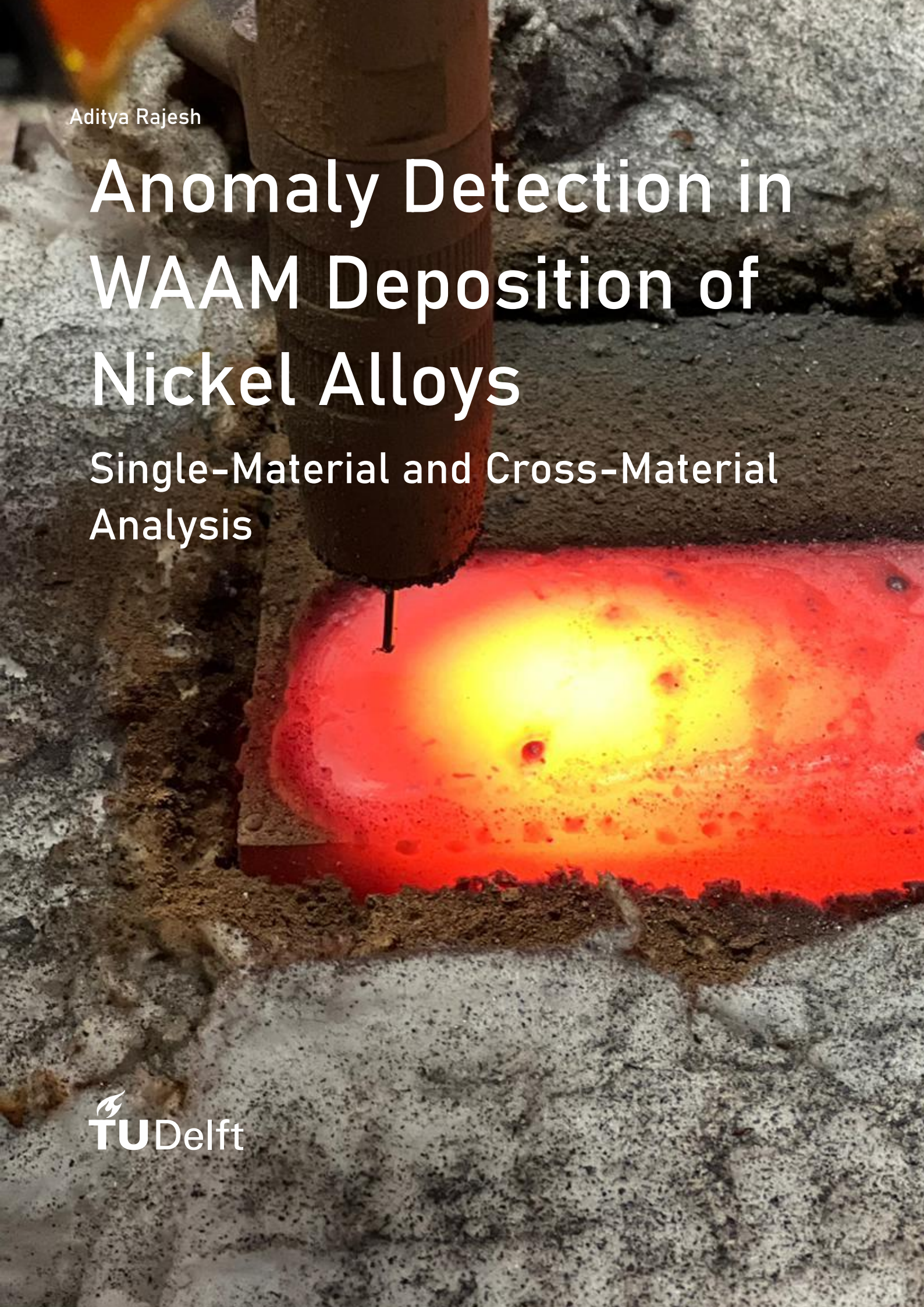


Aditya Rajesh

Anomaly Detection in WAAM Deposition of Nickel Alloys

Single-Material and Cross-Material Analysis

TU Delft

Anomaly Detection in WAAM Deposition of Nickel Alloys

Single-Material and Cross-Material Analysis

By

Aditya Rajesh

in partial fulfilment of the requirements for the degree of

Master of Science

in Materials Science and Engineering

at the Delft University of Technology,

to be defended publicly on Wednesday August 23, 2023, at 12:30 PM.

Supervisor:	Prof. dr. ir. M.J.M Hermans	
	Dr. ir. Wei Ya	
Thesis committee:	Prof. dr. Sid Kumar,	TU Delft
	Prof. dr. Jie Zhou,	TU Delft

TU Delft

Acknowledgements

My journey in science and research began with a spark that was lit by my loving parents and sister. Their unending support for my passion led to my pursuit of this Master program at TU Delft, which culminates in this thesis. The past 7 months working on this research has been an extremely interesting ride with highs and lows. I have gotten to learn so much about the practical aspects of wire arc additive manufacturing (WAAM) and have gotten the opportunity to deepen my knowledge of machine learning beyond being a “magic wand” to fix all problems and the potential pitfalls to avoid.

Working at RAMLAB has been a great experience, with all the opportunities to see industrial WAAM up close and deal with its difficulties firsthand. I extend my heartfelt gratitude to Vincent Wegener, who provided the opportunity to work with RAMLAB on this project. I also deeply thank my company supervisor Dr. Wei Ya, who has been a constant guide throughout the entire undertaking, helping me widen my knowledge base of additive manufacturing and always keep improving the quality of my research. I also thank Remco, José, and the other members of the RAMLAB team who helped me with the sensor setups and experimental details.

I would also like to thank my university supervisor, Dr. Marcel Hermans who has been very motivational through the entire process and helped me to shape the content of this thesis into a polished result. The members of the lab at TU Delft, including Remko, Durga, Kevork and Alice have been immensely helpful with the microstructural analysis and helping me refine my metallographic preparation skills. I also thank Arjun Sood for his insights into the behaviour of Invar 36 which helped me interpret my results better.

I also thank all of my friends Laura, Elena, Marloes, Vaibhav, Paul, Adithya and many more for being a huge rock of support for me not just through these 7 months, but through the entire 2 years of my Master program.

Thank you all for helping me so much on my journey so far.

Aditya Rajesh

Delft, August 2023

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AEL	Arc Energy per unit Length
ANN	Artificial Neural Network
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
CMT	Cold Metal Transfer
CNN	Convolutional Neural Network
DBSCAN	Density Based Spatial Clustering of Applications with Noise
DED	Directed Energy Deposition
FL	Focal Loss
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FPS	Frames per Second
GMAW	Gas Metal Arc Welding
GTAW	Gas Tungsten Arc Welding
HICC	Hydrogen-Induced Cold Cracking
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MCC	Matthew's Correlation Coefficient
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Mean Squared Error
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic – Area under Curve
SAWP	Super Active Welding Process
SE	Squeeze and Excitation
SGD	Stochastic Gradient Descent
SHAP	Shapley Additive Explanations
SNS	Spectral Noise Subtraction
SSD	Sum of Squared Distances
SVM	Support Vector Machine
TIG	Tungsten Inert Gas
TN	True Negative
TP	True Positive
WAAM	Wire Arc Additive Manufacturing
WPS	Welding Procedure Specification
YOLO	You Only Look Once

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 (welding voltage, welding current and welding audio) for predictions. A variety of types of prints, including ramp tests, single bead depositions, and walls were explored. Three different machine learning models are used – artificial neural networks (ANNs), K-Means clustering and random forests (RF), and the performances are compared. In addition to separate material analysis, cross-material predictions are conducted using two supervised models to investigate the prediction capabilities of such an approach. The results indicate that models are indeed capable of finding connections between welding parameters and defect formation, and 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. The cross-material predictions are significantly worse, with accuracies ranging from 20% to 27% and very weak correlation coefficients (less than 0.1). Analysis of the results indicates that the importance of audio sensor response depends on the nature of defect, and that additional sensors like spectrometers could give a wider range of information to cover more types of defects, potentially raising the performance of cross-material predictions. Between the models, random forest is found to perform the best overall, with ANNs coming in a close second. The versatility of ANNs indicates that increasing the dataset size and resolving the class imbalance could potentially tip the scales in the favor of ANNs.

1 Introduction

The development of different techniques of additive manufacturing, which is a manufacturing paradigm based around the addition of successive layers of material to form a complete product, has been the result of an increasing demand of complex geometries in various engineering fields like 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 generating complex geometries, heterogenous composition structures, and rapidly prototyping different components. While initially confined to prototyping research applications, additive manufacturing has since seen significant growth in mass manufacturing avenues [2].

Additive manufacturing techniques are typically classified based on the form of feed material and the energy source used [1, 3]. Feed material can be in the form of wire or as powder (which itself can be either in a bed or fed into the energy source), whereas energy sources can be lasers, electric arc or electron beams [1]. These processes are generally known as *directed energy deposition* (DED). This method utilizes an energy source like a laser to directly melt a wire/powder feedstock into a melt pool through a nozzle onto the substrate to form a bead of material after the melt pool solidifies [1, 4]. A track is generated when moving the heat source and feedstock in tandem. Overlapping multiple tracks results in a layer, and then a 3D object can be deposited layer by layer. An example of WAAM is shown in Figure 1.

Figure 1: Depiction of WAAM, courtesy RAMLAB

DED provides the advantage of a good deposition rate, ranging from 8.3 to over 330 g/min [5]. It is also capable of creating layers on uneven substrates [5].

DED techniques are further classified based on the type of feedstock and energy source used [5]. While powder-based DED typically involves only laser power, wire-based DED can use a variety of power sources, giving rise to wire arc additive manufacturing (WAAM), wire laser additive manufacturing (WLAM), and wire electron beam additive manufacturing (WEAM) [5, 6]. Out of the available energy sources, the electric arc is typically the most economical [3]. Thus, WAAM is one of the most popular DED processes among industries.

Since WAAM was developed on the basis of traditional arc welding, it can be conducted in a variety of modes depending on the type of arc welding being used. The two main arc welding processes that are most commonly used in industry are gas metal arc welding (GMAW) and gas tungsten arc welding (GTAW), and are briefly described below [1, 3].

Gas metal arc welding utilizes a *consumable* electrode acting as deposition metal. An arc is generated between the electrode and the substrate, which melts the electrode tip and transfers the molten metal to the melt pool on the substrate. Ideally, the electrode is fed in at a certain speed to offset the metal transfer by melting, such that the electrode stick-out length remains constant. Shielding gas is used to protect the melt pool from the atmosphere and may stabilize the arc. GMAW is reported to have a high deposition rate of about 3-4 kg/h [1], and has a relatively low setup cost, making it cost-effective and efficient [3, 5]. For larger wire diameter with a customized power setting, deposition rates as high as 5 kg/h can be achieved. A variant of GMAW called cold metal transfer (or CMT) achieves reduced heat input through the use of short circuiting [3, 10]. This kind of operation allows for higher arc stability and reduced spatter generation at the cost of deposition rate [5].

In gas tungsten arc welding, an arc is established between a *non-consumable* tungsten electrode and the substrate. A separate filler wire is fed into the arc in order to generate droplets that are deposited on the substrate. While GTAW has a lower deposition rate of around 1-2 kg/h [1], it shows reduced porosity and better surface finish [5]. GTAW also uses inert shielding gas for melt pool and electrode protection.

1.1 Defects in WAAM

The defects encountered in WAAM parts are very similar to the kind of defects seen in arc welding, since the same basic process is being used in both scenarios. Wu *et al.* [12] noted that some of the most common defects seen in WAAM-manufactured parts include (residual) stress-induced deformation, porosity, and cracking/delamination. Chen *et al.* [1] also mentioned additional types of defects like lack-of-fusion and humping. In addition, spatter formation may occur during wire-arc deposition. These defects will be described in greater detail below.

1.1.1 Spatter in WAAM

Spatter is defined as metal particles that are ejected from the melt pool or melting wire during the deposition process. Spatter particles may adhere to the surface of the top layer depending on the heat content of the particle, causing issues with surface finish and deposition of the next layer.

Cleaning such defects can be a difficult task at times as spatter may fuse with the substrate creating a strong bond. Furthermore, the removal of spatter reduces the deposition efficiency. An example of spatter is shown in Figure 2.

Figure 2: Weld bead with visible spatter, courtesy BOC Ltd.

Spatter can be generated due to a variety of mechanisms, including gas emission/explosion from the weld pool, molten metal scattering during arc re-ignition, and molten metal scattering due to repellent electromagnetic forces [21]. It can be seen that these factors depend on processing conditions like metal transfer mode, type of shielding gas, welding current, *etc*.

Shareef *et al.* [32] stated that arc instabilities are a prominent cause of spatter. When the arc fails to re-ignite properly, the feed electrode will touch the substrate/workpiece. This completes the electrical circuit, leading to rapid resistance heating and melting of the electrode. The increase in current during the short circuit and the associated Lorentz force may lead to an explosive rupture of the liquid bridge. This kind of sudden melting can lead to ejection of the molten metal, generating spatter. They also reported that shielding gas bubble bursts like CO_2 explosions can lead to spatter generation. Conventional GMAW welding is generally considered to have lower arc stability as compared to other welding processes like CMT and GTAW, making it more prone to spatter defects in general [3].

Wu *et al.* [12] reported that complex deposition paths can induce a higher risk of spatter ejection. This can be due to the higher probability of arc instabilities and deflection when tracing a complex path.

1.1.2 Porosity in WAAM

Porosity is defined as the presence of cavities within the material [51]. The presence of porosity reduces the effective material density, and results in diminished effective strength. Pores can also act as crack initiation points, making them detrimental to the fatigue properties of the printed part (to a lesser extent than cracks) [16]. An example of porosity in aluminum alloys is seen in Figure 3.

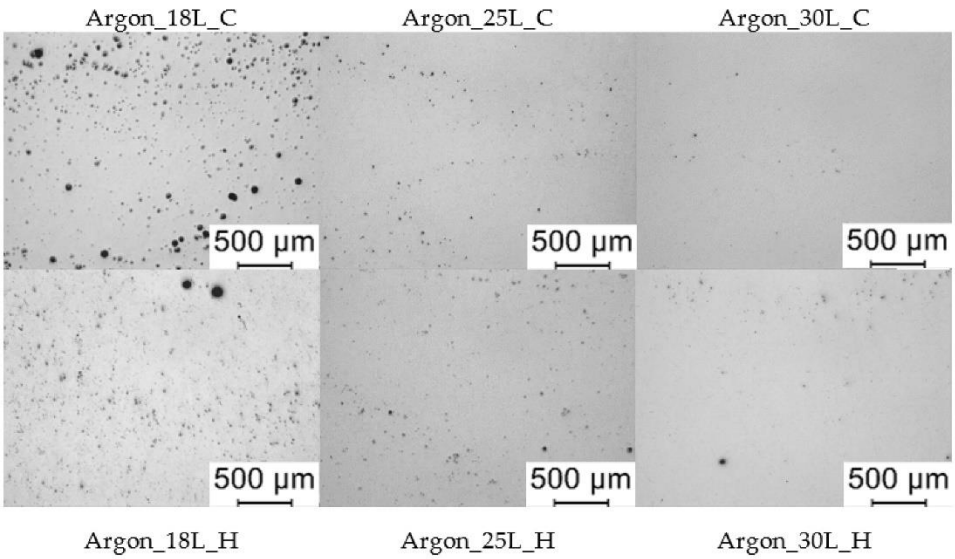


Figure 3: Porosity seen in AA536 WAAM samples [20]

Welding pores can be classified into two broad types, namely *hydrogen pores* and *process pores* [13, 14, 17]. Some researchers generalize hydrogen pores under a broader category of *material-induced pores* [1, 16]. These aforementioned types are explained briefly below.

1.1.2.1 Hydrogen/Material-induced Pores

Hydrogen pores are voids filled with hydrogen within the deposited material. Hydrogen is typically generated during the welding process from different contaminants on the surface of the feed wire. Moisture, or grease and other hydrocarbon-based contaminants, when exposed to the intense heat of the welding process, get converted to atomic hydrogen which then enters the melt pool [13, 15]. When this hydrogen is unable to escape molten deposited metal before solidification, pores are generated.

In aluminum and aluminum alloys, hydrogen shows a significant difference in solubility in liquid and solid phases (0.71 cm^3/100g in liquid aluminum at melting point and 0.043 cm^3/100g in solid aluminum), which acts as the main driving force behind the formation of hydrogen pores in aluminum structures deposited using WAAM [13, 18]. Such voids are typically spherical in nature and relatively homogenously distributed in the material [13].

Hydrogen can also migrate into shrinkage voids, which are formed during the cooling of molten metal [14]. The latter case can also be considered as a lack-of-fusion defect since shrinkage voids imply lack of proper joining. In fact, some authors consider lack-of-fusion to be a subtype of porosity, called interlayer porosity [1, 6].

1.1.2.2 Process Pores

Process pores are formed when instabilities in welding/deposition lead to lack-of-fusion and material spatter [1, 12]. This induces irregular voids in the built component. As seen in the previous section about hydrogen pores, such voids can also be considered hydrogen pores if

hydrogen migration occurs. Their distribution is relatively heterogenous due to this mechanism of formation being more random in nature [13].

Another possible mode of process porosity formation is the entrapment of shielding gas in the molten metal prior to solidification [6, 13]. This is very similar to the formation of hydrogen pores, with the only difference being the formation mechanism of the gas that generates the pores. Thus, such pores are also spherical in nature. The distribution of such pores is still random as compared to hydrogen pores since the entrapment of shielding gas does not happen in a uniform fashion [6].

1.1.3 Lack-of-Fusion in WAAM

Lack-of-fusion is the phenomenon of insufficient melting of the substrate material, leading to lack of proper joining between filler metal and substrate. This leads to void generation within the material, and in some cases, complete layer delamination of the deposited material [1, 3]. As seen earlier, lack-of-fusion can be considered a subtype of porosity in some cases.

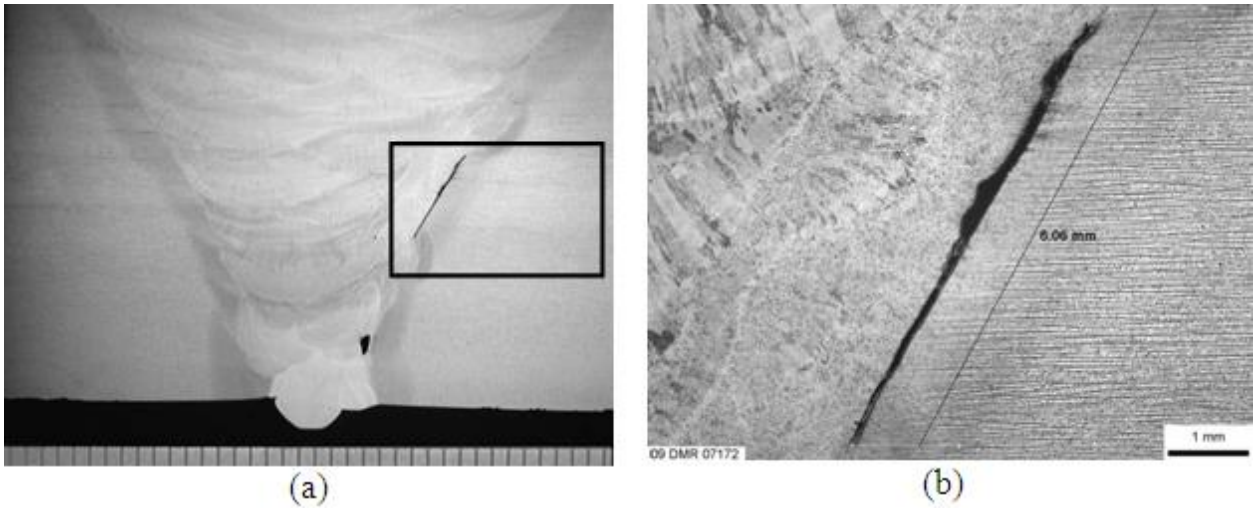


Figure 4: Lack of fusion seen in TIG welding samples [109]

One of the most obvious reasons for lack-of-fusion is the lack of sufficient heat input to melt the substrate material insufficiently, as can be seen in Figure 4 for a TIG welded sample. This may happen for certain deposition conditions which involve low heat input, as is for instance also the case in the cold metal transfer process. Cong *et al.* [19] reported about the generation of lack-of-fusion defects when using CMT-P-ADV (Pulse-advanced CMT) when depositing beads in parallel fashion for the printing of block structures.

Arc instabilities can also lead to lack-of-fusion alongside other defects like porosity [20]. This is because problems with the arc directly translate to problems with melting the feed wire. This leads to large voids in the specific regions of the print where the problem occurred.

Lack-of-fusion defects are irregular in shape due to their nature of formation, leading to multiple potential crack initiation points. In addition, the lack of sufficient bonding between layers due to this defect can compromise the mechanical integrity of the final part.

1.1.4 Cracking in WAAM

Cracks can be formed in a material due to various reasons. These defects can extend through the material upon the application of a certain threshold stress, leading to complete failure of the component. Cracks can be generated by locally exceeding the tensile strength of the material, which can happen due to a variety of phenomena. An example of solidification cracks is shown in Figure 5. In welding and additive manufacturing processes, cracking can be classified into two main types, namely *hot cracking* and *cold cracking* [1, 3, 22]. These cracking types are summarized below.

Figure 5: Microscopic solidification cracks in repair material samples

1.1.4.1 Hot Cracking

Hot cracking is cracking that occurs at elevated temperatures during the deposition process. A combination of mechanisms can lead to the formation of hot cracks in the final stages of solidification [1].

