned, the trend from clean to defective beads is found to be increasing for both materials, and the range of magnitudes is found to be higher for Invar 36 compared to Inconel 718. From this perspective, a model trained on Inconel 718 will still find most Invar 36 beads to be defective (due to the higher values of AEL) and a model trained on Invar 36 will find most Inconel 718 beads to be clean (due to the lower values of AEL). To support this, the results from Tables 7 and 9 show that, for Invar 36, the detection of positives (defects) is difficult. Thus, it is safe to say that the model somehow either finds the trends of AEL without including a welding speed parameter or manages to combine the three kinds of input parameters (welding current, welding voltage and welding audio) in a fashion that mimics the trend of the AEL. Assessing the impact of audio data in this scenario, one can expect that the relative importance of audio features is slightly increased compared to Inconel 718. This may be understood when considering the nature of defects seen in Invar 36, i.e., bead overflow, which happens during the melting process. While the root of the issue still lies in the thermal input of the printing process, the presentation of the defect is topological in nature (unlike cracks, which are fully internal). The change in the surface geometry of the bead induced by molten metal runoff causes significant variation in the instantaneous contact tip distance of the welding torch from the metal surface. This variation can induce instabilities in the arc which directly translate to clear audio signatures. Thus the discrepancies in audio are expected to have a bigger impact on the prediction capabilities of a model trained on Invar 36 and tested on Inconel 718 compared to earlier.

To quantify this, it is useful to compile the precision, recall and F1 scores of both kinds of predictions using both neural network (ANN) and random forest (RF) models. The F1 score is added as a summary of the precision and recall since it is the harmonic mean of both metrics. The scores are shown in Table 15.

Table 15. Precision, recall and F1 scores of cross-material predictions between Inconel 718 and Invar 36.

| ML Model | Trained on Inconel 718 Tested on Invar 36 | | | Trained on Invar 36 Tested on Inconel 718 | | |
|----------|--|--------|-------|--|--------|-------|
| | Precision | Recall | F1 | Precision | Recall | F1 |
| ANN | 0.181 | 0.962 | 0.303 | 0.906 | 0.144 | 0.234 |
| RF | 0.162 | 0.958 | 0.277 | 0.636 | 0.093 | 0.267 |

The F1 values are seen to be slightly higher when predicting on Invar 36 from Inconel 718 compared to the reversed situation. This could be due to the difference in relative importance of audio parameters, as explained earlier, but the difference is small (the increase is close to 30% for the ANN model and approximately 4% for the RF model). It is interesting to note that when looking at the effect of class imbalance, there is not much

difference seen depending on which class is in the minority (since Inconel 718 and Invar 36 have opposite classes as the minority class). Both Inconel 718-Invar 36 and Invar 36-Inconel 718 predictions have comparable metrics, and the MCC only differs by approximately 0.023. It would be useful in the future to look into cases where the same class is the minority for both materials in order to see whether the prediction quality changes.

4.7. Predictions on Inconel 625

On the one hand, going off the trend of AEL seen in Table 14, one would expect that a model trained on Inconel 718 would predict most datapoints of Inconel 625 to be defective (since the average AEL of Inconel 625 is higher than the entire range of values for Inconel 718). On the other hand, a model trained on Invar 36 would predict a fair number of Inconel 625 datapoints to be clean since the AEL of Inconel 625 is closer to the average of clean Invar 36 beads compared to the defective beads. This expected result is confirmed when looking at Tables 11 and 12. Since the precision is zero in all cases (due to the lack of ground truth positives), the false positive rate can be examined to assess the tendency of the models to erroneously predict positives. The compiled results are shown in Table 16.

Table 16. False positive rates of cross-material predictions on Inconel 625.

| ML Model | Trained on Inconel 718 | Trained on Invar 36 |
|----------|------------------------|---------------------|
| ANN | 0.922 | 0.319 |
| RF | 1.000 | 0.455 |

5. Conclusions

Through the course of this research work, many interesting aspects were found between various sensor responses and defects. Physically consistent reasonings were deduced for most observations, but some still need further investigations. The primary conclusions made from this work can be summarized as shown below.