Solidification cracking occurs due to the differential in volumetric change of different phases of the melt pool during cooling [1, 23, 24], *i.e.*, solidification shrinkage and thermal contraction, generating stresses within the material (microscopic level). The stress state of the component is also affected by the clamping conditions of the printed part (macroscopic level). In addition, during solidification segregation may takes place, affecting the liquidus temperature of the remaining liquid. During the cooling of just-deposited liquid metal, solidification is accompanied by the trapping of liquid metal in inter-dendritic regions, *i.e.* a mushy zone is created. The strength of this mushy zone, which is the semisolid phase before complete cooling, is comparatively low due to the presence of liquid metal in the inter-dendritic zones and the elevated temperature. Initially, the amount of liquid is sufficient to accommodate the generated strains via grain rearrangement, but towards the end of solidification, the liquid fraction is too low to allow such rearrangement.

At this point, the residual stresses lead to the generation of intergranular cracks within the material that persist after complete solidification [25]. Seow *et al.* [26] mentioned that microsegregation during the solidification stage can increase the risk of cracking due to microstructural heterogeneities.

Hot cracking can also occur due to liquation at grain boundaries. Certain classes of alloys, like superalloys (being comprised of many different elements), are prone to forming low melting phases (like Laves phases) under the processing conditions of WAAM [26]. Such phases can segregate to grain boundaries and melt, leading to highly reduced strength. Subsequently, cracks will generate under developing tensile stresses in the cooling phase of the thermal cycle. This leads to intergranular cracking similar to solidification cracking [27, 28]. Grain size affects cracking susceptibility, as evident from the mechanism of crack generation. Coarse grains have less freedom of movement and rearrangement, leading to lower ductility which increases the cracking risk, while finer grain structures have lower cracking susceptibility due to enhanced ductility [23, 29]. Chen *et al.* [1] stated that hot cracks are typically irregular in shape, which is logical, considering that such cracks are intergranular in nature.

1.1.4.2 Cold Cracking

Cold cracking occurs eventually after deposition when the metal part is completely solidified. This kind of cracking is caused by residual stresses within the material after the thermal cycle of welding [22]. When this stress exceeds the tensile strength of the printing material, cold cracks are generated inside the printed part [1].

Residual stresses within the material can be magnified by the presence of discontinuities like slag inclusions and porosity since these act as stress concentrators. In addition, if the ductility of the printed material is low, this increases the cracking risk due to lack of accommodation of residual stresses as plastic strain. Finally, the presence of hydrogen within the weld can increase the cold cracking risk due to its embrittling effect [30]. The latter is specifically called hydrogen-induced cold cracking, or HICC.

Chen *et al.* [1] reported that cold cracks are typically straight in shape. Cold cracks are associated with high crack propagations velocities and the sudden release of energy, generating sound. In addition, it should be mentioned that cold cracking may manifest itself after an extended period of time after deposition.

1.2 Multi Sensor-based Defect Detection

A skillful welder monitors the welding process making use of eyesight (monitoring the pool shape and the deposition of the consumable) and on sound (unstable processes generate irregular noises). He is capable of directly acting upon his observations. The automation of this kind of monitoring could be done using sensors.

All kinds of signals could be considered to link to the occurrence of defects. In order to differentiate between defects, ‘fingerprints’ unique to each defect type should be extracted from the measured data. Increasing the number of sensors can be expected to increase the reliability of detection.

The concept of a multi-sensor monitoring system in welding is not a very recent approach. Bhattacharya *et al.* [33] combined 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 in 2011. Kurtosis is a measure of signal peakedness (the tailedness of the distribution), with high values indicating lots of outliers and vice versa [34]. It can be a useful tool to extract information of outliers from raw audio data, making computations more efficient. The study of Bhattacharya *et al.* [33] was focused on welding deposition efficiency rather than specific defect detection, but there is a connection between deposition efficiency and certain classes of defects, making this a valued reference for further research. Alfaro *et al.* [35] studied the combination of acoustic and optical infrared signals along with welding current and voltage in short-circuit GMAW (GMAW-S) done on AISI 1020 steel plates using ER70S-6 filler wire. The 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 Kalman filter is a linear optimal state estimation theory based on Bayesian inference [52]). While numerical assessments of the extent of detection were not presented, the plots of quality parameters against different weld parameters do indicate that data fusion is a beneficial approach.

Vandone *et al.* [36] used an interesting approach of combining online signals (specifically melt pool images and tool position/trace) and offline data (a 3D scan of the actual deposited material which acts as the "ground truth") during the laser metal deposition of SS316L powder. The study focused on key process indicators like surface roughness and attempted to connect the predictions made using online parameters and the actual measured surface topography. The models used were able to adapt process parameters in about 50 ms based on surface profile deviations (quantified by a track degradation parameter), indicating the viability of such a system for automation of the printing process.

Xu *et al.* [37] similarly designed a monitoring system for WAAM consisting of sensors like a pyrometer, wire feed speed sensor, current sensor, shielding gas flow sensor, oxygen sensor, and laser profilometer. This study also used Kalman filters to denoise the obtained signals. Defect detection tests were not conducted, but data acquisition was successfully achieved in real-time, making this study more of a proof-of-concept rather than an actual defect detection test.

Hsieh *et al.* [38] conducted a multivariate time series anomaly detection study using a new unsupervised LSTM (long short term memory)-based autoencoder neural network. While the sensors used were not specified, the study did attain an anomaly detection accuracy of about 90% using their model. The authors indicated an important point that the lack of labeled data makes it difficult to use binary classification, since anomalies can be completely random and thus may have too many variations to individually identify and label. Thus, the approach taken in this study was to train the model on data indicating normal operation only. This way, by learning the time-based patterns present during normal printing operation, the model would become capable of detecting deviations from such patterns, making it theoretically possible to detect any kind of defect. Transfer learning was used to retrain the model on datasets containing anomalies, and the final accuracy of 90% was achieved for both precision and recall. Other models like convolutional neural networks (CNNs) were used for comparison, and those methods achieved accuracies in the range of 70-85% for the same data.

Reisch *et al.* [39, 40] conducted multiple studies on multisensor systems, initially using a combination of welding current sensor, arc voltage sensor and imaging, and later using a more comprehensive setup additionally including a spectrometer, microphone, structural acoustic

sensor, and gas flow sensor to detect various defects. The initial study in 2020 [39] compared three different neural network models (LSTM, Conv1D and Autoencoder) and achieved a minimum training error (called loss) on the order of 10^{-3}. The authors noted that a lot of optimization concerning hyperparameters (parameters which define the structure and working of a model) was required to achieve the results. They also noted that voltage data already provides a lot of information to detect anomalies in the welding process, but the addition of other kinds of data adds redundancy and thus increases monitoring reliability. The second study done in 2022 [40] took things a step further through the utilization of a "Digital Twin", which is a virtual representation of the printing setup. While the physical machine would provide sensing data, the digital machine would analyze it using models to assess the quality of the ongoing print. Bidirectional communication between the two entities would provide a system of continuous modification of parameters to correct for anomalies occurring during the printing process, as shown in Figure 6. The study found that intentionally introduced continuity defects could be rectified by the system using tool repositioning with a 93.4% success rate. It would be interesting to see such a system being applied to more complex defects which require optimization of multiple weld parameters simultaneously.

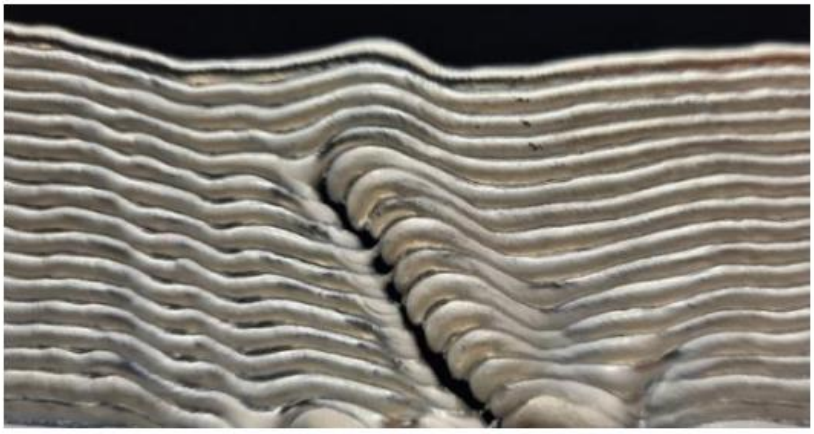


Figure 6: Correction of intentionally introduced discontinuity using bidirectional digital twin, Reisch et al. [40]

Cui *et al.* [41] studied the use of arc sound signals, welding current and arc voltage to track weld penetration in keyhole TIG (K-TIG) welding. 18 features were extracted from these three sensors (including kurtosis), and principal component analysis was used to extract the most important components for input data of training models. As mentioned before, mel frequency is used because it mimics the human ear response, thereby making use of the fact that weld process anomalies are usually audible to human ears in the form of arc sound discontinuities [31]. A support vector machine (SVM) with error correction and cross-validation was used to successfully classify different cases of weld penetration (too much, too little or normal) with an accuracy of 98.7%. Verification was done by modifying different parameters that affect heat input, which makes the model more reliable. Interestingly, the authors focused on *both* classification accuracy and classification speed, meaning that a balance would be necessary to be considered in the ideal model. In this paper, a new method called spectral noise subtraction (SNS) is used for denoising the acoustic signal.

Pringle *et al.* [42] designed an arc analyzer which could measure arc voltage, welding current, sound, light intensity, radio signals and temperature data. The wire feed speed was varied in order to destabilize the arc and induce extinction events in order to analyze the sensor responses. The

findings indicated that sound, radio signals and light were the most important sensors in context of arc anomalies since these signals showed the most variation. Current and voltage were found to be comparatively steady and non-indicative of any problems. Arc extinctions generally do show up in current and voltage signals, but this study seemed to indicate that such variations were less prominent as compared to those in other signals.

1.3 Machine Learning in WAAM

Nele *et al.* [7] used multilayer perceptron networks (MLPs) to analyze weld current, weld voltage and wire feed speed and classify welds having defects like lack-of-fusion. The results indicated a high classification accuracy of 97%, indicating high viability of MLPs in weld defect classification. The authors used a validated Welding Procedure Specification (WPS) as the baseline values for indicating a "good weld". The values obtained from the experiments conducted by the authors were compared with these baseline values, and the errors were utilized for neural network training. This indicates that having an idea of a "boundary value" of different welding parameters can be useful in order to train machine learning models better.

Rohe *et al.* [8] used two-dimensional convolutional neural networks (CNNs) to examine acoustic signals and detect irregularities. The results showed an overall accuracy of 84%, with some samples showing low accuracy due to noise. The obtained acoustic data was processed through a mel-band filter bank to simulate human perception and was subjected to a Fourier transform to convert it into a two-dimensional spectrogram, as shown in Figure 7. The performance indicates that CNNs might be a good idea for multidimensional datasets, with some further hyperparameter optimization.

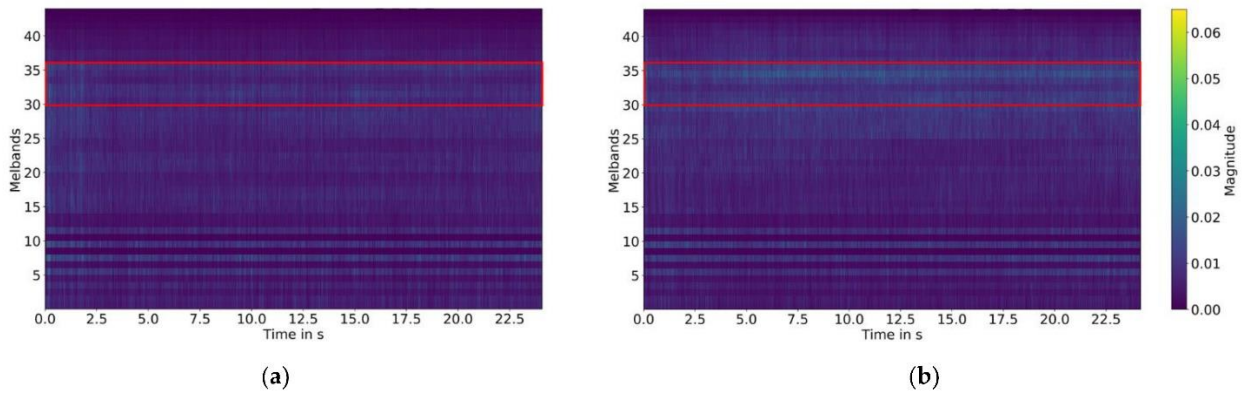


Figure 7: Mel acoustic spectrograms from GMAW WAAM of ER70S-6 wire with (a) no shielding gas (b) shielding gas flow rate of 15 L/min, Rohe et al. [8]

Similarly, Zhang *et al.* [9] used CNNs to evaluate inline weld images in pulsed GTAW of an Al alloy . A defect classification accuracy of 99.38% was achieved, indicating the suitability of CNNs for processing optical data. Combining this with the study of Rohe *et al.* [8], there exists the possibility of combining imaging and acoustic data to form a new input dataset which could potentially point out connections between the two classes of data. In addition, instead of filtering out bright spots in imaging data, the spots were converted into features that were input into the training model. Considering the reported high accuracy, it indicates that certain aspects of sensor data that are usually filtered out as noise could be re-examined to look for additional details regarding the

welding process. The authors pointed out, however, that their approach is limited to detecting surface defects since none of the sensors used were capable of depth scanning below the surface.

Li *et al.* [43] used the concept of incremental learning to monitor weld current and weld voltage to detect arc-on, arc-off and weld pool shift (narrowing of the melt pool due to impurities) events. Incremental learning is a different approach where the learning process never stops. The model is initially trained on a certain amount of data. Following this, each new print is evaluated for quality and incrementally added to the model for additional training. In this fashion, the model gradually learns better classification boundaries with increasing numbers of samples. In this study, a support vector machine (SVM) was used as the base ML model, and the results showed detection rates of above 90% for the abovementioned defects and required about 250 training instances/iterations to reach such accuracies. Such a system would be very interesting to use in an industrial application, since every printing job carried out by the machine can act as a training iteration for the model, essentially making it more and more accurate over time.

Xiao *et al.* [44] developed a novel machine learning framework based on a neural network structure that has the capability of inversely determining optimal windows of process parameters based on the desired weld bead geometry. The model was able to estimate bead dimensions to within 5% of the actual printed geometry and had prediction accuracies of over 95%. This approach was similar to the digital twin approach of Reisch *et al.* [40], since such a model was capable of taking feedback to improve upon its own results.

Kwak *et al.* [45] used reinforcement learning to analyze wire feed speed, travel speed and standoff distance in order to control bead geometry and temperature field. Reinforcement learning typically works based on a reward system, whereby the model iteratively learns the best courses of action based on the reward obtained by doing random actions. The specific type of reinforcement used was Q-learning, which is a method where a constantly updated quality matrix is used to determine the next action taken by the system. The results showed a convergence on a combined error of around 1.2 mm for bead height and width. The authors noted that this kind of learning method takes a lot of time/iterations, which is an indication that it might not be suitable for real-time applications. Additionally, it was noted that the deterministic nature of the model could make it unsuitable to handle random occurrences/disturbances.

Wu *et al.* [11] used an object detection model named YOLO, or *you only look once* (specifically YOLOv3) to analyze slag and hole defects on the weld surface via melt pool monitoring, as shown in Figure 8. The model achieved a mean average precision (mAP) of 98% at a rate of 59 frames per second (FPS), which is great for real-time monitoring. For image-based monitoring, the selection of an appropriate bounding box is important, since that determines the regions that are monitored for defects. In this study, a K means ++ algorithm was used that iteratively deduces the best bounding boxes for each image. The authors used some additional modifications like a focal loss (FL) function and a squeeze and excitation (SE) module to improve the performance of the YOLOv3 model. Focal loss is a loss function that can deal with class imbalances, *i.e.*, the over- or under-representation of certain classes of objects in a classification problem. In the scenario of defect detection, it is very common that datasets will have imbalances, since perfect welds are difficult to get as compared to defective welds (thereby making the dataset deficient in perfect samples), making the focal loss function a useful addition to the machine learning model used. Squeeze and excitation (SE) is a technique to allow a convolutional neural network to adjust its parameters for each feature separately. The advantage is that each feature can be assigned a certain level of importance which increases classification accuracy, all with a very minimal

increase in computational cost. Comparisons of the modified YOLOv3 model with the original YOLOv3 model in the study showed improved mAP along with improved FPS values.

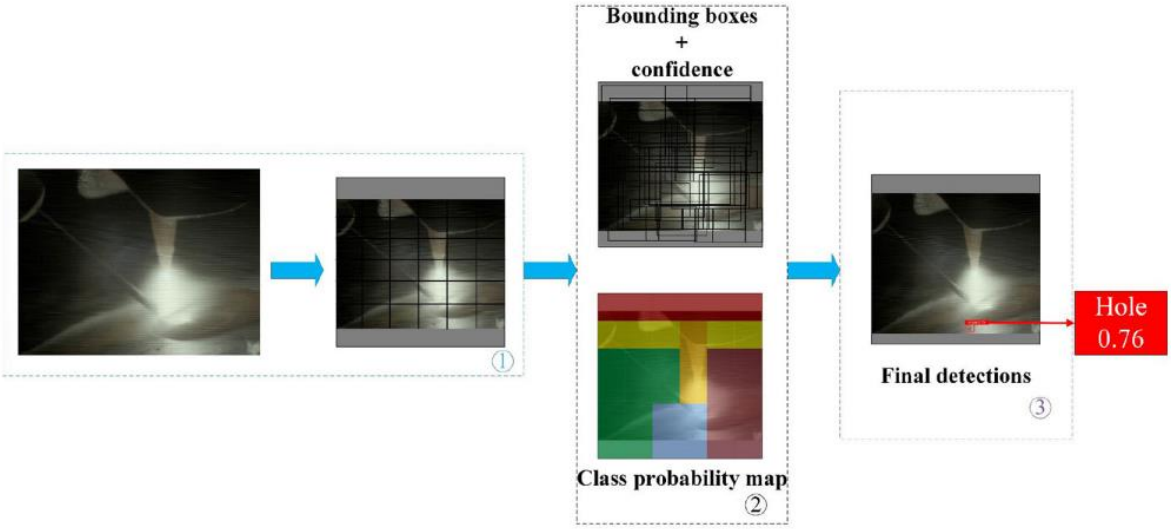


Figure 8: Depiction of YOLOv3 classification algorithm, Wu et al. [11]

Qin *et al.* [46] studied a single sensor response – weld voltage in order to classify the droplet transfer mode in WAAM. The voltage signal was analyzed to extract 11 features which acted as input to the ML model. The model used in this case was an LSTM (long short term memory) neural network, which is designed for time series data processing. The results indicated a defect detection accuracy ratio of over 90% for both single pass and oscillating walls. There was reasonable correlation found between all the voltage features and metal transfer mode (via the Pearson correlation coefficient). An interesting point to note is that a class imbalance could be observed in the dataset of this study, since the percentages of each transfer mode represented in the dataset were 28.05%, 50.17%, 18.96% and 2.82%. This indicates that the usage of a focal loss function akin to the one used by Wu *et al.* [11] might help improve the accuracy of the model.

A review of the usage of machine learning models in additive manufacturing conducted by Qin *et al.* [47] showed that deep learning models (which includes 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. An example of thermal image processing from the paper is shown in Figure 9. It is also interesting to note that support vector machines are fairly commonly used, whereas Gaussian processes are rarely used. This makes sense, considering that while Gaussian-based models can be highly accurate, they are computationally expensive, making them a less-desired choice for real-time applications. The authors noted, however, that deep learning models typically require a large amount of data to be useful, making it difficult to utilize in certain situations. Additionally, the paper pointed out the problem of many of these models being "black boxes" which are not physics-based. Some authors like Ko *et al.* [48] constructed knowledge-based models using graph-based networks for different additive manufacturing processes, indicating the possibility of a knowledge-based approach to investigate any potential accuracy improvements. Another review by He *et al.* [49] also corroborated the popularity of neural networks and support vector machines in the field of additive manufacturing. The authors additionally noted that support vector machines can be trained on small datasets to get reasonably good accuracies, giving a solution for dataset availability problems. It is further observed that feature extraction

should be carefully examined to ensure no loss of important markers for certain defects. This is especially relevant when looking at denoising since there is the potential for loss of important patterns in the removed "noise".

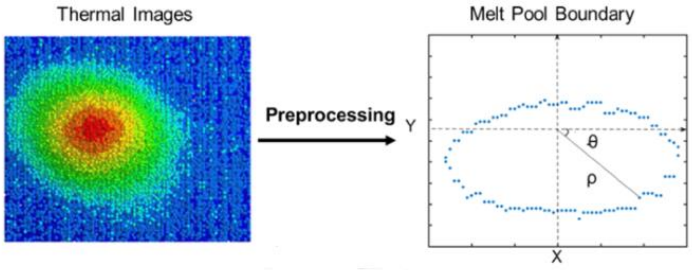


Figure 9: Converting a thermal image of a melt pool to a computer-usable contour using functional principal component analysis, Qin et al. [47]

To summarize, it is clear that the primary objectives to be satisfied for the inline usage of machine learning models in WAAM are high accuracy and low latency time. An optimum should be found between the two since high accuracy typically requires more computation, which increases latency time. Surveying the research around this area, it is seen that deep learning models are highly favored due to their ability to learn complex patterns with low computational cost and high efficiency. Different types of neural networks are designed for different kinds of data (for example, recurrent neural networks handle time series data, whereas convolutional neural networks higher-dimensional data), which means that there is high adaptability of the neural network approach for handling different varieties of sensor responses.

Support vector machines (SVMs) are also found to be popular for this application. Ding *et al.* [50] states that SVMs have numerous advantages over artificial neural networks (ANNs), including lower training times, better convergence on global solutions and lesser extent of overfitting. The authors also mention that studies comparing SVMs and ANNs in an experiment to predict weld penetration showed better results for the SVM. The high "customization" potential of neural networks for different kinds of data, though, means that they are the focus of this study.

1.4 Motivation and Thesis Structure

The research work in this thesis is inspired by the fact that the majority of literature works on ML-based anomaly detection have confined themselves to working on one material at a time. While such an approach does shed light on the nature of defect formation and its relation to welding parameters for the particular material, it limits the utility of such a detection paradigm on an industrial/practical level.

In addition, it is very interesting to consider the similarities between different classes of materials when attempting to characterize the formation of a particular type of defect. This can be extended to looking at different defect classes altogether. If such commonalities can be found, that paves the way for a universal anomaly detection algorithm, which would be of great interest.

Finally, it is important to consider how the process of data collection for additive manufacturing can be a costly and time-consuming process. Keeping this in mind, the likelihood of class

imbalance (whereby there is a high discrepancy between the number of datapoints of clean and defective beads) in the obtained datasets for such research is fairly high. The impact of this imbalance is important to understand in order to design data-collection experiments more efficiently for future work.

Thus the research questions being investigated in this thesis can be summarized as shown below.

1. *Is it possible to make predictions on defect formation during the WAAM deposition of one material (say, Material A) with an ML model that is trained on the defect formation behavior of another material (say, Material B)?*
2. *Is the usage of welding voltage, welding current and audio response sufficient to capture defect formation behavior during WAAM deposition? If not, what other parameters might be necessary and why?*
3. *How severe is the effect of class imbalance on the quality of predictions made for anomaly detection in WAAM? Does the effect vary depending on which class comprises the majority of the dataset?*

This thesis begins with an introduction to the WAAM process and the defects most commonly seen in it. This includes a review of the state-of-the-art in current research on multisensor-based defect detection and machine learning approaches in WAAM. This is followed by the experimental protocols used in this research work, including the hardware, materials used and methodology of data evaluation. The machine learning models used for analysis are then discussed, along with the methods used to optimize their hyperparameters.
Once the experimental conditions have been discussed, the results are presented. The physical detections of defects in the micrographs are presented, followed by data feature observations. This is followed by the machine learning results, along with the evaluation metrics used for the same. The three models used for the machine learning approach are all discussed along with their performance on each material. Following this, cross-material analyses done with neural network and random forest models are discussed using the same error metrics.
Finally, the welding input parameter trends are analyzed to determine the physical reasons behind the observations made. The possibility of expanding the approach to other classes of alloys is discussed and the thesis concludes with a summary of the results and recommendations for future work.

2 Experimental Details and Data Processing Methodology

This chapter describes the experimental aspect of this research work. All facets ranging from the printing hardware to the details of materials used for printing are discussed. The processing pipeline for extracting machine learning input parameters from the various sensor responses is also discussed.

2.1 Experimental Setup and Materials

2.1.1 Experimental Machinery

All printing experiments were conducted at RAMLAB BV on a Panasonic TM1400 WGIII Robot. A picture of the robot is shown in Figure 10.

Figure 10: Panasonic TM1400 WGIII robot used for printing experiments

The robot has inbuilt current and voltage sensors that are capable of logging current and voltage signals during the printing process 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 was attached to the welding arm. An attachment in this fashion would ensure zero relative velocity between the microphone and weld tip, 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 also shown in Figure 10.

In some deposition experiments, thermocouples were spot-welded to the base plate, and in some cases, to the wall itself in order to monitor the temperature of the build. This was utilized in prints where the interpass temperature had to be kept within a certain range in order to avoid thermal defects like hot cracking. The thermocouples were connected to a Picolog datalogger which was connected to a laptop in order to record temperature variation with time. An example of one such thermocouple connection is depicted in Figure 11.

Figure 11: Thermocouple connection to a WAAM print

2.1.2 Materials Used in Study

All materials were designed as consumable wires of 1.2 mm diameter. The wires were sourced from Voestalpine Böhler Welding.

2.1.2.1 Inconel 718

Inconel 718 is a nickel-based superalloy that is designed for high-temperature corrosion-resistant applications that was developed by Special Metals Corporation [55]. The datasheet suggests that the performance of welded joints of this superalloy is good, indicating the suitability of processing this material using wire arc additive manufacturing [55].

The composition of Inconel 718 is shown in Table 1 [55].

Table 1: Chemical composition of Inconel 718

Element	Fe	Ni	Cr	Nb	Mo	Ti	Co
Weight Percentage (%)	17	50 - 55	17 - 21	4.75 - 5.50	2.80 - 3.30	0.65 - 1.15	≤ 1

One of the most prominent issues encountered during the printing/welding of Inconel 718 is *hot cracking* [56]. This includes both solidification and liquation cracking. Research suggests that one of the main reasons for hot cracking is the formation of secondary phases (like γ' and γ" phases) throughout the matrix [56]. This has been observed to reduce the tensile performance of IN718 by 40 – 60%.

2.1.2.2 Invar 36

Invar 36 is an iron-nickel alloy that is well-known for a very low thermal expansion coefficient (1.2×10^{-6} K^{-1}), which is about a tenth of that of carbon steel (11.7×10^{-6} K^{-1}) [57]. It was

discovered by CE Guillaume, for which he received the Nobel Prize in Physics in 1920 [58]. The composition of Invar 36 is shown in Table 2 [57].

Table 2: Composition of Invar 36

Element	Fe	Ni	Mn	Si
Weight Percentage (%)	63	36	0.35	0.20

When considering the printing of Invar 36, the biggest problem encountered is the low heat transfer coefficient. Experiments conducted during the course of the research indicated that high input energy per unit length or too small print dimensions resulted in the viscosity of the melt bead being too low and cooling being too slow, leading to the bead flowing off the surface and generating a misshapen structure.

2.1.2.3 Inconel 625

Inconel 625 is a nickel-based superalloy that is designed for high-temperature high-strength applications that was developed by HL Eiselstein and DJ Tillack [59]. Like Inconel 718, the performance of welded joints of this superalloy is also very good, indicating the suitability of processing this material using wire arc additive manufacturing [60].

The composition of Inconel 625 is given in Table 3 [60].

Table 3: Composition of Inconel 625

Element	Fe	Ni	Cr	Nb	Mo	Mn	Si
Weight Percentage (%)	≤ 5	≥ 58	20 – 23	3.15 – 4.15	8 - 10	≤ 0.5	≤ 0.5

Similar to Inconel 718, the primary issue observed during the printing/welding of Inconel 625 is hot cracking [79, 80]. Research indicates that grain boundary aggregation of carbides can generate eutectic phases (specifically the Laves/γ eutectic) which is low-melting and can form cracks under residual stress [79].

2.2 Experimental Details

Throughout the course of data collection for this thesis, two different welding modes were used, which are listed below.

1. *Super Active Weld Process (SAWP)*
2. *Pulsemix Mode*

SAWP is a welding process developed by Panasonic which functions using repeated short circuiting to stabilize the arc and minimize spatter [53]. A typical super active weld process waveform of the voltage is as shown in Figure 12. The waveform together with the back-and-forth wire movement allows material to transfer through a stable short arc.

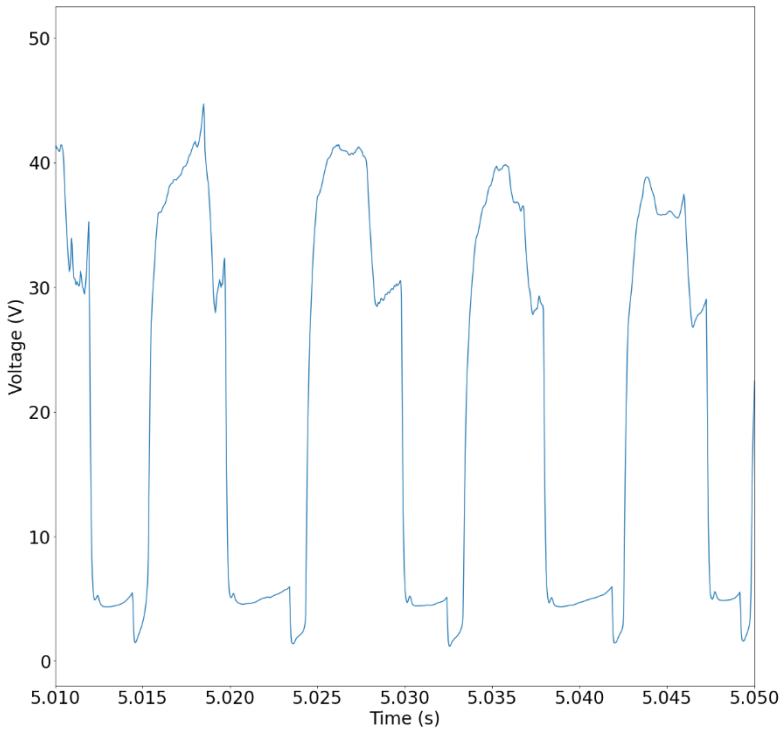


Figure 12: SAWP typical waveform of voltage vs. time plot

Pulsemix mode is a welding process that is a combination of pulsed welding and short-circuit welding. A typical pulsemix weld process waveform of the voltage is shown in Figure 13. Pulsemix welding allows for lower heat input as compared to traditional pulsed welding, which can be useful for the deposition of materials with high thermal sensitivity of defects or low thermal conductivity.

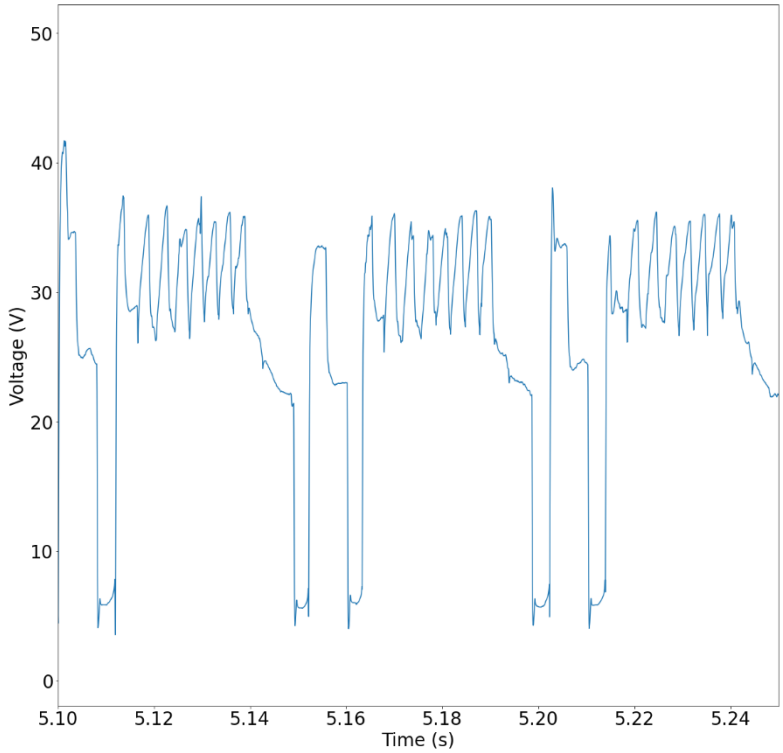


Figure 13: Pulsemix typical waveform of voltage vs. time plot

Different experiments were performed which have been categorized as follows.

1. *Ramp Tests* – It is a single bead deposition with one parameter (like voltage, for instance) being varied along the length of deposition while keeping other parameters constant (like wire feed speed, current, travel speed, etc.). Ramp tests allow for a quick navigation of desired process parameters [108].
2. *Single Beads* – It is a single bead deposited using selected process parameters remaining constant throughout its length. Weaving motion was used to deposit a "wider" bead for some experiments.
3. *Single Bead Walls* – It is a wall made with single beads deposited on top of each other in a bidirectional fashion.
4. *Block Walls* – It is a wall with overlapping of multiple beads, *i.e.*, structures printed layer by layer, where each layer consists of more than one bead
5. *Pyramids* – It is an analogue of a block wall, with each layer from the bottom consisting of two less beads than the previous. It is used to evaluate the material response to the process when deposited on the feed material itself. Due to the similarity between these and block walls, they can be considered as one category.

Due to the large number of experiments, the processing parameters for each test are listed in Appendix A. The following tables only summarize the type of experiment performed for each material in this research work.

The details of the Inconel 718 tests have been summarized in Table 4.

Table 4: Types of experiments performed using Inconel 718 wire

Sl. No.	*Test Type*	*Total No. of Beads Deposited*	*Remarks*
1	Ramp Tests	10	• None
2	Single Beads	25	• Includes weaving single and layer beads
3	Single Bead Walls	0	• None
4	Block Walls/Pyramids	413	• Includes beads deposited for preheating the base plate • Includes weaving block wall

The details of the Invar 36 tests have been summarized in Table 5.

Table 5: Types of experiments performed using Invar 36 wire

Sl. No.	*Test Type*	*Total No. of Beads Deposited*	*Remarks*
1	Ramp Tests	6	• None
2	Single Beads	11	• None
3	Single Bead Walls	138	• None
4	Block Walls/Pyramids	24	• No preheating done

The details of the Inconel 625 tests are summarized in Table 6.

Table 6: Types of experiments performed using Inconel 625 wire

Sl. No.	*Test Type*	*Total No. of Beads Deposited*	*Remarks*
1	Ramp Tests	0	• None
2	Single Beads	6	• Includes layer beads
3	Single Bead Walls	0	• None
4	Block Walls/Pyramids	652	• Includes beads deposited for preheating the base plate

2.3 Input Data Evaluation

2.3.1 Voltage Data Processing

In order to analyze voltage signals for defect detection, it was imperative that the monitoring paradigm be independent of the weld mode. This would increase the applicability across different welding operations.

The simplest possible operation would be to take a batchwise average of n points in order to eliminate the effects of pulse shape and only show value-based variations. While this approach is beneficial in terms of computational cost, it is too reductive and leads to the loss of information about individual pulse abnormalities (which can, for instance, occur due to wire feed variations caused by excessive friction in feeding tubes). To illustrate this, a bead of Invar 36 is taken as an example. This bead was deposited as part of a series of experiments to determine the ideal weld parameters for single bead-width wall deposition of the alloy. A picture of the bead is shown below in Figure 14.

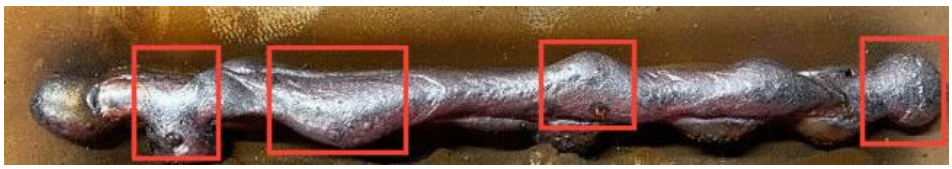


Figure 14: Invar 36 bead with irregularities caused by bead overflow (superactive mode, 15.6V, 0.25 m/min travel speed)

The red rectangles indicate regions of the bead where clear deviation is seen from the straight line along the print direction. Now the average voltage for this bead can be examined (the batch size is taken to be 220 points for each average) in Figure 15.

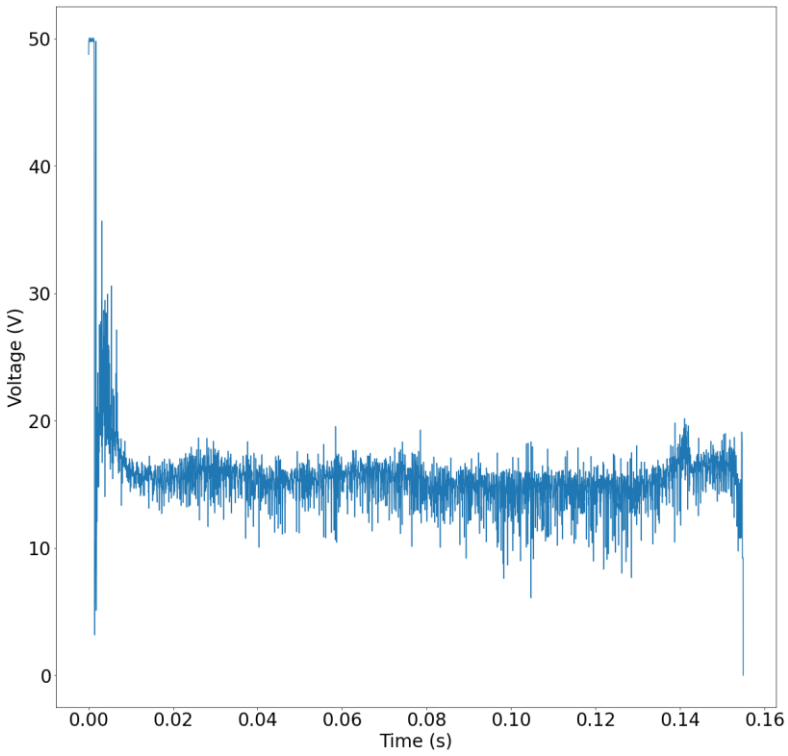
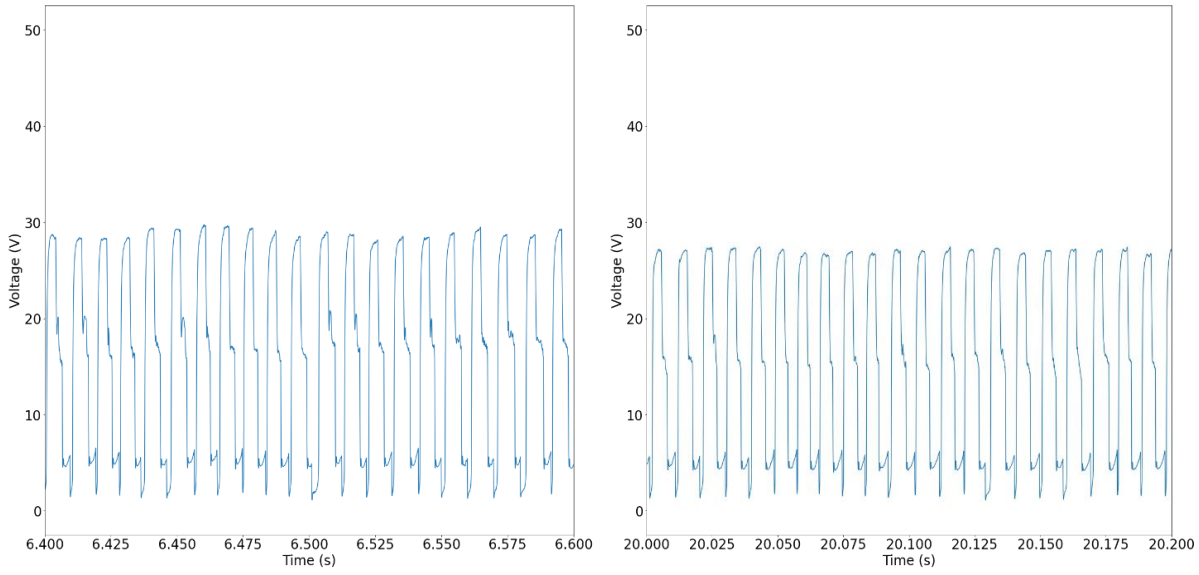


Figure 15: Average voltage of Invar 36 bead in Figure 14

It is clear that variations specific to the four marked regions in Figure 14 are not visible in this plot. While there is a small increase in variance in the plot towards the end, this is not seen to correspond directly with the anomalies in the observed bead. Thus, simple averaging was discarded and an alternative parameter was investigated.

Since it was postulated earlier that individual pulse variations can play a role, the individual pulses in different regions of the weld bead were compared. Two such regions are shown below in Figure 16.



(a) (b)

Figure 16a-b: (a) Voltage pulses from deviant section of the bead depicted in Figure 14, (b) Voltage pulses from straight section of bead

While the crests and troughs of the pulses appear consistent across the two sections, it can be noticed that many of the pulses in the deviant section have secondary peaks, whereas very few of the pulses in the straight section do. This means that a tracking parameter derived from voltage needs to display this feature.

A straightforward way of finding the number of such peaks within a single pulse is to track the change in voltage between consecutive points (basically, tracking ΔV). A certain threshold value of ΔV can be defined above which the variation is deemed to be abnormal. Having multiple peaks would lead to more occurrences of large values of ΔV as compared to a smoother pulse. Counting the number of instances of excessive ΔV for each bead can lead to plots that potentially show variations directly corresponding to visible deviations.

The plot of the number of excessive ΔV instances ($n_{\Delta V}$) for the bead in Figure 14 is shown below in Figure 17.