1. Out of the two supervised ML models used for predictions on the single-material defect status of WAAM deposits, for the most part, the random forest model is found to have better results compared to the ANN in terms of accuracy (91.1% vs. 89.4% for Inconel 718), F1 (92.6% vs. 90.3% for Inconel 718) and MCC (83.0% vs. 79.6% for Inconel 718) scores.
2. Even though welding speed as a parameter is not introduced to the supervised models at any stage of training, the trends seen in the predictive nature of these models show similarities to the energy/length values, which would theoretically require knowledge of the welding speed.
3. Cross-material predictions between Inconel 718 and Invar 36 are found to heavily depend on the nature of the defect in either material. Training with Inconel 718 is found to lead to overprediction of defects in Invar 36 (recall > 95% and precision < 20%) and Inconel 625 (false positive rate > 92%), whereas training with Invar 36 is found to lead to underprediction of defects in Inconel 718 (precision > 60% and recall < 15%) and Inconel 625 (false positive rate < 46%).
4. In both random forest and ANN models with a reduced amount of minority class datapoints in the training set (33.3% of total), the testing performance of Inconel 718 was found to be better than Invar 36 in terms of accuracy (91.1% vs. 61.1% for random forest), F1 (92.6% vs. 22.2% for random forest) and MCC (83.0% vs. 27.1% for random forest) scores.
5. The usage of welding current, welding voltage and audio signals is found to provide information about the stability of the arc. However, tracking additional details such as heat input, weld contaminants, etc. will require additional parameters or sensor responses. One possibility is to include spectroscopic sensing in order to gather arc temperature and arc plasma composition data. In addition, attempting to extract more parameters from the voltage, current and audio data can be helpful.

Based on the current research work, it can be seen that there is good potential for the application of machine learning to complex processes such as gas metal arc welding-based additive manufacturing. However, much fundamental research work has to be carried out for better understanding or for establishing sets of rules, so that the machine learning techniques can be adopted for achieving inline usage, contributing to building next-generation smart systems. Future work would mainly involve the expansion of both defect domain and material domain to include more variety and generalization in the modeling approach. In addition, new models using incremental learning could be potentially useful for continuous learning [41].