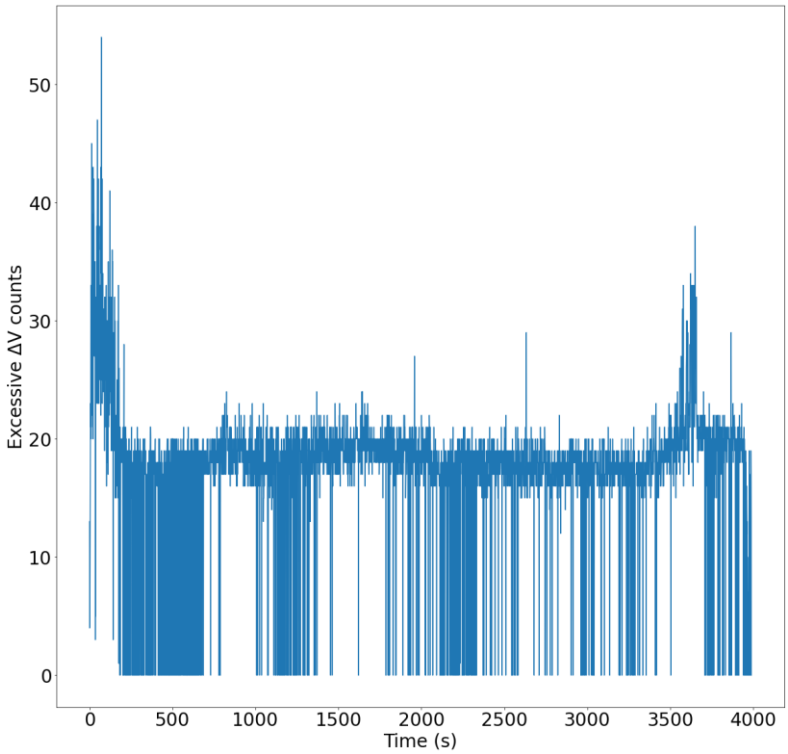


Figure 17: $n_{\Delta V}$ of Invar 36 bead in Figure 14

From the outset, this plot shows significantly more variation than the average voltage plot in Figure 15. Since the objective is to see regions of high ΔV counts, the regions where the plot drops to 0 can be considered as non-defective regions. It is visible that there are clear regions where the ΔV is high, and these roughly correspond to the deviant regions in the bead. While the correspondence is not perfect, modifications in the threshold value can likely improve the results. This indicates that this parameter is a useful tracking tool for voltage.

It is also visible in Figure 17 that the beginning and end portions of the plot show maxima corresponding to the arc start and arc stop. Both these phenomena are expected to be relatively unstable compared to the remaining part of the printing process (the arc stop section can be somewhat stabilized using good crater filling conditions like increasing crater weld time, but a certain level of instability is always expected). This means that any calculations attempting to quantify the overall variance of ΔV must take the arc start and arc stop into account in order to prevent an erroneously inflated value. Examination of the variance values indicated that the inflation induced by arc start and stop is small enough to ignore, since the duration of both events is very small.

The next step is normalization. Considering that different materials require different voltages for proper deposition, it is natural to expect that the average ΔV count would vary significantly. This can affect the magnitude of deviation/variance seen, which can give erroneous results for a model that should work across a wide range of materials. To eliminate this issue, the mean of the plot is subtracted from each point, giving the plot in Figure 18 (approximately centered around the zero line).

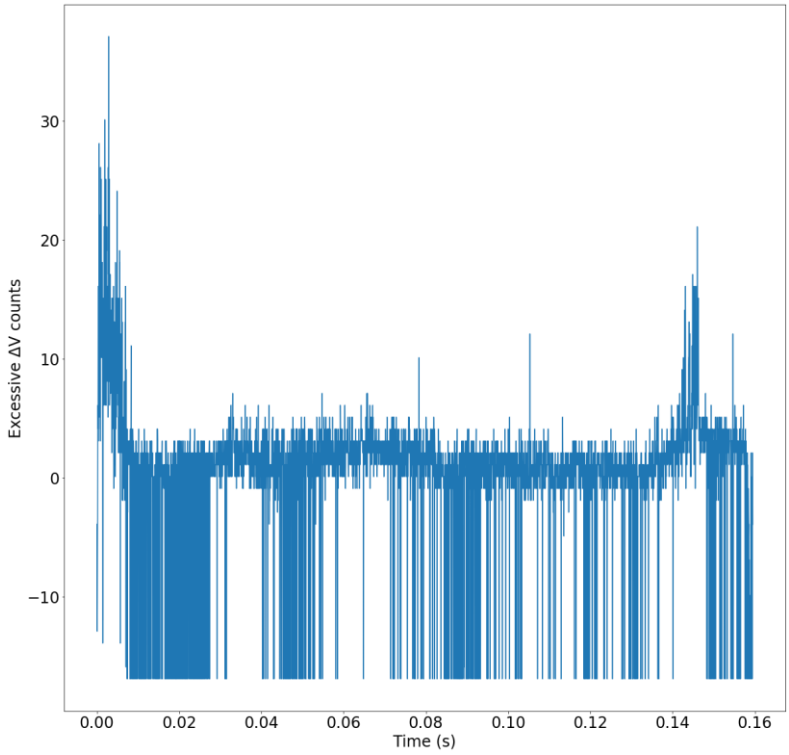
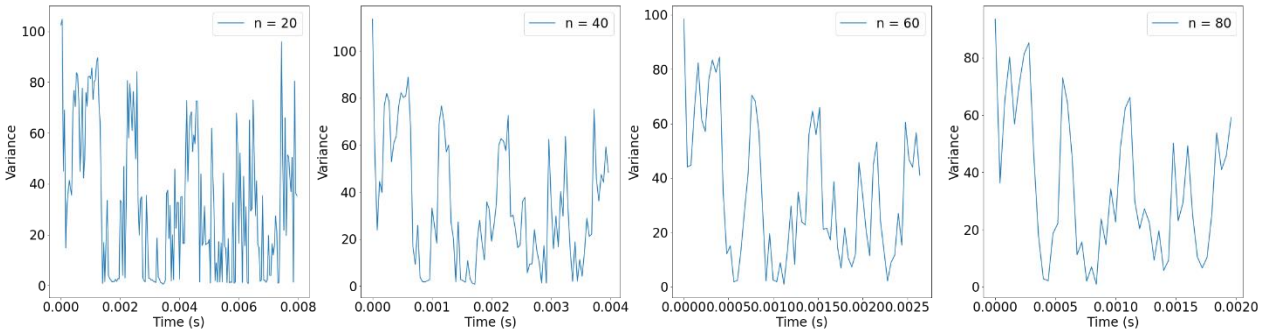


Figure 18: Normalized $n_{\Delta V}$ of Invar 36 bead in Figure 14

In order to quantify the deviations obtained in Figure 18, a simple method would be to slice the data into chunks of n elements and calculate the variance of each chunk individually. The chunk-wise variances obtained for different values of n are shown in Figure 19.



(a) (b) (c) (d)

Figure 19a-d: Chunk-wise voltage variance for Invar 36 bead in Figure 14 – (a) n = 20, (b) n = 40, (c) n = 60, (d) n = 80

For the smallest value of n, it is clear that there is too much variation to detect any useful regions of deviation. As the value gets higher, however, it is visible that three distinct peak regions are visible (ignoring the arc start and arc stop). Comparing this to Figure 14, it is possible to correlate these regions to the marked deviant portions of the bead.

As clearly seen, the usage of $n_{\Delta V}$ can provide height-based information of voltage, whereas width-based data is still missing. While the width of each pulse may have a weak correlation with $n_{\Delta V}$ (simply put, a pulse having a larger number of secondary peaks would be likely to be wider to accommodate those peaks), it is still important to have a dedicated width parameter in order to ensure better connection between the sensor response and defects.

The calculation of pulse width is fairly straightforward and involves the definition of a threshold value which acts as the midpoint of the pulse crest and trough (somewhat akin to calculating FWHM, or full width at half maximum). Counting the number of points between consecutive crossings of such a threshold can give the width of each pulse. It is possible to normalize the obtained point number immediately using sample rate (thereby giving pulse width in time units), but it is not necessary.

The determination of a threshold can be realized in two ways – one would be to select a conservative constant value threshold based on the data that has been collected, while the other would be to dynamically decide the threshold as an average of the maximum and minimum values of the entire dataset. Each approach has the potential for errors, but examination of the data indicated that the errors from using a constant threshold would be rare enough to be statistically insignificant. One such instance of peak width detection using a constant threshold of 15V has been portrayed in Figure 20.

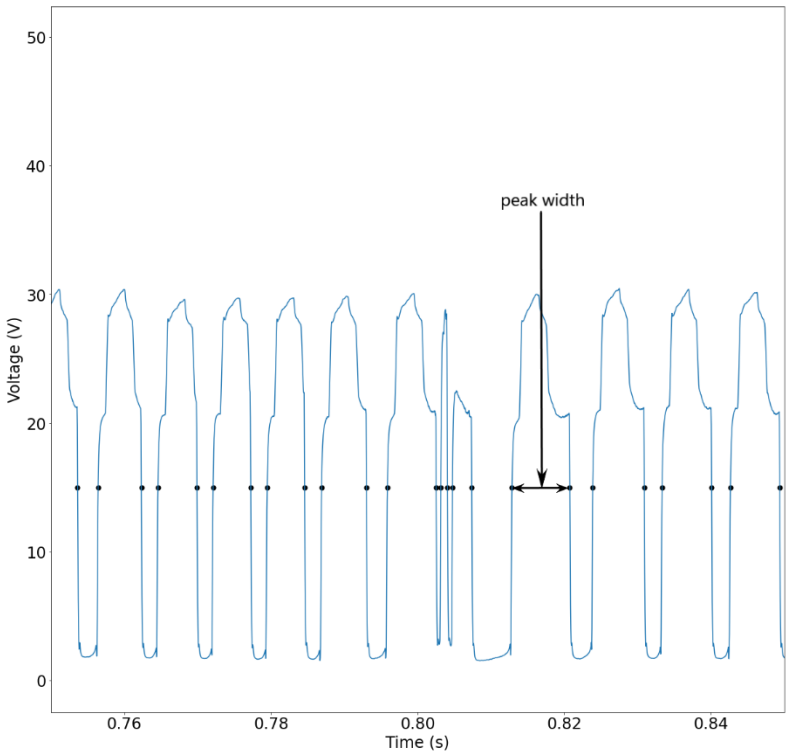


Figure 20: Peak width measurement using constant voltage threshold of 15V

In this fashion, both dimensions of the pulses can be considered, ensuring a better chance of identifying abnormalities.

2.3.2 Current Data Processing

As mentioned in the Section 2.3.1, two different modes of welding were used during this thesis. Comparing the current waveforms of the two modes can also be helpful for the analysis. The typical variations in the current waveform for a superactive weld process is provided in Figure 21.

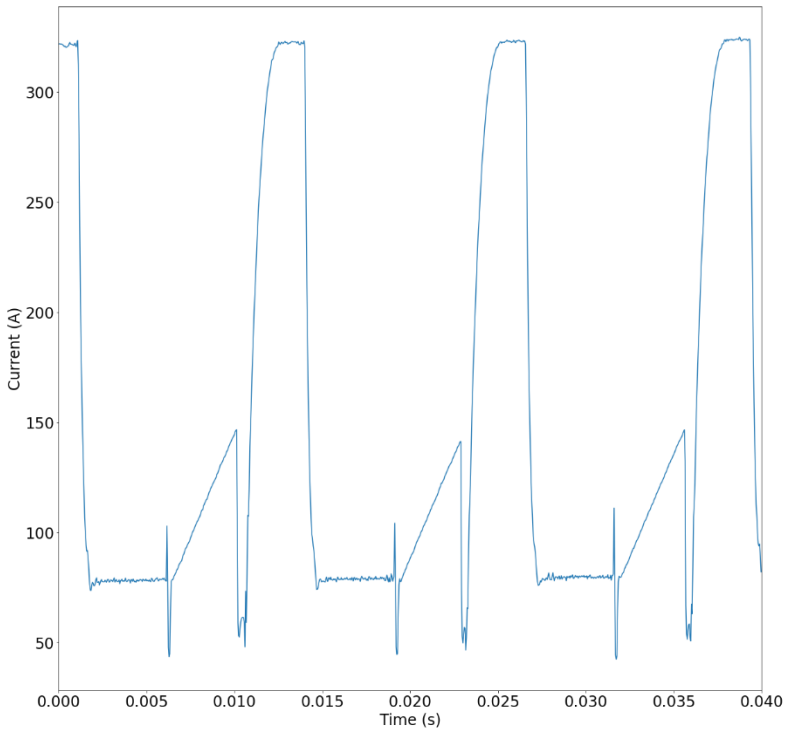


Figure 21: Superactive current vs. time plot

In comparison, the current variation in pulsemix mode is presented in Figure 22.

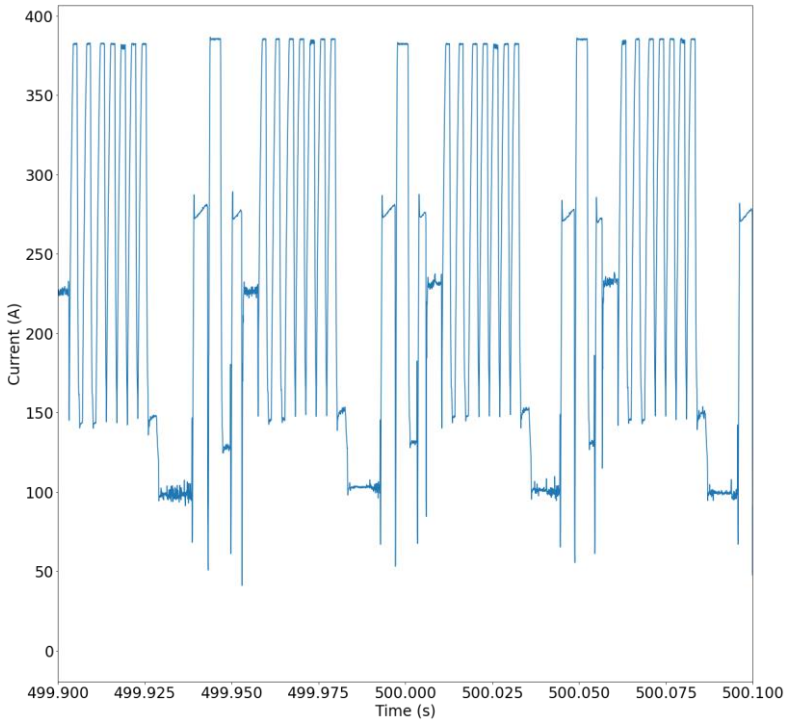
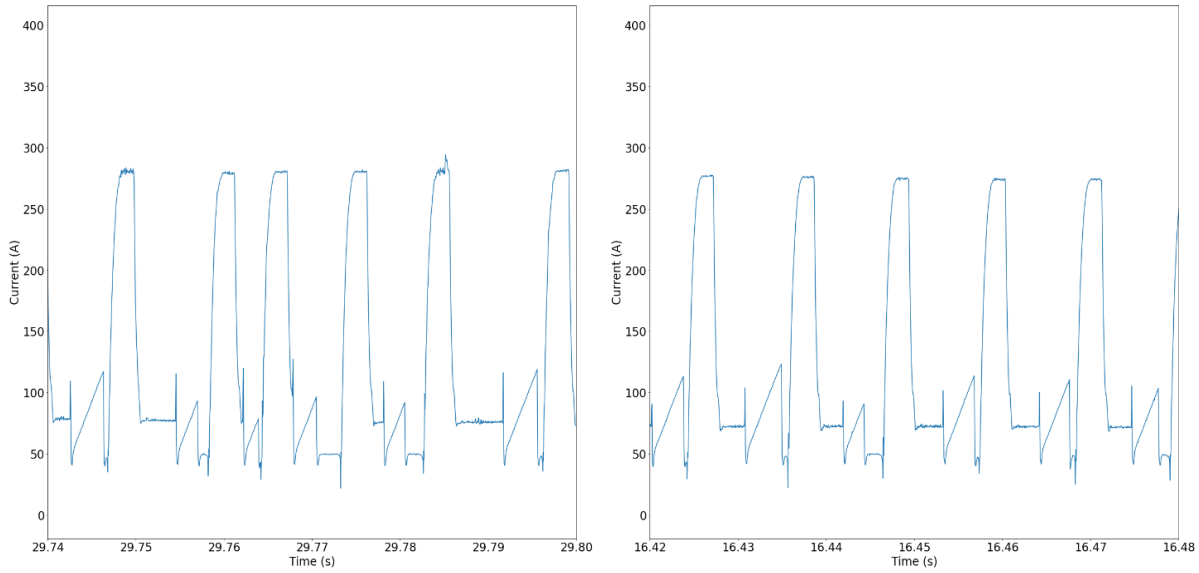


Figure 22: Pulsemix current vs. time plot

Similar to the analysis of the voltage signal, the next step is to compare pulses from different sections of a bead to identify signature features to track. Pulses from two such sections are shown in Figure 23.



(a) (b)

Figure 23a-b: (a) Current pulses from deviant section of bead depicted in Figure 14, (b) Current pulses from straight section of bead

It is visible that the deviant sections have pulses with secondary peaks whereas the straight sections have pulses that are more consistent in shape. Considering the similarity with voltage pulses, it stands to reason that an analogous parameter can be used to track current deviations (in this case, $n_{\Delta I}$). In this case, another bead of Invar 36 produced in superactive mode can be examined, which is shown in Figure 24.

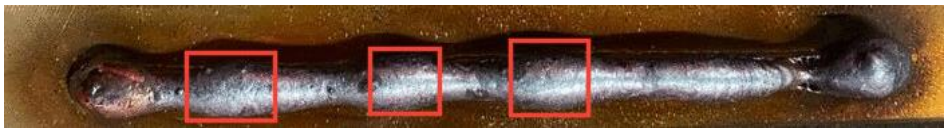


Figure 24: Invar 36 bead with overflow defects (superactive mode, 12.8V, 0.25 m/min travel speed)

The plot of excessive ΔI instances ($n_{\Delta I}$) for the bead in Figure 24 is shown below in Figure 25.

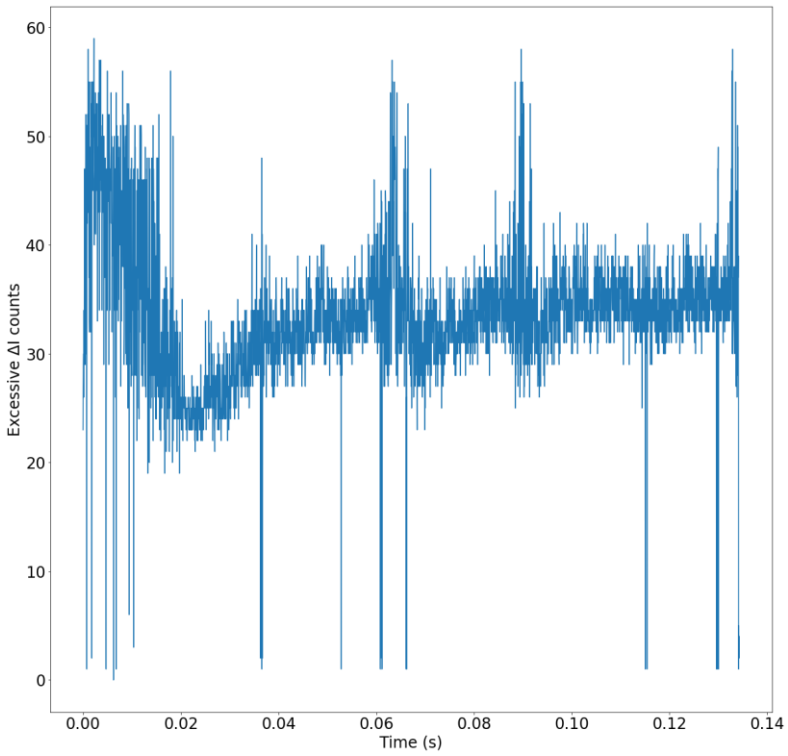
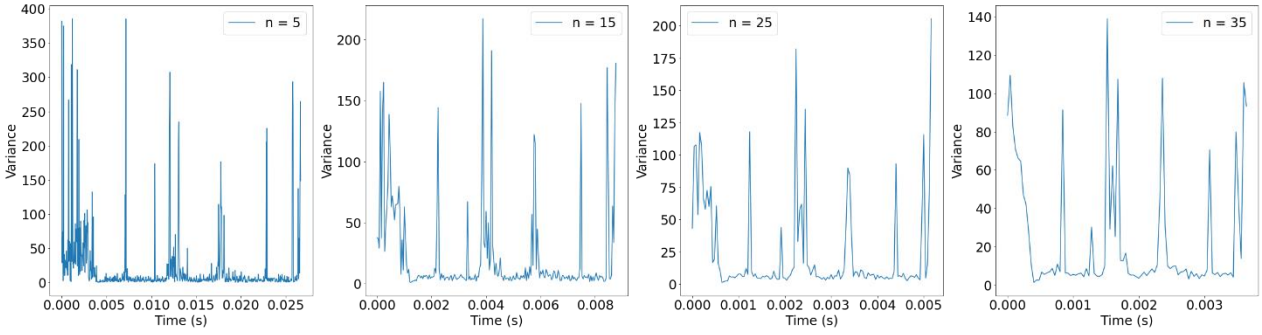


Figure 25: $n_{\Delta I}$ of Invar 36 bead in Figure 24

Ignoring the arc start and arc stop portions, it is visible that there are two large regions of deviation, with some more minor deviations present. These two major regions could be correlated to the two central rectangles marked in Figure 24, indicating the viability of current to be utilized in a manner similar to voltage.

The same normalization procedure as for the voltage analysis is applied here, and a chunk-wise variance can be obtained, as shown in the plots of Figure 26.

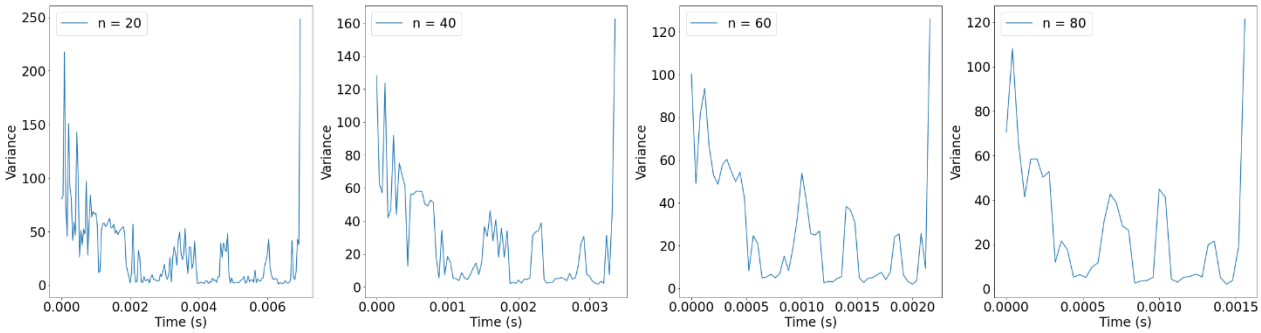


(a) (b) (c) (d)

Figure 26a-d: Chunk-wise current variance for Invar 36 bead in Figure 24 – (a) n = 5, (b) n = 15, (c) n = 25, (d) n = 35

Interestingly, the variance plots show multiple peaks, indicating more arc disturbances than expected. Similar to the voltage, it is also seen that too low values of n leads to excessive "noise" which impedes identification of defects, and too high values of n leads to washing out of potential minor defects.

For comparison, it is useful to look at the corresponding chunk-wise variance plots for the voltage of the same bead. This is visualized in Figure 27.

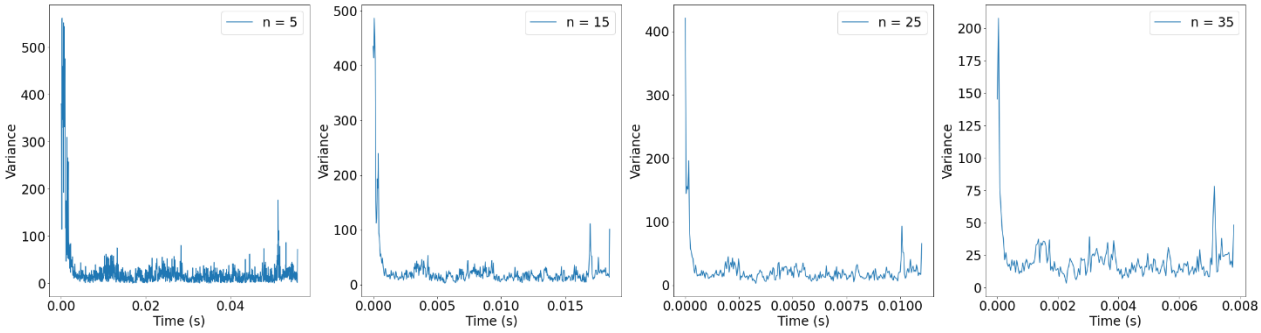


(a) (b) (c) (d)

Figure 27a-d: Chunk-wise voltage variance for Invar 36 bead in Figure 24 – (a) n = 20, (b) n = 40, (c) n = 60, (d) n = 80

It can immediately be seen that the three peaks towards the end of the plots are corresponding well to the defective portions of the bead. In addition, there is a significant region of high variance towards the start that could mean additional defects alongside the arc start.

Additionally, it is useful to examine the current plots for the bead in Figure 14 to compare the results with the voltage plots. The chunk-wise voltage variance plots for the bead in Figure 14 are shown in Figure 28.



(a) (b) (c) (d)

Figure 28a-d: Chunk-wise current variance for Invar 36 bead in Figure 14 – (a) n = 5, (b) n = 15, (c) n = 25, (d) n = 35

It is visible that there are two slightly elevated regions towards the beginning of the bead (ignoring the arc start and arc stop). These could correspond to a couple of marked regions on the bead, but it is clear that for this example, the current does not give as much information about the bead as the voltage does.

Similar to voltage, width-based information is also necessary to examine. Once again, either a static or a dynamic threshold can be selected, and an example of a static threshold of 150A has been shown in Figure 29.

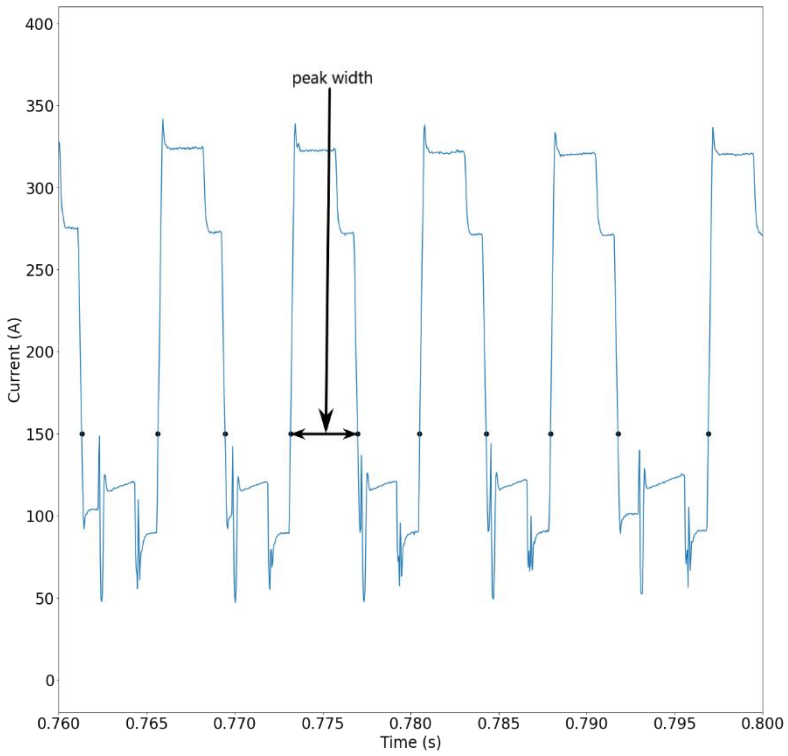


Figure 29: Peak width measurement using constant current threshold of 150A

Considering the significant similarities between the behavior of current and voltage, it is possible to use the parameters deduced from both in conjunction for conducting predictions on defect occurrence.

2.3.3 Audio Data Processing

The acquisition of audio data for the welds produced during the course of this study required the attachment of a Devine M-Mic USB BK microphone to the welding robot arm. The integration of the audio data channel to the backend of the sensor hub required some time, which means that audio data was available only for a certain portion of the prints made during the experimental period.

The first step is to denoise the signal. There are different methods of denoising mentioned in literature, and the method used in this study was *spectral gating*. This method is a type of *noise gating*, which splits the spectrogram of the audio data into frequency bands and applies noise thresholds to each band. Bands that are lower than the threshold get eliminated, which essentially removes noise from the signal. Controlling the threshold can give varying levels of noise reduction which can be finetuned for specific scenarios.

To visualize this, the audio signal for the bead shown in Figure 14 can be examined. The raw audio data for this bead is shown in Figure 30. It is evident that there are no discernible “hotspots” of deviation that might indicate problems in the bead. A simple spectral gating-based denoising can be performed with Python using a threshold value of 1 (according to the scale used by the denoising library), leading to the plot shown in Figure 31.

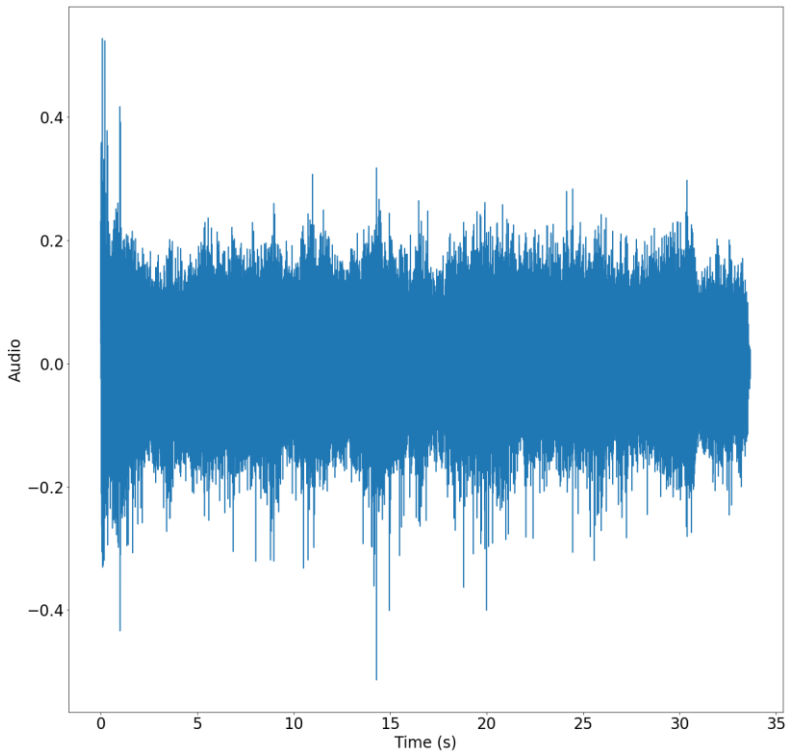


Figure 30: Raw audio signal of the Invar 36 bead shown in Figure 14

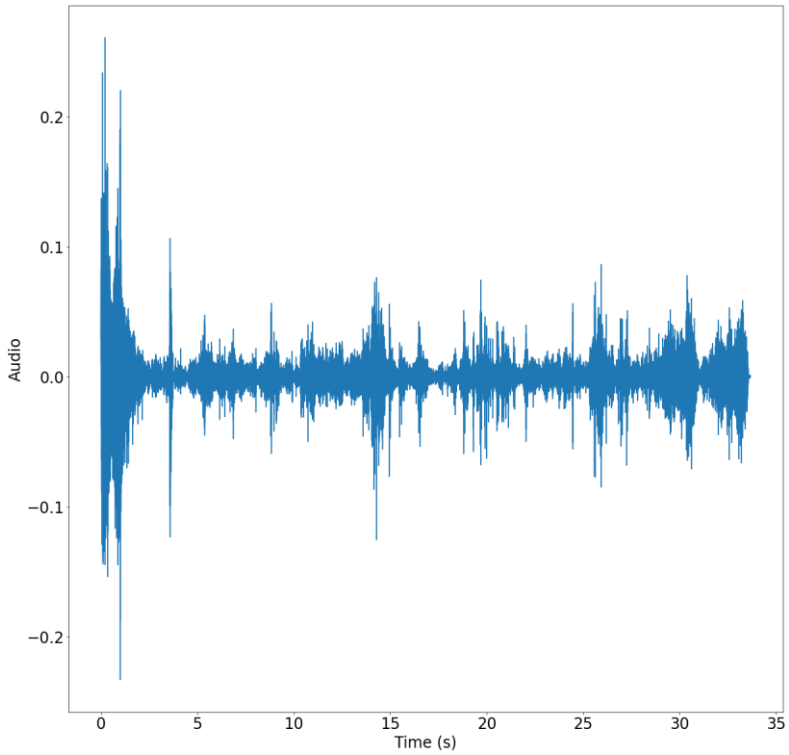
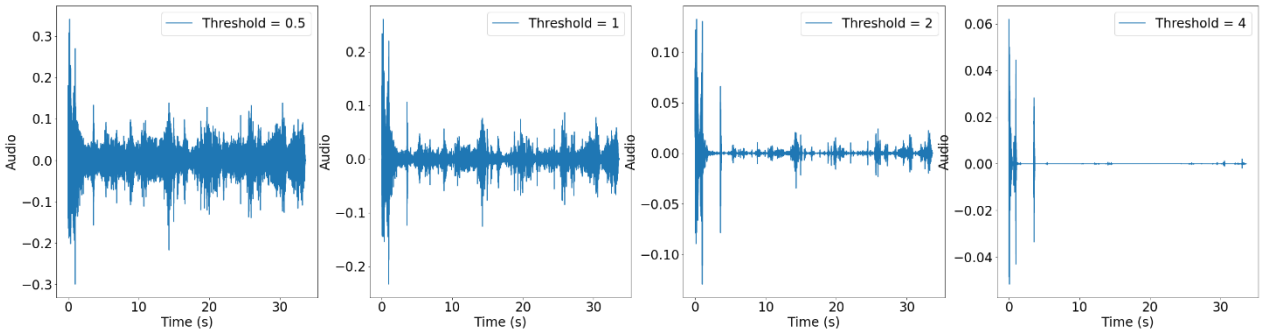


Figure 31: Spectral gating-denoised audio signal of Invar 36 bead in Figure 14 (threshold = 1)

This signal already looks a lot cleaner, with marked regions of increased amplitude. Two to three regions of interest can be identified (excluding the regions of arc start and arc stop) at locations

that roughly correspond to the deviant sections marked in Figure 14. It is possible to tweak the value of the threshold to find the best results. A few examples are visible in the plots of Figure 32.



(a) (b) (c) (d)

Figure 32a-d: Denoised audio signal for Invar 36 bead in Figure 14 – (a) Threshold = 0.5, (b) Threshold = 1, (c) Threshold = 2, (d) Threshold = 4

As seen in Figure 32a-d, keeping the threshold too low leads to insufficient denoising, *i.e.*, too much noise to delineate instabilities sufficiently. Excessively high thresholds lead to filtering out all but the most severe disturbances (which, in most cases, is just the arc start and arc stop), which is also undesirable. Thus, a threshold value between 1 and 2 is considered to be most viable, with a value of 1.6 being chosen for this study.

The next step is to convert this one-dimensional audio data into a spectrogram in frequency-time space. Rather than using a simple linear frequency scale, the different frequency bands in the spectrogram are represented using the *mel scale* [54]. 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.

Construction of a mel spectrogram of the filtered signal is done and shown in Figure 33. The arc start and arc stop regions are clearly visible at the beginning and end of the spectrogram. Apart from these, a thin region of high intensity is seen in the region between 0-512 Hz at time around 0.08 seconds. To get a better picture, the same kind of spectrogram for different values of threshold can be examined in Figure 34.

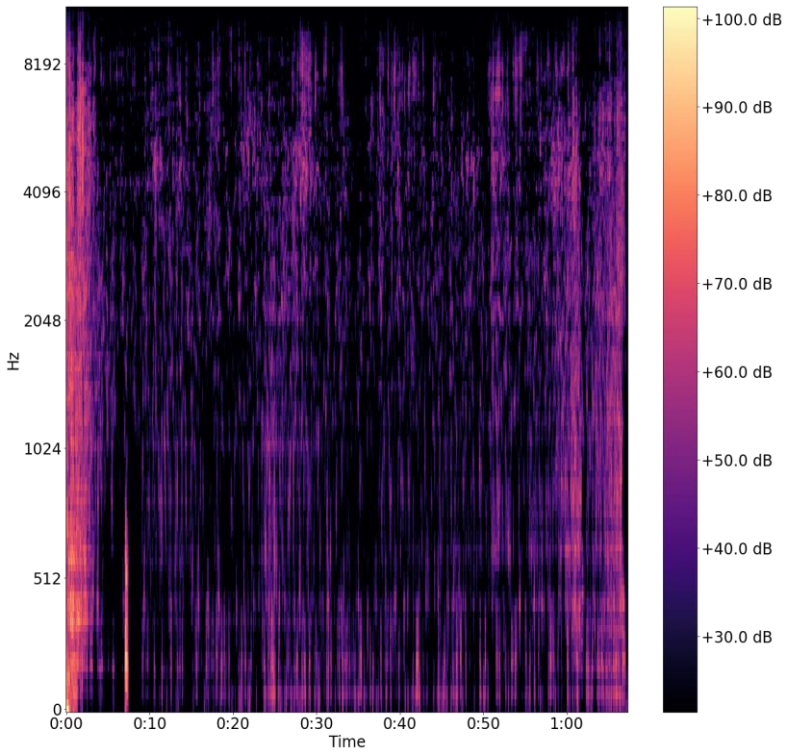
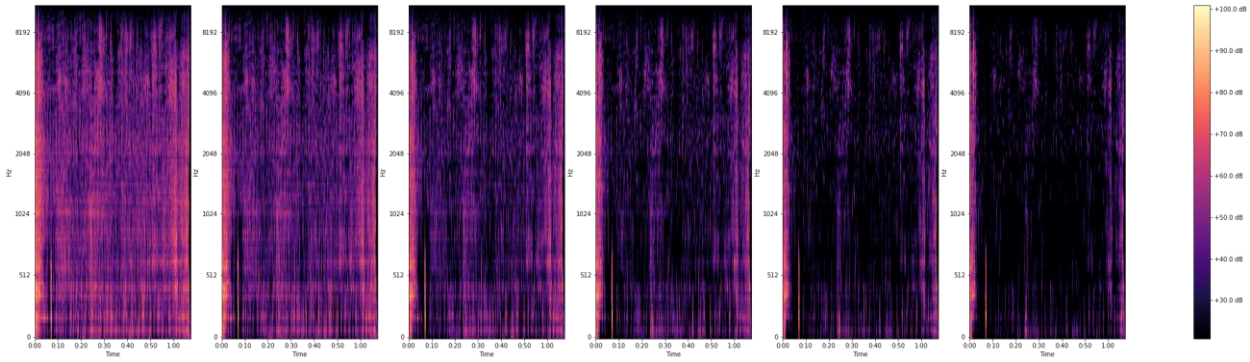


Figure 33: Mel spectrogram of filtered audio signal of the Invar 36 bead presented in Figure 14



(a) (b) (c) (d) (e) (f)

Figure 34a-f: Mel spectrograms of Invar 36 bead in Figure 14 - a) Threshold = 1, b) Threshold = 1.2, c) Threshold = 1.4, d) Threshold = 1.6, e) Threshold = 1.8, f) Threshold = 2

As shown, there are bright colored portions in the spectrogram that roughly correspond to the defective regions of the bead. Another observation is that these variations are best seen in the band between 3072-8192 Hz. This indicates that this particular frequency band needs to be isolated from the spectrogram and analyzed on its own. An added advantage of such an isolation would be the conversion of 2D data to 1D data, which can reduce computation time for a subsequent machine learning algorithm.

While spectrograms are very descriptive, it can be argued that their analysis through machine learning would be computationally intensive compared to one-dimensional or zero-dimensional data. This means it can be useful to examine such alternate parameters that could be extracted from audio.

A useful one-dimensional parameter for audio analysis that has been used in literature is kurtosis [33]. Kurtosis is a measure of "peakedness" of data which can be useful to identify anomalies in print audio signals [34]. Bhattacharya *et al.* [33] used audio kurtosis in combination with current and voltage data to predict weld deposition. The utility of such a parameter is the reduction in dimension (one-dimensional audio data is converted into a zero-dimensional point), which is useful to reduce computational cost. In Python, kurtosis can be calculated using the *scipy* module after denoising. A comparison of kurtosis and spectral variance for some beads of Inconel 718 can be seen in Figure 35.

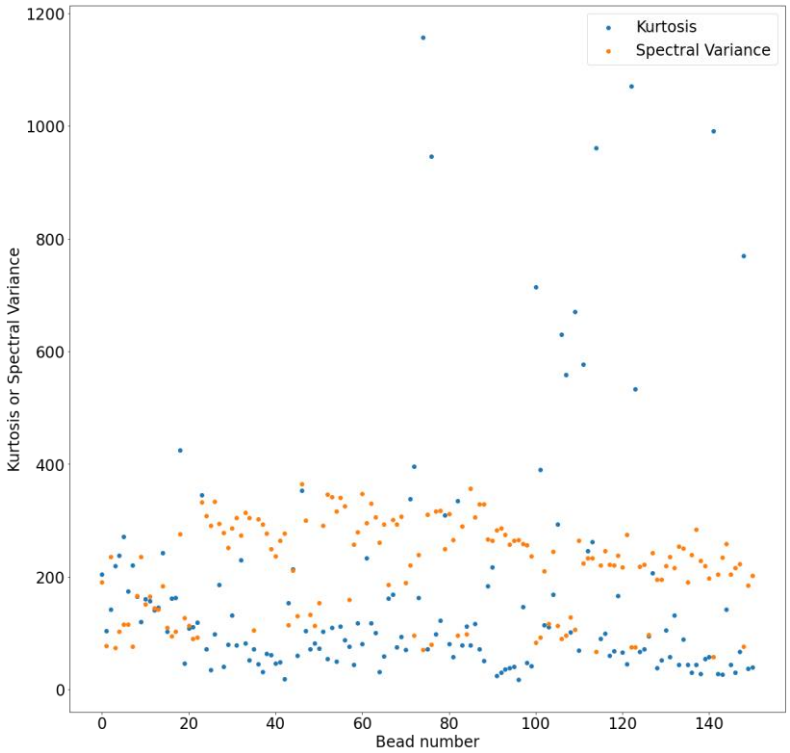


Figure 35: Kurtosis and spectral variance scatter plot of Inconel 718 beads

2.3.4 Parameter Summary

Since the calculation methods of all parameters have been described in Sections 2.3.1, 2.3.2 and 2.3.3, a table of input parameters can be compiled, as shown in Table 7. A vector comprising of these 6 parameters is used to represent each input datapoint after appropriate scaling in the subsequent ML approaches.

Table 7: Input parameters used for ML models

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 3072-8192 Hz frequency band of audio Mel spectrogram

2.4 Data Labeling Methodology

Printed bead datasets need to be labeled to have a ground truth for supervised machine learning approaches. In this context, labeling involves the identification of defects and classifying each bead as clean or defective based on the presence and severity of defects.

The assessment of defects in Inconel 718 and Inconel 625 was done using ASME standards [81]. All cracks and lack-of-fusion defects are considered unacceptable regardless of size, and since cracking was the primary defect observed in Inconel samples, this was used in order to label Inconel samples.