Author Contributions: Conceptualization, A.R. and W.Y.; methodology, A.R. and W.Y.; software, A.R.; validation, A.R., W.Y. and M.H.; resources, W.Y.; writing—original draft preparation, A.R.; writing—review and editing, W.Y. and M.H.; visualization, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: This research was carried out under a master project of Delft University of Technology in collaboration with RAMLAB BV (www.ramlab.com). The authors would like to express sincere gratitude and thank all researchers from Delft University of Technology and RAMLAB who provided valuable support and input to this research. A special thanks to Vincent Wegener for providing additional financial support, and special appreciation for Remco Rook for technical support from RAMLAB. Voestalpine Böhler Welding is acknowledged for providing the wires. AirProducts® is acknowledged for the provision of the shielding gas mixture.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, X.; Kong, F.; Fu, Y.; Zhao, X.; Li, R.; Wang, G.; Zhang, H. A review on wire-arc additive manufacturing: Typical defects, detection approaches, and multisensor data fusion-based model. *Int. J. Adv. Manuf. Technol.* **2021**, *117*, 707–727. [CrossRef]
2. Zhang, T.; Liu, C.T. Design of titanium alloys by additive manufacturing: A critical review. *Adv. Powder Mater.* **2022**, *1*, 100014. [CrossRef]
3. Workman, G.L. (Ed.) *Nondestructive Testing Handbook*, 3rd ed.; American Society for Nondestructive Testing: Columbus, OH, USA, 2012; Volume 10
4. Bhattacharya, S.; Pal, K.; Pal, S.K. Multi-sensor based prediction of metal deposition in pulsed gas metal arc welding using various soft computing models. *Appl. Soft Comput.* **2012**, *12*, 498–505. [CrossRef]
5. Qiu, W.; Murphy, W.J.; Suter, A. Kurtosis: A New Tool for Noise Analysis. *Acoust. Today* **2020**, *16*, 39–47. [CrossRef]
6. Alfaro, S.C.A.; Cayo, E.H. Sensing Fusion Data from the Optic and Acoustic Emissions of Electric Arcs in the GMAW-S Process for Welding Quality Assessment. *Sensors* **2012**, *12*, 6953–6966. [CrossRef] [PubMed]
7. Qin, J.; Hu, F.; Liu, Y.; Witherell, P.; Wang, C.C.; Rosen, D.W.; Simpson, T.W.; Lu, Y.; Tang, Q. Research and application of machine learning for additive manufacturing. *Addit. Manuf.* **2022**, *52*, 102691. [CrossRef]
8. Ko, H.; Witherell, P.; Lu, Y.; Kim, S.; Rosen, D.W. Machine learning and knowledge graph based design rule construction for additive manufacturing. *Addit. Manuf.* **2021**, *37*, 101620. [CrossRef]
9. He, F.; Yuan, L.; Mu, H.; Ros, M.; Ding, D.; Pan, Z.; Li, H. Research and application of artificial intelligence techniques for wire arc additive manufacturing: A state-of-the-art review. *Robot. Comput. Integr. Manuf.* **2023**, *82*, 102525. [CrossRef]
10. INCONEL Alloy 718. Technical Report, Special Metals, New York, USA. 2021. Available online: <https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-718.pdf> (accessed on 21 August 2023).
11. INVAR 36. Technical Report, Salomon's Metalen BV, Groningen, The Netherlands. 2020. Available online: https://salomons-metalen.nl/datasheets/Invar_36.pdf (accessed on 21 August 2023).
12. INCONEL Alloy 625. Technical Report, Special Metals, New York, USA. 2021. Available online: <https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-625.pdf> (accessed on 21 August 2023).
13. The Arc Welding Robot System TAWERS. Technical Report, Panasonic, Osaka, Japan. 2018. Available online: https://industrial.panasonic.com/content/data/WS/PDF/201801_TAWERS_E.pdf (accessed on 21 August 2023).
14. Lin, Z.; Ya, W.; Subramanian, V.; Goulas, C.; di Castri, B.; Hermans, M.; Pathiraj, B. Deposition of Stellite 6 alloy on steel substrates using wire and arc additive manufacturing. *Int. J. Adv. Manuf. Technol.* **2020**, *111*, 411–426. [CrossRef]
15. Bevans, B.; Ramalho, A.; Smoqi, Z.; Gaikwad, A.; Santos, T.G.; Rao, P.; Oliveira, J. Monitoring and flaw detection during wire-based directed energy deposition using in-situ acoustic sensing and wavelet graph signal analysis. *Mater. Des.* **2023**, *225*, 111480. [CrossRef]

16. Ramalho, A.; Santos, T.G.; Bevans, B.; Smoqi, Z.; Rao, P.; Oliveira, J. Effect of contaminations on the acoustic emissions during wire and arc additive manufacturing of 316L Stainless Steel. *Addit. Manuf.* **2022**, *51*, 102585. [CrossRef]
17. Hauser, T.; Reisch, R.T.; Kamps, T.; Kaplan, A.F.; Volpp, J. Acoustic emissions in directed energy deposition processes. *Int. J. Adv. Manuf. Technol.* **2022**, *119*, 3517–3532. [CrossRef]
18. Sainburg, T.; Thielk, M.; Gentner, T.Q. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Comput. Biol.* **2020**, *16*, e1008228. [CrossRef] [PubMed]
19. O’Shaughnessy, D. *Speech Communication: Human and Machine*; Addison-Wesley: Boston, MA, USA, 1987; p. 150.
20. *2017 ASME Boiler & Pressure Vessel Code Section VIII Division*; ASME: New York, NY, USA, 2017.
21. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
22. Menze, B.H.; Kelm, B.M.; Masuch, R.; Himmelreich, U.; Bachert, P.; Petrich, W.; Hamprecht, F.A. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of Spectral Data. *BMC Bioinform.* **2009**, *10*, 213. [CrossRef] [PubMed]
23. Grossi, E.; Buscema, M. Introduction to artificial neural networks. *Eur. J. Gastroenterol. Hepatol.* **2008**, *19*, 1046–1054. [CrossRef] [PubMed]
24. Rithani, M.; Kumar, R.; Doss, S. A review on big data based on deep neural network approaches. *Artif. Intell. Rev.* **2023**, *56*, 14765–14801. [CrossRef]
25. Buda, M.; Maki, A.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* **2018**, *106*, 249–259. [CrossRef]
26. Leevy, J.; Khoshgoftaar, T.; Bauder, R.; Seliya, N. A survey on addressing high-class imbalance in big data. *J. Big Data* **2018**, *5*. [CrossRef]
27. Johnson, J.; Khoshgoftaar, T. Survey on deep learning with class imbalance. *J. Big Data* **2019**, *6*, 27. [CrossRef]
28. Elmrabbit, N.; Zhou, F.; Li, F.; Zhou, H. Evaluation of Machine Learning Algorithms for Anomaly Detection. In Proceedings of the 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Dublin, Ireland, 15–19 June 2020; pp. 1–8. [CrossRef]
29. Aminikhanghahi, S.; Cook, D.J. A survey of methods for time series Change point detection. *Knowl. Inf. Syst.* **2016**, *51*, 339–367. [CrossRef] [PubMed]
30. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [CrossRef] [PubMed]
31. Zoubir, A.; Iskander, D. Bootstrap Methods and Applications. *Signal Process. Mag. IEEE* **2007**, *24*, 10–19. [CrossRef]
32. Nandeshwar, A.R. Models for Calculating Confidence Intervals for Neural Networks. Master’s Thesis, College of Engineering and Mineral Resources at West Virginia University, Morgantown, WV, USA, 2006.
33. Lundberg, S.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:cs.AI/1705.07874.
34. Fiocchi, J.; Casati, R.; Tuissi, A.; Biffi, C.A. Laser beam welding of coccufemni high entropy alloy: Processing, microstructure, and mechanical properties. *Adv. Eng. Mater.* **2022**, *24*, 10. [CrossRef]
35. Raza, S. Superalloys: An introduction with thermal analysis. *J. Fundam. Appl. Sci.* **2015**, *7*, 364. [CrossRef]
36. Tinoco, J.; Fredriksson, H. Solidification of a modified Inconel 625 alloy under different cooling rates. *High Temp. Mater. Process.* **2004**, *23*, 13–24. [CrossRef]
37. Obidigbo, C.N.; Gockel, J. A Numerical and Experimental Investigation of Steady-State and Transient Melt Pool Dimensions in Additive Manufacturing of Invar 36. Master’s Thesis, Wright State University, Dayton, OH, USA, 2017. Available online: https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=2962&context=etd_all (accessed on 21 August 2023).
38. Artaza, T.; Bhujangrao, T.; Suárez, A.; Veiga, F.; Lamikiz, A. Influence of Heat Input on the Formation of Laves Phases and Hot Cracking in Plasma Arc Welding (PAW) Additive Manufacturing of Inconel 718. *Metals* **2020**, *10*, 771. [CrossRef]
39. The formation and control of Laves phase in superalloy 718 welds. *J. Mater. Sci.* **1997**, *32*, 1977–1984. [CrossRef]
40. Scipioni Bertoli, U.; Wolfer, A.J.; Matthews, M.J.; Delplanque, J.P.R.; Schoenung, J.M. On the limitations of Volumetric Energy Density as a design parameter for Selective Laser Melting. *Mater. Des.* **2017**, *113*, 331–340. [CrossRef]
41. Li, Y.; Polden, J.; Pan, Z.; Cui, J.; Xia, C.; He, F.; Mu, H.; Li, H.; Wang, L. A defect detection system for wire arc additive manufacturing using incremental learning. *J. Ind. Inf. Integr.* **2022**, *27*, 100291. [CrossRef]

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.