The case of Invar 36 is more complicated because of the fact that the defect being examined is not conventionally common. Internal defects such as porosity were observed on occasion and analyzed according to ASME standards [81], but the biggest anomaly observed was bead overflow (which could be classified under part dimension deviation). Looking at the relevant additive manufacturing defect formation standards from ASTM [82], part dimension deviations were indeed mentioned, but they were primarily concerned with stepping defects (*i.e.* stepped surfaces formed due to lack of proper slicing and path generation), but overflow defects were not discussed. Thus in such a scenario, a specific standard in terms of extent of deviation from the base geometry had to be defined in order to label deviations in the Invar 36 beads. For this thesis, an overflow defect was defined as follows – *An instance of bead overflow is considered to be unacceptable, if the normal distance of the outermost point of the overflow from the nearest point on the expected bead geometry is more than 40% of the bead width. Any instance of complete bead detachment and flow is also considered unacceptable.*

An example of this can be seen in the collection of Invar 36 beads shown in Figure 36. Once all beads were identified based on the above definitions, a binary labeling system was used where 0 corresponds to a clean bead, and 1 corresponds to a defective bead.

Figure 36: Unacceptable (3, 4) and acceptable (5) overflow defects seen in Invar 36 beads

3 Machine Learning Models and Hyperparameters

This chapter deals with the various ML models that are utilized during the course of this research work. Different ML approaches and the models appropriate for each approach are discussed, and the methods of optimizing hyperparameters for each kind of model to achieve optimum performance are explained.

3.1 Model Choice for Each Parameter

A useful avenue to consider as a starting point is unsupervised learning models. Unsupervised learning refers to the paradigm of training an ML model using unlabeled data [62]. In the context of this thesis, it means not informing the model whether the given information corresponds to a defective bead or clean one. The hypothesis is that this method would be advantageous, since the nature of defects that can arise in a bead may be unpredictable.

Multiple types of unsupervised approaches exist, one of which is clustering. As the name suggests, clustering involves separating the given training data into n clusters based on parameters that the model learns from the data. The number of clusters is defined by the user. Some examples of clustering approaches are indicated below.

1. K-Means Clustering
2. Mean-shift Clustering
3. DBSCAN

The idea is to use such approaches on the available bead data with the number of clusters being a rough approximation of the number of classifications/defects expected (+1 to include clean beads) and looking at the quality of the results.

Following an unsupervised analysis, supervised approaches can be used with bead labels based on micrograph observations. The micrographs can help decide the labels for the supervised

models, and the results of the unsupervised approach can be used to add additional details in the labeling if required.

Literature suggests the usage of neural networks and support vector machines for this kind of application. For this thesis, neural networks were chosen due to their high customizability, as mentioned in Section 1.3. Random forest models are also useful from the perspective of feature importance identification (understanding which features are most important for classifying beads), and thus will be analyzed too.

3.2 Hyperparameter Optimization

With regard to unsupervised clustering, one of the main hyperparameters is the number of clusters. This value decides the quality of segregation of data points, which is important for good results.

Literature suggests that the elbow method can be used to determine the optimal number of clusters for a K-Means model [63, 64, 65]. The method consists of fitting the K-Means model on the training data for a range of values of cluster numbers. For each cluster number, a distortion score (the sum of squares of distances of each point from the center of the cluster it is assigned to). The curve of distortion score versus cluster number is then analyzed, and the "elbow point" is identified. The idea is to find a certain threshold number of clusters beyond which no further useful information may be obtained. It is also possible for a situation where the curve is too smooth to find an elbow, as shown in Figure 37.

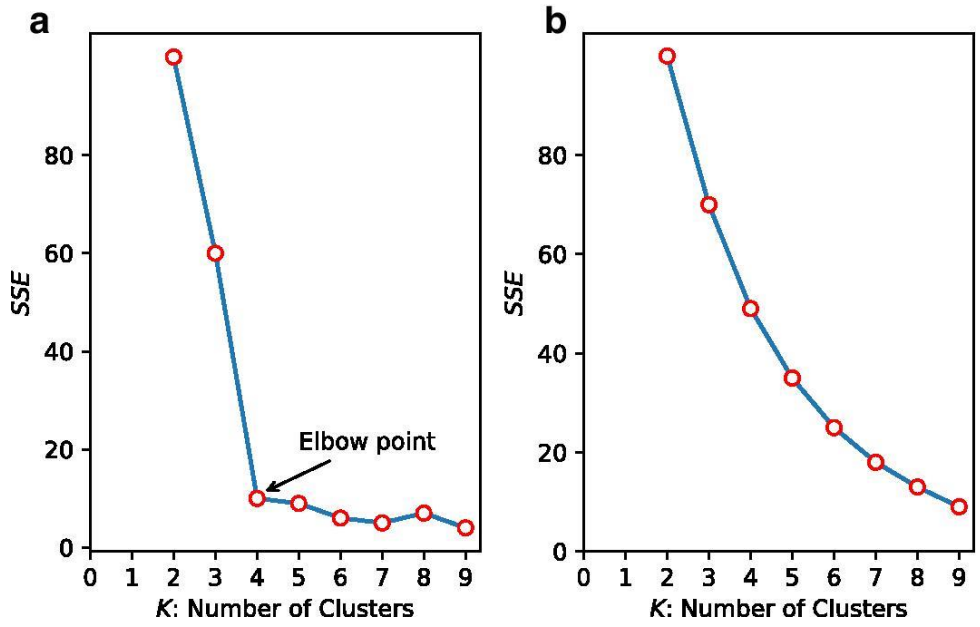


Figure 37a-b: a) Distortion plot with clearly identifiable elbow point b) Distortion plot without clearly identifiable elbow point. Reproduced from [66]

Another clustering hyperparameter is the number of initializations, *i.e.*, the number of times the model attempts to initialize centroids before finding the most converging value. A higher value essentially gives more attempts to find the most optimal solution, thereby giving a better quality of segregation.

For the supervised learning models, neural networks can be primarily examined. Neural networks tend to have a large number of hyperparameters since they are highly customizable. Some of these are described subsequently.

Two of the most obvious hyperparameters for any neural network are the number of hidden layers and the number of neurons in each layer. These two values directly influence the complexity of the model, and thus the classification accuracy. Care needs to be taken to avoid excessive values since that makes the model too complex, leading to possible overfitting.

Once the model structure has been decided based on the aforementioned hyperparameters, the next step is to determine the method and extent of optimization. The hyperparameters associated with this include the choice of optimizer, learning rate, number of epochs and batch size.

Multiple optimizers exist for neural networks, with one of the most commonly used ones being the Adam optimizer [67]. Others include SGD (stochastic gradient descent), Adagrad and RMSprop [68]. Optimizers generally work based on iterative gradient descent, which relies on finding the negative gradient of the function needed to be minimized (in this case, the loss function) in order to find the optimal solution [69]. On one hand, SGD and related optimizers (like SGD with momentum) work with a stochastic approach, which selects a random number of points from the dataset rather than the whole dataset itself, which reduces computational load [70]. The extent of change in each iteration is controlled by one of the hyperparameters, namely learning rate. On the other hand, Adam, Adagrad, RMSprop, etc. work based on adaptive learning, where the learning rate itself is modified individually for each parameter based on the magnitude of gradient of each parameter [71].

The number of epochs refers to the number of passes over the entire dataset done during the training process [72]. An insufficient number of iterations can lead to lack of sufficient learning, leading to bad predictions, whereas excessive number of iterations can lead to the model learning the noise within the data, leading to a problem known as *overfitting* [73]. Research done by Afaq *et al.* [73] suggested that the selection of number of epochs is a dataset-specific problem, and a general rule of thumb is difficult to decide. In general, a useful way of deciding the ideal number of epochs is to manually test different values and monitor the validation error performance. Unlike training error, validation error is calculated on data that the model has not seen, meaning that it can more accurately detect overfitting. An increase in validation error beyond a certain number of epochs will signal that the number of epochs is likely too high for the problem.

The final hyperparameter, *i.e.*, batch size is the number of samples that go through the neural network in each pass [74]. Having the batch size be the same as the total training dataset can give the most accurate picture in terms of gradient descent, but also takes more computational power (since more data is being processed). Thus models often use smaller batch sizes to attempt to improve efficiency without significant cost in terms of accuracy. Research done on CNNs suggests that larger batch sizes generally lead to higher accuracies [74]. Some authors have examined multilayer perceptrons and found that for a limited range of sizes, it is possible to use greedy algorithms to decide the best batch size [75]. Beyond this range, however, more intensive approaches like a grid search might be necessary. Smith *et al.* [76] examined Resnet-50 (a convolutional neural network) and found that increasing the batch size can improve accuracy (to a similar extent as compared to reducing the learning rate). For the purposes of this thesis, considering the comparatively smaller amount of data being analyzed (as compared to the aforementioned papers), the batch size can be set to the total input dataset size without significant computational cost.

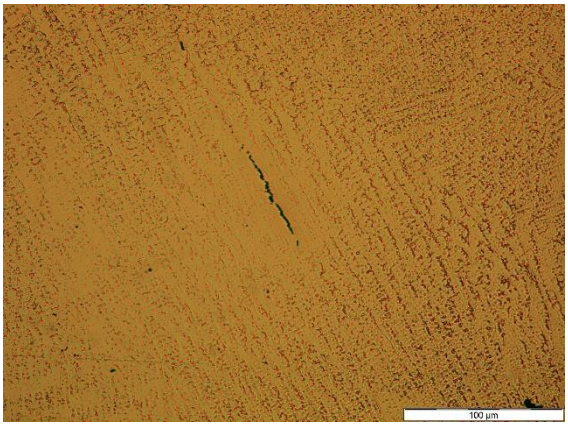
4 Results and Discussion

In this chapter, the results from the experiments are presented in detail. The physical defect observations are presented first in the form of micrographs, showing the most common defects seen in each material. This is followed by the data features observed for each material from which patterns are attempted to be deduced. The predictions made by the machine learning models detailed in Chapter 3 are then analysed using error metrics that are explained at the beginning of the section. A cross-material analysis using some of the ML models is also done. The reasons for different defect occurrences and the different signatures in the input data that correspond to these defects are discussed, which will help to further deduce whether analytical relationships can be made between different defects that occur in nickel-class alloys.

4.1 Defect Detections for each Material

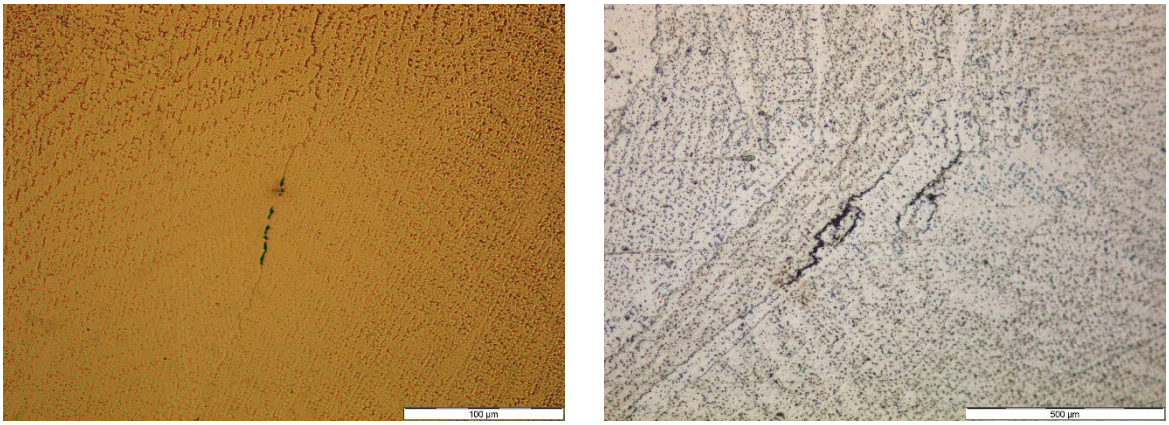
4.1.1 Defects in Inconel 718 Deposition

Among Inconel 718 samples printed at RAMLAB, the most common defects observed were solidification cracking and liquation cracking. This holds in line with the behavior of welded/printed Inconel 718 in literature [28, 77, 78]. Some examples of such hot cracking can be seen in the micrographs shown in Figure 38.



(a)

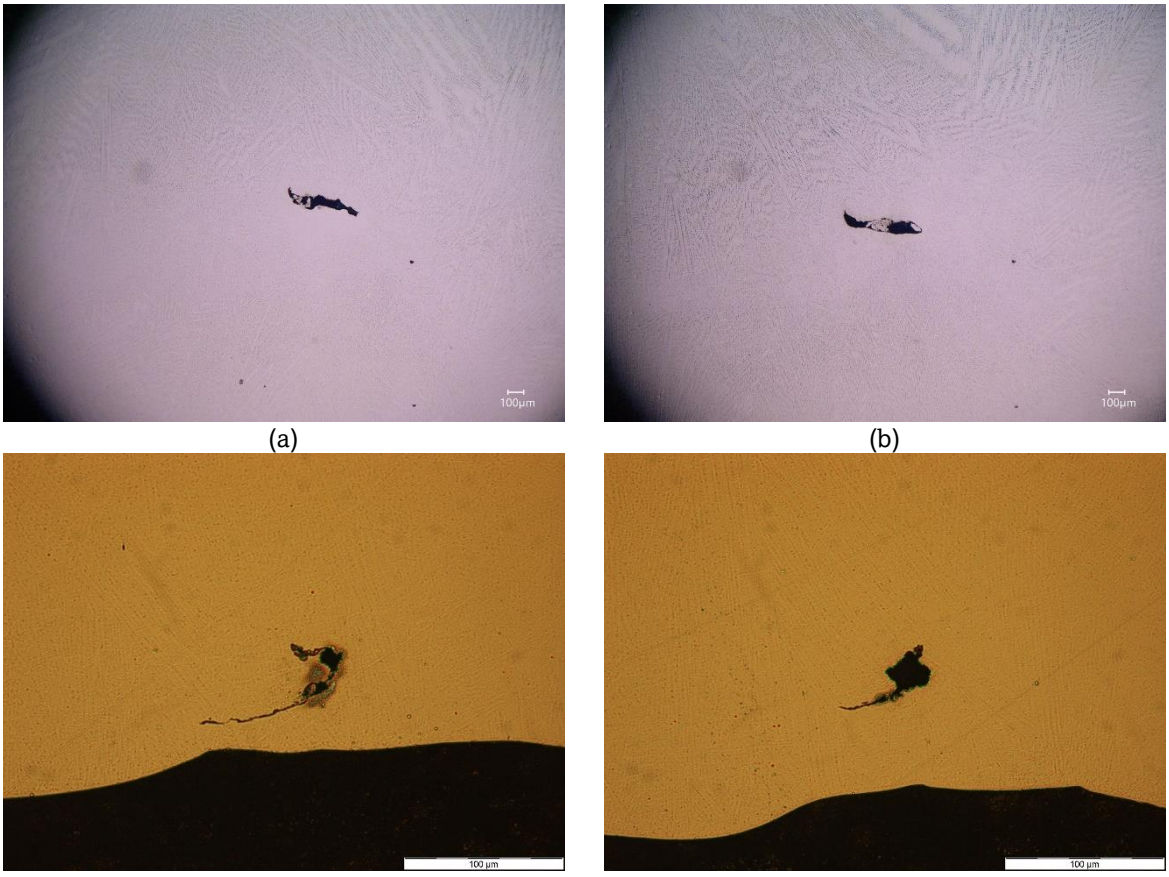
(b)



(c) (d)

Figure 38: Micrographs of various hot cracks observed in printed Inconel 718 samples – (a) solidification crack, (b)-(d) liquation crack

The defect locations appeared commonly at the interlayer positions (*i.e.*, between successive printed layers) are more susceptible to cracking. In addition to this, lack of fusion defects were also observed between the beads and commonly appeared around the weld toe where remelting can be difficult sometimes due to insufficient arc energy when the process is disturbed, as evidenced by the micrographs shown in Figure 39.



(a) (b)

(c) (d)

Figure 39: Micrographs of various lack-of-fusion defects observed in printed Inconel 718 samples – (a)-(b) lack-of-fusion in bulk, (c)-(d) lack-of-fusion in weld toe

4.1.2 Defects in Invar 36 Deposition

The most common defect seen in Invar 36 samples printed at RAMLAB was bead overflow. This is defined as a projection of the bead volume outside the expected print shape and volume due to insufficient solidification rate and associated liquid overflow. This is thought to happen due to the very low thermal conductivity of Invar 36 (10.49 W/mK [57] as compared to 54 W/mK for AISI 1024 steel [110], for instance). Supporting this hypothesis is the fact that the first couple of beads deposited on a conventional steel base plate do not show such overflow (which would be explained by the high conductivity of steel enabling fast diffusion of heat from the melt pool), whereas subsequent beads gradually begin to show this defect (which is explained by low conductivity of the Invar beads below which slows down heat conduction away from melt pool). One good example of this overflow is seen in Figure 14. Another example of this can be seen in Figure 36. Such overflow did not show significant internal defect indications in the micrographs of most Invar 36 beads. In some cases, porosity could be seen, as shown in Figure 40. While the possibility of cracks due to high heat retention was considered, no such cracking was observed in any of the micrographs.

Figure 40: Significant porosity seen in a single-bead-thickness wall of Invar 36

Measurement of the tensile properties of the final large Invar 36 wall deposition (shown in Appendix D) indicated improvements over the typical mechanical properties of the material. The results for longitudinal and transverse tensile specimens are summarized in Table 8 alongside the properties of Invar 36 from the wire manufacturer datasheet. The tensile results for Inconel 718 and Inconel 625 are still under processing as of now.

Table 8: Mechanical properties of large Invar 36 wall deposited and datasheet from wire manufacturer

Mechanical Property	*Longitudinal Specimen*	*Transverse Specimen*	*Datasheet*
Yield Strength, 0.2% (MPa)	325	328	280
Tensile Strength (MPa)	485	486	350
Elongation (%)	34	33	25

4.1.3 Defects in Inconel 625 Deposition

As mentioned in Section 2.1.2.3, Inconel 625 is also susceptible to hot cracking similar to Inconel 718. However, the prints constructed during the course of this thesis did not show such cracks, indicating that the parameters used were satisfactory. Additionally, a large number of Inconel 625 beads using near-optimal conditions were constructed as part of another project, whose data was used for this thesis.

4.2 Data Features for each Material

Before moving into each material, it is useful to evaluate the class composition, *i.e.*, the number of samples which are clean versus defective. Class imbalance 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 was calculated for each material, defined as the ratio of number of samples in majority class (class with higher number of samples) to number of samples in minority class (class with lower number of samples). This ratio, along with the numbers of samples in each class for all the materials in this study has been presented in Table 9.

Table 9: Class composition for all materials (Inconel 718, Invar 36, and Inconel 625)

Material	*Inconel 718*	*Invar 36*	*Inconel 625*
Clean	30	145	658
Defective	418	34	0
Imbalance ratio	13.933	4.265	∞

It is immediately clear that 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 literature to identify some kind of limit of imbalance ratio beyond which prediction performance gets affected.

Buda [95] conducted a study on different kinds 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 around 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 done by Leevy *et al.* [96] suggested that high imbalance ratios generally lie between 100:1 and

10000:1, indicating that the ratios in Table 9 do not qualify to be extreme in nature (except Inconel 625). Johnson *et al.* [97] 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 (not just the imbalance ratio).

4.2.1 Inconel 718 Data Features

During the various experimental prints of Inconel 718 conducted at RAMLAB, solidification cracks were extensively observed throughout, as shown in Figure 39. In total, 448 individual beads of Inconel 718 were deposited. The first step is to calculate the total population variance of $n_{\Delta V}$ for each Inconel 718 bead and represent all these bead variances in a scatter plot. The plot is shown in Figure 41. It is quickly evident that while most of the values are low and between 0 and 200 (by visual estimation), some values are much higher. A careful examination reveals that the errant beads belong to pulsemix ramp tests and small plate bead wall tests. Plots of the batchwise variances of the $n_{\Delta V}$ of all these beads are done in one plot and shown in Figure 42. It is clearly visible that the variance values are much higher than those visible in Figure 41, for reference. Since it has been established that the high variance beads are primarily pulsemix welds, it is only logical to separate superactive and pulsemix beads and examine them separately.

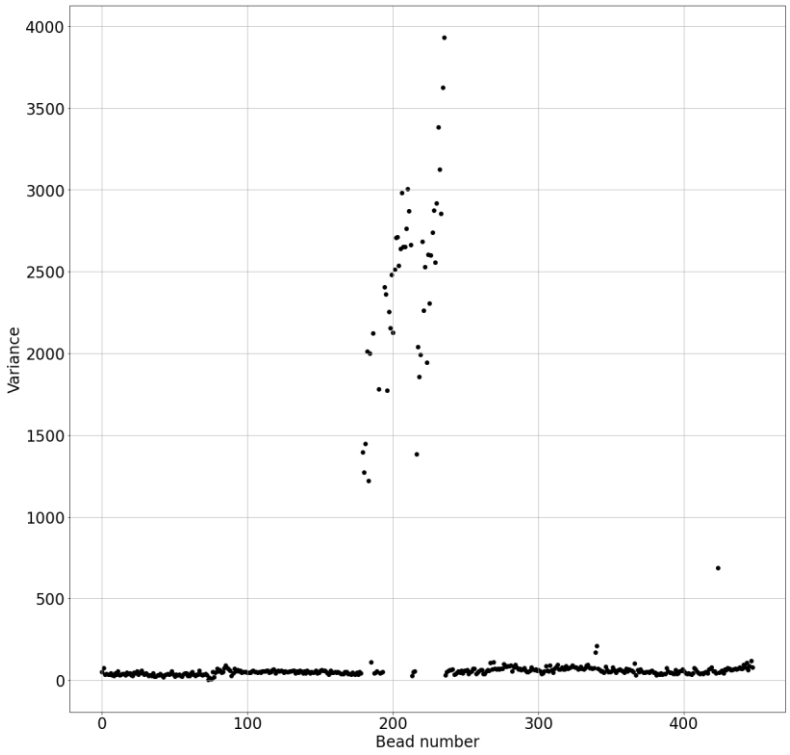


Figure 41: Variance of $n_{\Delta V}$ of all Inconel 718 beads

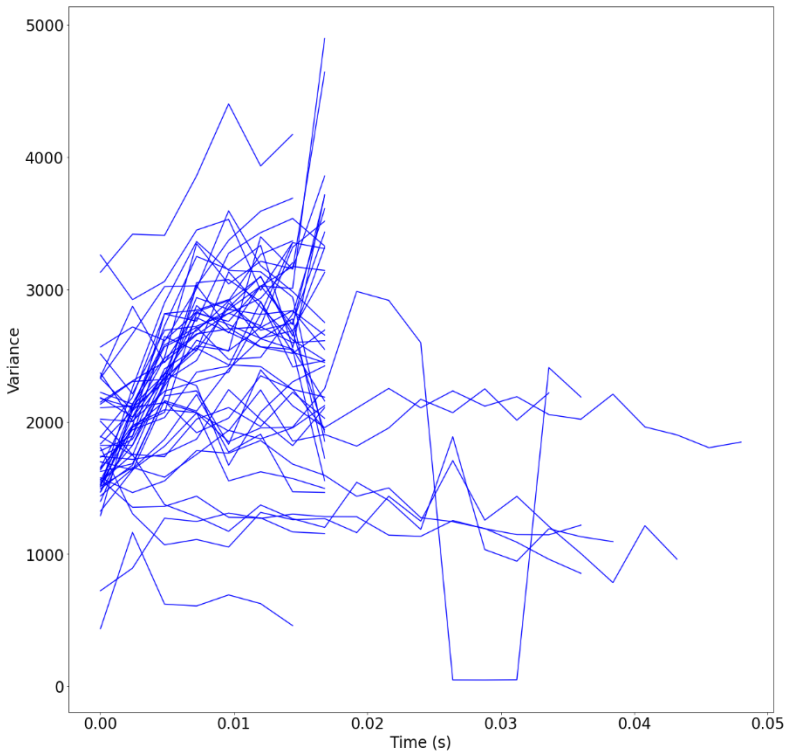
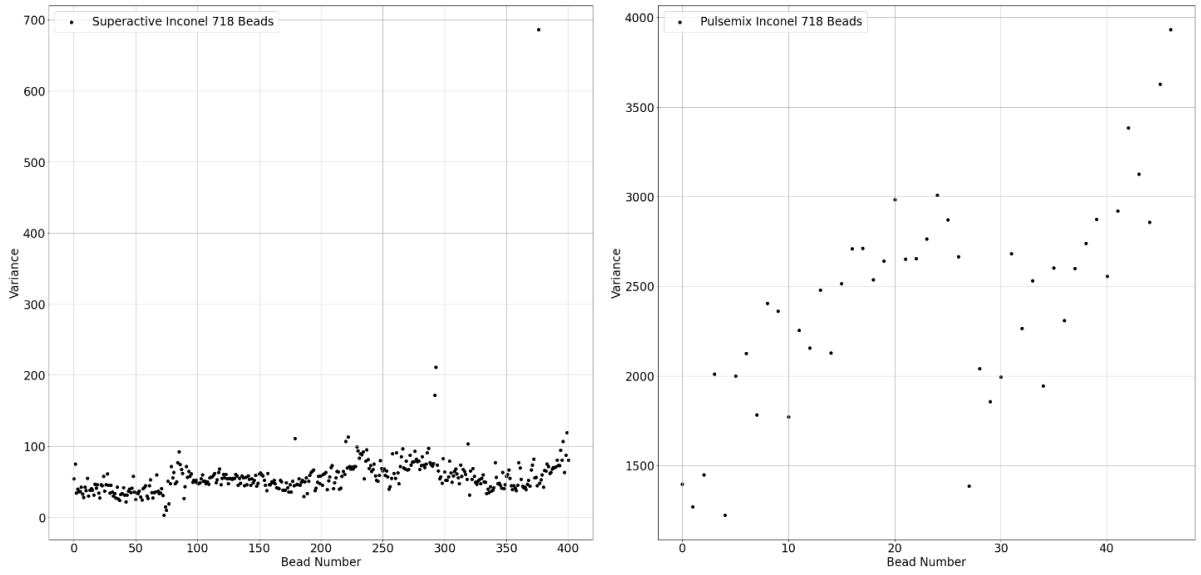


Figure 42: Chunk-wise $n_{\Delta V}$ variance of Inconel 718 pulsemix beads for n = 60

Considering superactive and pulsemix beads separately, constructing a variance plot (similar to Figure 41) gives Figure 43. The majority of values in Figure 43a are seen to lie between 0 and 100, with a few values going outside these bounds. In Figure 43b, the range of values is much bigger, going between 1000 and 4000. Using a similar approach for current, the population variance plot of $n_{\Delta I}$ obtained for all beads without separation is shown in Figure 44. It is interesting to note that contrary to voltage, the difference between superactive and pulsemix modes is not so significant when looking at current. It is clearly seen that when examining the corresponding bead numbers, the difference between superactive and pulsemix beads is smaller than in voltage. The superactive and pulsemix beads have been plotted separately in Figure 45.



(a) (b)

Figure 43a-b: Variance of $n_{\Delta V}$ of all (a) superactive Inconel 718 beads, (b) pulsemix Inconel 718 beads

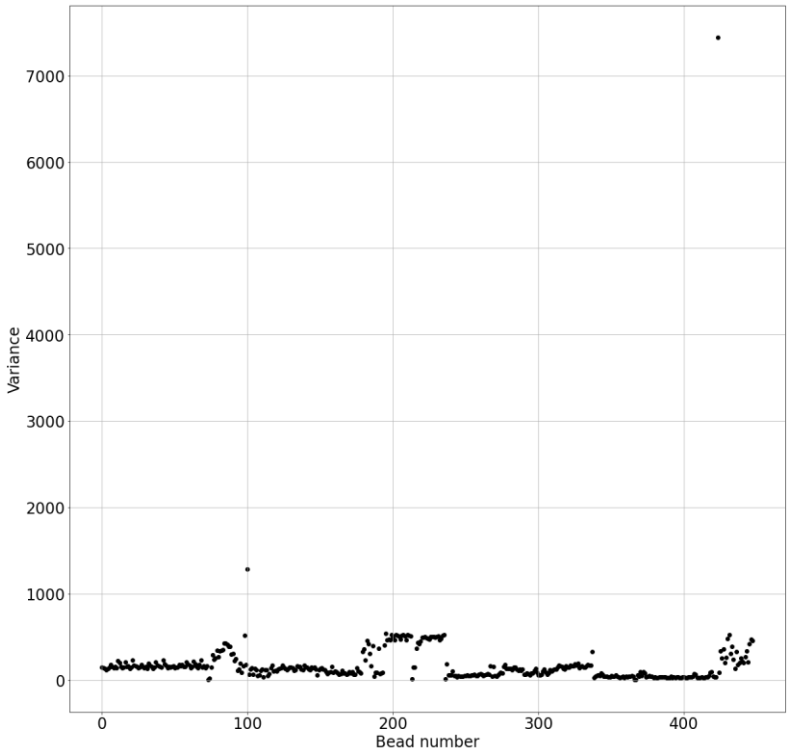
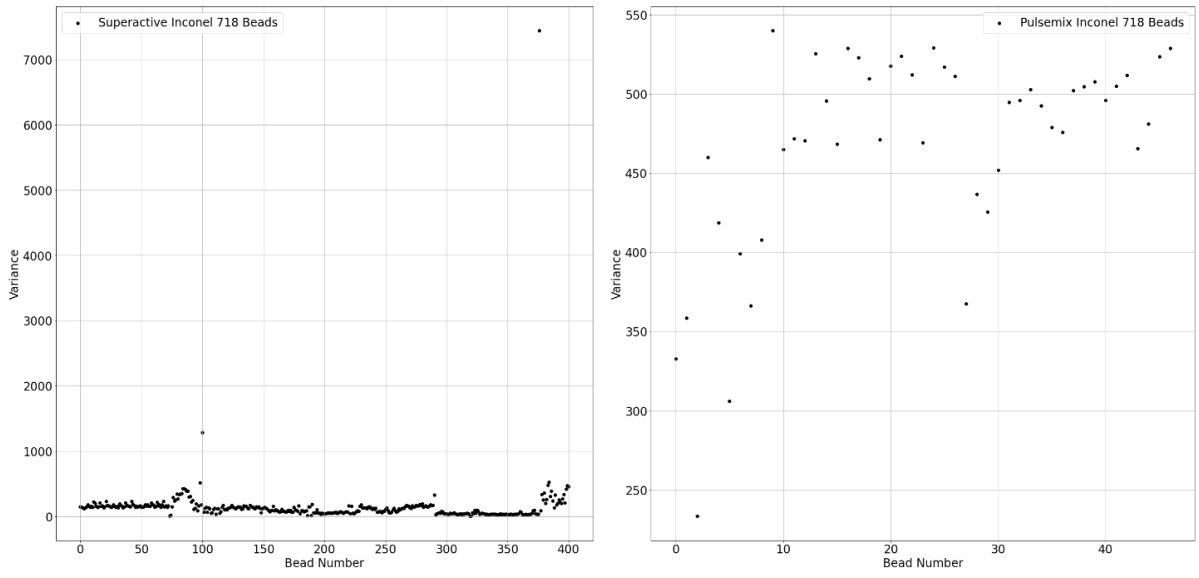


Figure 44: Variance of $n_{\Delta I}$ of all Inconel 718 beads

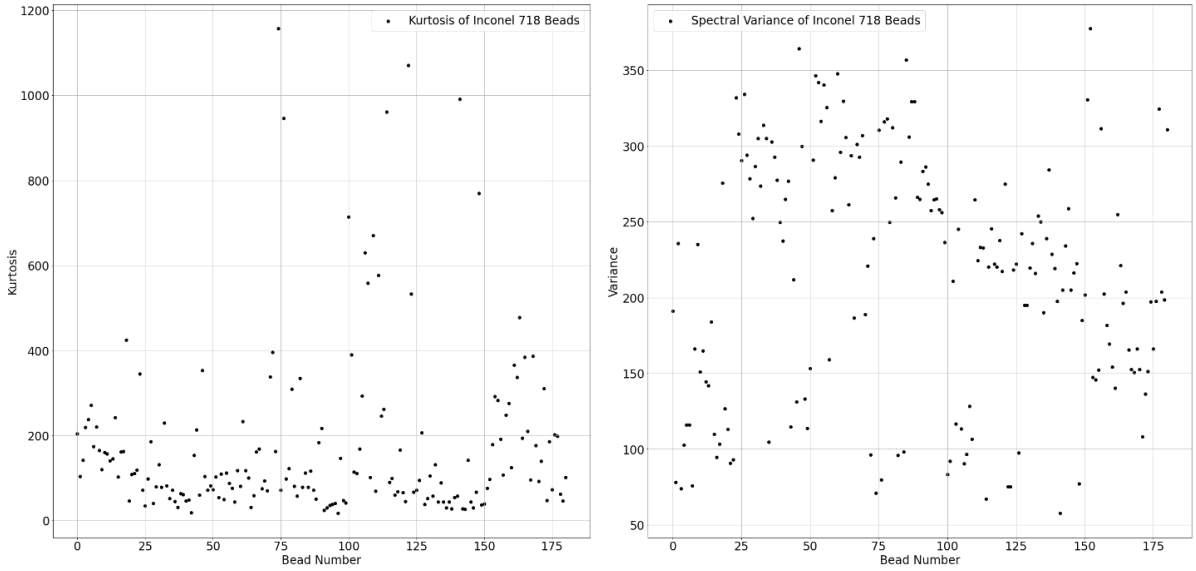


(a) (b)

Figure 45a-b: Variance of $n_{\Delta I}$ of all (a) superactive Inconel 718 beads, (b) pulsemix Inconel 718 beads

The last parameter to examine is audio. In this case, the raw audio is denoised as indicated in Section 2.3.3 and the kurtosis and spectral variance are taken for each bead. As mentioned earlier, audio data was not available for the entirety of beads printed, and thus the available data is

presented in Figure 46. The spectral variance does not show a clear trend, whereas the audio kurtosis has most values lying between 0 and 400, with some outliers. All the beads for which audio data was available were deposited in superactive mode as it was the mode of interest for an industrial partner of RAMLAB, so a mode-based separation is not done, which can be included in future research.



(a) (b)

Figure 46a-b: (a) Kurtosis, (b) Spectral variance of all Inconel 718 audio signals

4.2.2 Invar 36 Data Features

Examination of different beads showed bead overflow to be the most likely defect in Invar 36. Cracking was not observed and even the presence of porosity was only in the case of severe overflow. Since arc disturbances *need not happen* for excessive energy input, it is expected to be a little more difficult to accurately detect such defects. One example of such an overflow can be seen in Figure 47.

Figure 47: Invar 36 bead showing overflow

Throughout the course of experiments, a total of 179 individual beads of Invar 36 were deposited. Similar to Inconel 718, a total variance plot of $n_{\Delta V}$ for each bead can be calculated and the resulting scatter plot is shown in Figure 48. It is visible that most of the variances are visually approximately between 0 and 200, with some values being significantly higher. It is also observed (via experimental logs) that the majority of high-value points belong to beads generated by pulsemix welding.

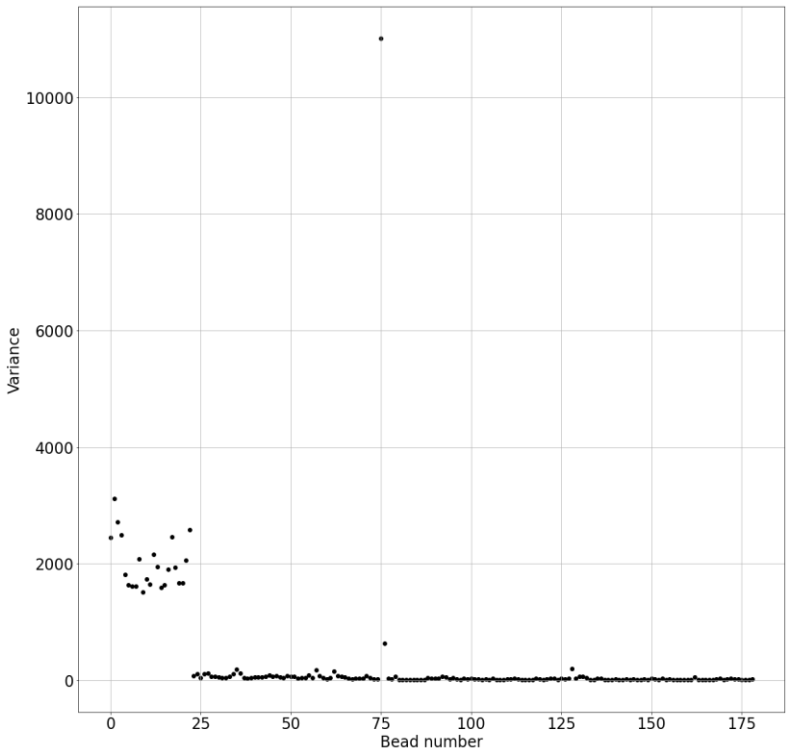
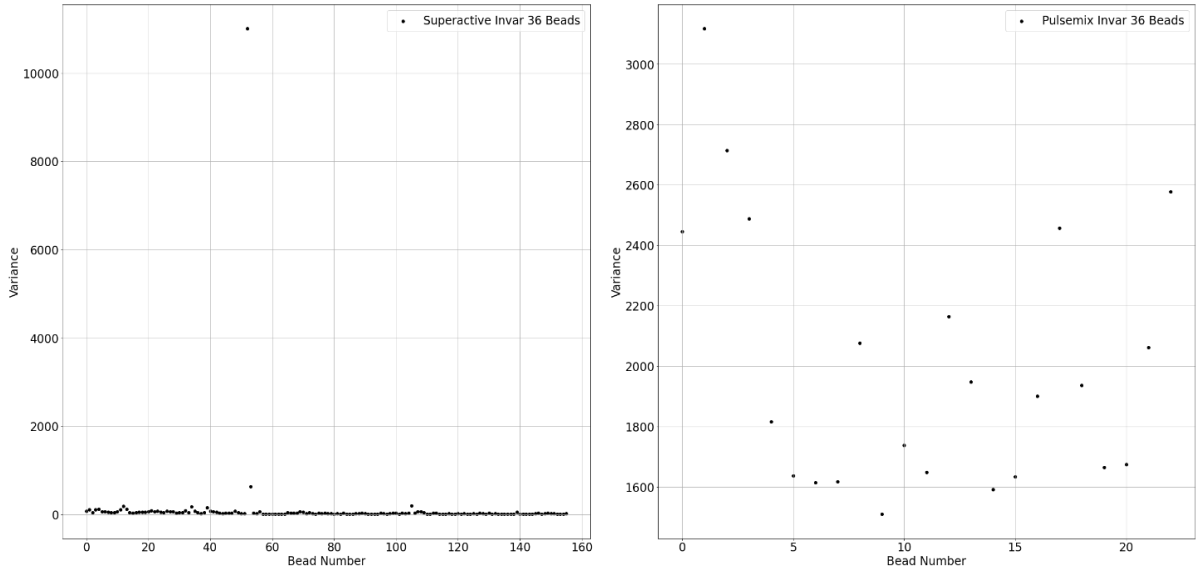


Figure 48: Variance of $n_{\Delta V}$ of all Invar 36 beads

Analogous to Inconel 718, it is useful to separately plot the $n_{\Delta V}$ of all the superactive and pulsemix beads for further examination. This results in Figure 49. Similar to the case of Inconel 718, it is visible that the majority of beads have a variance below 100, with a few going above this value. Shifting focus to current, the variance plot of $n_{\Delta I}$ obtained for all beads without separation is shown in Figure 50. Similar to Inconel 718, the difference between superactive and pulsemix modes is smaller when looking at current. It is, however, still clear that the values of pulsemix welds seem to be higher overall than superactive mode.



(a) (b)

Figure 49a-b: Variance of $n_{\Delta V}$ of all (a) superactive Invar 36 beads, (b) pulsemix Invar 36 beads

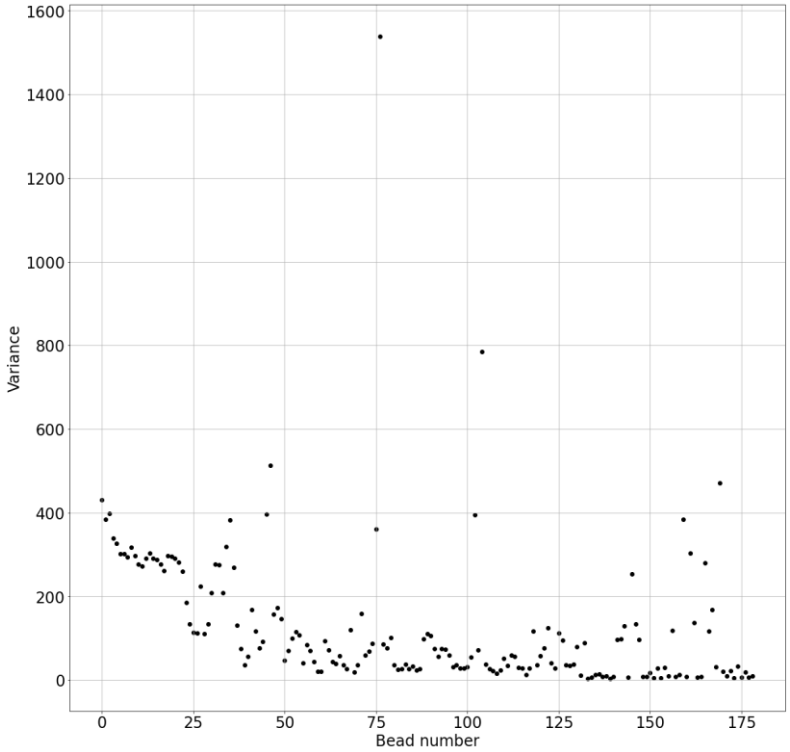
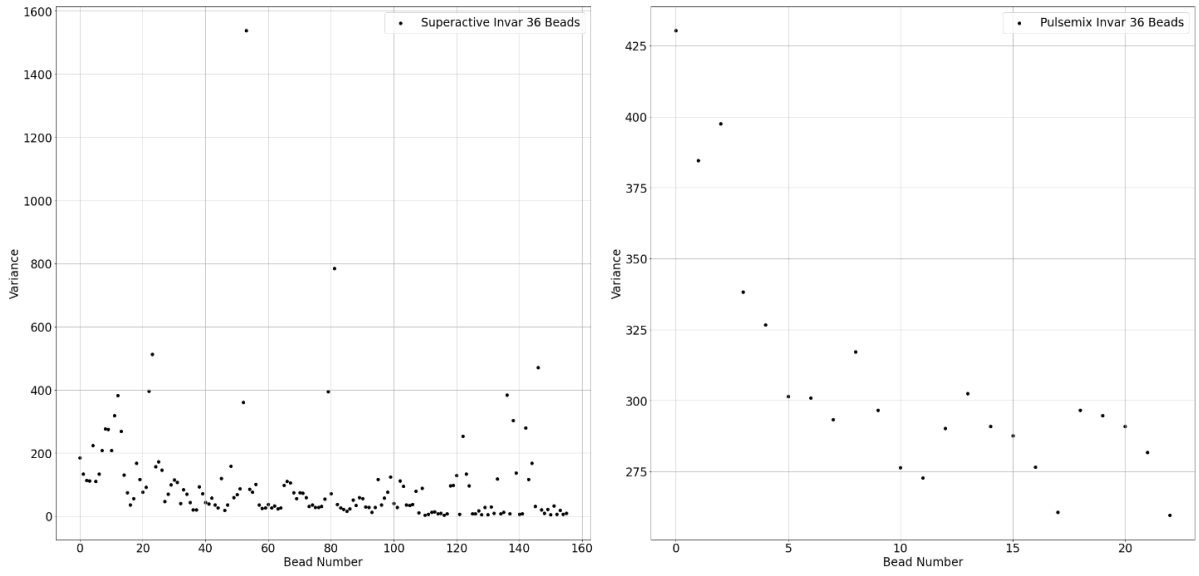


Figure 50: Variance of $n_{\Delta I}$ of all Invar 36 beads

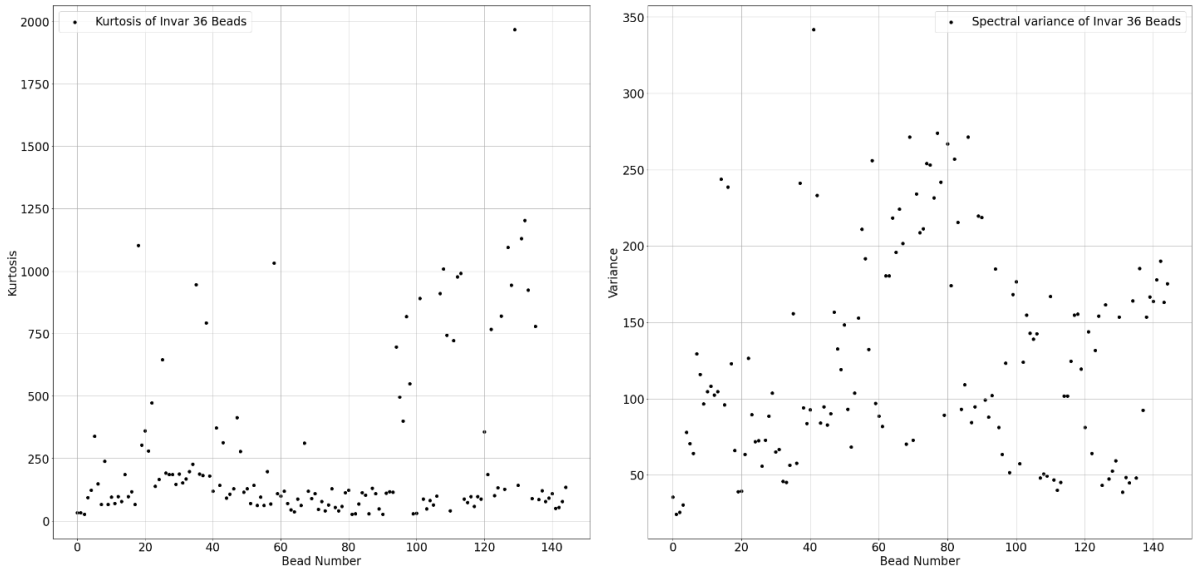
The $n_{\Delta I}$ of superactive and pulsemix welds are plotted separately and shown in Figure 51. The majority of superactive values are seen to lie between 0 and 200, with a few values exceeding these limits.



(a) (b)

Figure 51a-b: Variances of $n_{\Delta I}$ of all (a) superactive Invar 36 beads, (b) pulsemix Invar 36 beads

The audio data including kurtosis and spectral variance of all the beads is presented in Figure 52. The kurtosis plot shows most of the values lying between 0 and 250, with some exceptions. The spectral variance does not show a clear visible trend.



(a) (b)

Figure 52a-b: (a) Kurtosis, (b) Spectral variance of all Invar 36 audio signals

4.2.3 Inconel 625 Data Features

The tests conducted on Inconel 625 comprised of a few single beads, small walls and a couple of large walls. A total of 658 beads of Inconel 625 were deposited. Micrographs indicated a normal extent of substrate dilution (5-10%), but no major defects were observed. In general, literature indicates that Inconel 625 may also show the possibility of weld cracking similar to Inconel 718 [61].

The variance of $n_{\Delta V}$ of all the beads is presented in Figure 53. Most of the points are visibly below 400, with a few that are above, but not much higher (less than 50% higher at most). This makes sense, considering that all the beads were printed using superactive mode. Interestingly enough, it is visible that the variances of the beads printed on small plates while making small builds (the rightmost 22 points) is much lower than the variances of the beads printed as part of the big walls.

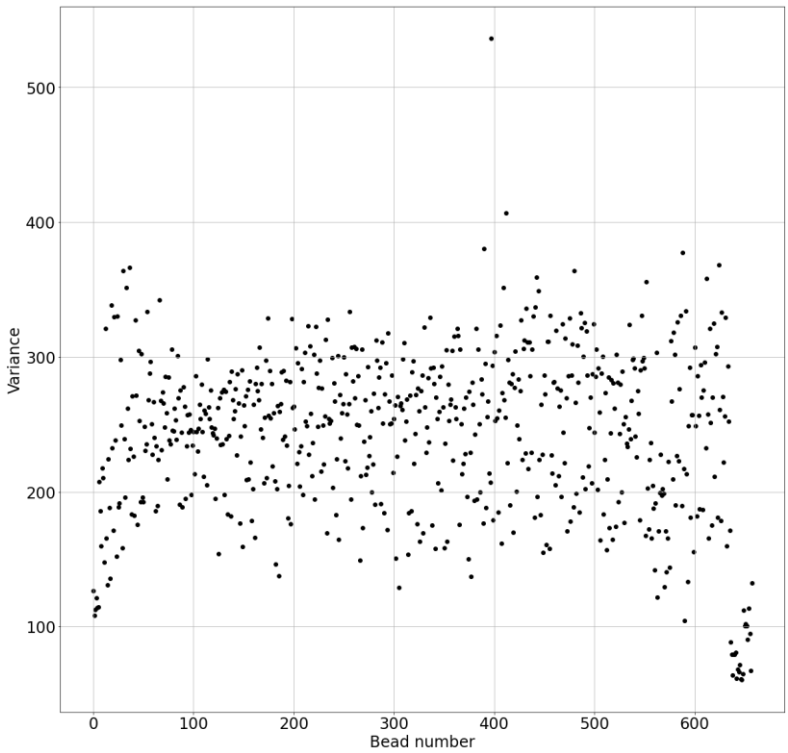


Figure 53: Variance of $n_{\Delta V}$ of all Inconel 625 beads

The variance of $n_{\Delta I}$ of all the beads is shown in Figure 54. Most of the values are below 350, but in a few beads, the variance of the current exceeds this value. This is also attributable to the fact that all the bead weld modes were superactive in nature.

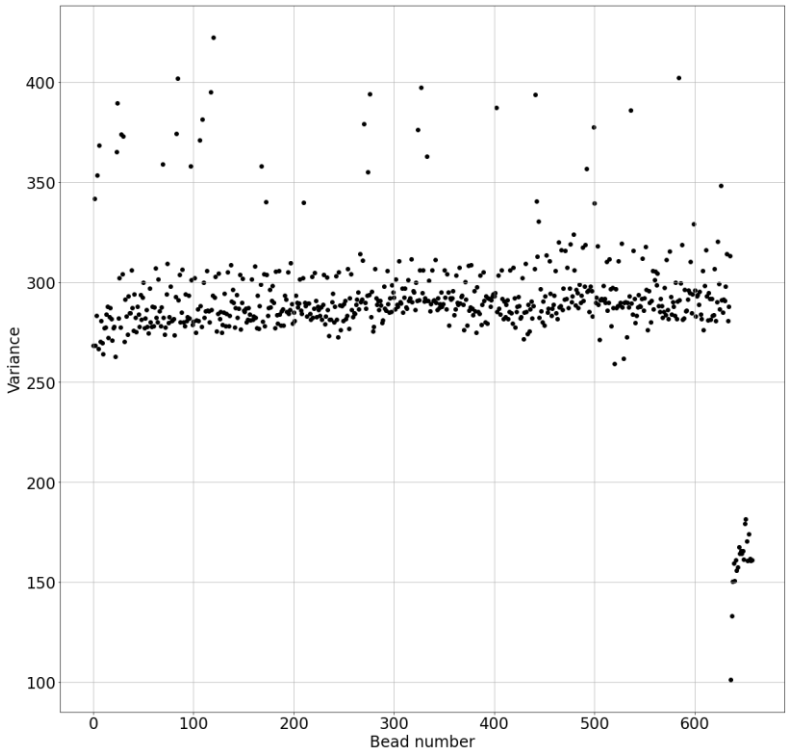
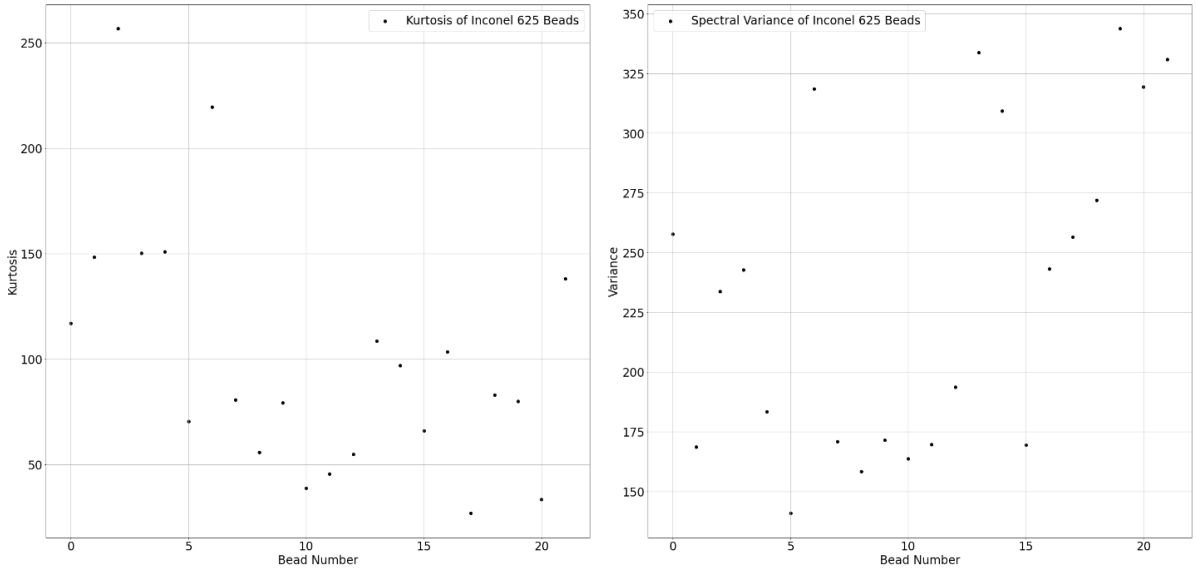


Figure 54: Variance of $n_{\Delta I}$ of all Inconel 625 beads

The kurtosis and spectral variance of all the beads (for which audio data was available) are shown in Figure 55. A clear pattern in the variance in the audio signals is not visible for this material either. It is important to note that audio data was not available on the beads printed for a large wall (amounting to 636 beads in total), so they have not been depicted in Figure 55.



(a) (b)

Figure 55a-b: (a) Kurtosis, (b) Spectral variance of all Inconel 625 audio signals

4.3 Machine Learning Results

As mentioned in Section 4.2, the class imbalance in the available data needs to be considered in order to accurately evaluate the performance of the models used. For this purpose, 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.

In addition, scaling was done on all input parameter vectors to ensure equal weightage of every parameter in the model. Scaling was done using the StandardScaler function from scikit-learn (in Python), 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 9, 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. This way, the ability of the model to correctly ascertain the relation between input parameters and class labels for the minority class could be assessed.

4.3.1 Single Material Evaluation

4.3.1.1 Evaluation Metrics

Machine learning (ML) model evaluation is a wide research topic in itself. A vast amount of research is being conducted and published on the different evaluation methodologies of evaluation with their advantages and disadvantages [83, 84, 85, 86]. The "black box" approach of most ML models can make it difficult to accurately identify if the model predictions are applicable in practical cases.

Elmrabit *et al.* [83] analyzed multiple ML classification models to evaluate various instances of network behavior in order to identify cyberattacks. In this work, the authors used seven metrics to analyze their 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 done into time series anomaly detection by Aminikhanghahi *et al.* [87] details the aforementioned parameters and mentions that they can be useful in classification-related problems. Chicco *et al.* [88] interestingly notes that another metric called the Matthew's Correlation Coefficient (MCC) can be better used than F1 score and accuracy in order to evaluate binary classification problems. The authors conducted tests on highly class-imbalanced datasets and found that the MCC score could better assess the poor quality of prediction as compared to accuracy and F1 score (both of which were erroneously high in some cases). Considering that the current thesis 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 for this thesis. A summary of the evaluation metrics used for classification in this research and their formulae has been shown in Table 10.

Table 10: Classification evaluation metrics summary

Metric	*Formula*	*Meaning*	*Range*
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Degree of closeness of predicted value to ground truth	{0, 1} Close to 1 is good
Precision	$\frac{TP}{TP + FP}$	Quality of positive (defect) prediction	{0, 1} Close to 1 is good
Recall	$\frac{TP}{TP + FN}$	Ability to predict positives (defects) correctly	{0, 1} Close to 1 is good
False Positive Rate (FPR)	$\frac{FP}{FP + TN}$	Extent of false prediction of positives (defects)	{0, 1} Close to 0 is good

F1 Score	$\frac{TP}{TP + 0.5 * (FP + FN)}$	Harmonic mean of precision and recall	{0, 1} Close to 1 is good
Matthew's Correlation Coefficient (MCC)	$\frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$	Summarizes the contributions of the entire error matrix	{-1, 1} Close to 1/-1 is strong positive/negative correlation

In Table 10, TP and TN represent true positives and true negatives, which are the predicted positives (defects) and negatives (clean) that match with the ground truth. FP and FN represent false positives and false negatives, which are the predicted positives and negatives that do not match with the ground truth.

As for regression, the metrics used are generally mean squared error (MSE), root mean squared error (RMSE) or R^2 score. Since this is non-linear regression, R^2 score is invalid and an extremely poor indicator of model quality due to high susceptibility to number of parameters [91]. Thus, the error metrics being used will be mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). Mean absolute percentage error (MAPE) is not considered since there are ground truth class labels with value 0, meaning a division by 0 would occur.

While some reviews state that these measure have flaws too, it is also agreed upon that an "ideal" metric does not exist [100]. Thus the metrics that are selected will be analyzed while being mindful of the factors that might cause bias towards excellent or poor results. A summary of the evaluation metrics used for regression in this research along with their definitions and formulae has been shown in Table 11. Here, $\hat{y}_i$ refers to the i^{th} ground truth value, whereas y_i refers to the i^{th} predicted value.

Table 11: Regression evaluation metrics summary

Metric	*Formula*	*Meaning*	*Range*
Mean Squared Error (MSE)	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Average squared difference between predicted and ground truth values	{0, ∞} Close to 0 is good
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	Root of average squared difference between predicted and ground truth values	{0, ∞} Close to 0 is good
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	Average absolute difference between predicted and ground truth values	{0, ∞} Close to 0 is good

4.3.1.2 Random Forest Evaluation

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 [98]. Each tree runs its own analysis, and the collection of trees eventually lands at a decision.

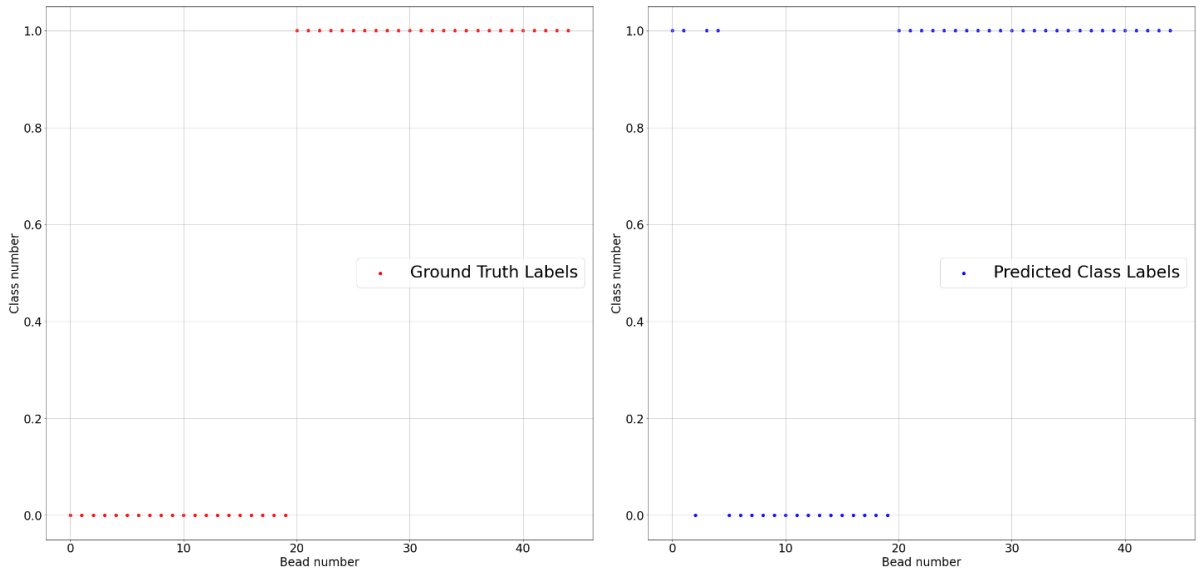
Random forests can be extremely useful to identify the relative importance of features, since the decisions made at each node in each tree depends on the input features. Python offers the functionality of calculating the relative importance of each feature based on the mean decrease in impurity (or Gini 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 good extent of division of classes (*i.e.*, the parameter that determines the question asked by the node is of high importance).

The hyperparameters used for the random forest analysis have been shown in Table 12.

Table 12: Hyperparameters of Random Forests model used

Hyperparameter	*Value*
Number of trees (n_estimators)	1000
Maximum tree depth (max_depth)	8
Criterion (criterion)	Entropy

The available criteria are entropy and Gini, both of which are effectively the same for most applications [99]. Running the experiments on Inconel 718 led to the predictions shown in Figure 56.



(a) (b)

Figure 56a-b: (a) Ground truth, (b) Predicted class labels for Inconel 718 beads

Visual inspection of Figure 56 indicates the results are fairly accurate. To view this in a numerical perspective, the error metrics of this model on the testing set are shown in Table 13.

Table 13: Evaluation metrics for Random Forest classification of Inconel 718 beads

Evaluation Metric	*Score*
Accuracy	0.911 (91.1%)

Precision	0.862 (86.2%)
Recall	1.000 (100.0%)
False Positive Rate (FPR)	0.200 (20.0%)
F1	0.926 (92.6%)
MCC	0.830 (83.0%)

The results are great, with a high MCC indicating very strong positive correlation between the prediction and ground truth. The accuracy, precision, recall and F1 are also very high. The false positive rate is 20%, which might be too high for an industrial application. Ideally, an FPR of less than 10% should be targeted at the very least (the decision of an appropriate FPR is typically done on a case-to-case basis). 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 such datapoints for the most part. Using the inbuilt feature_importances_ attribute (in Python) on each tree in the random forest, a bar plot of feature importances can be constructed as shown in Figure 57.

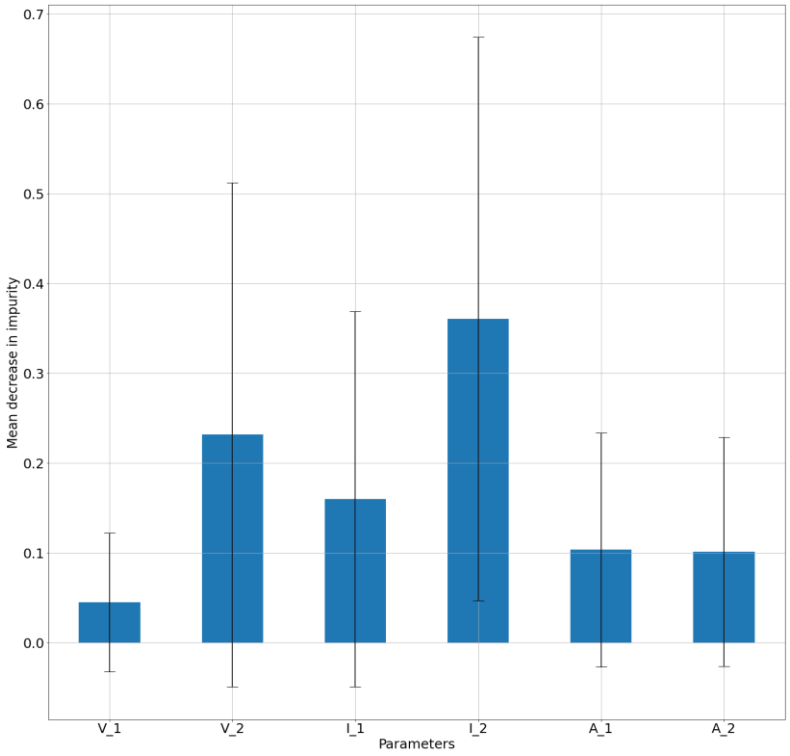
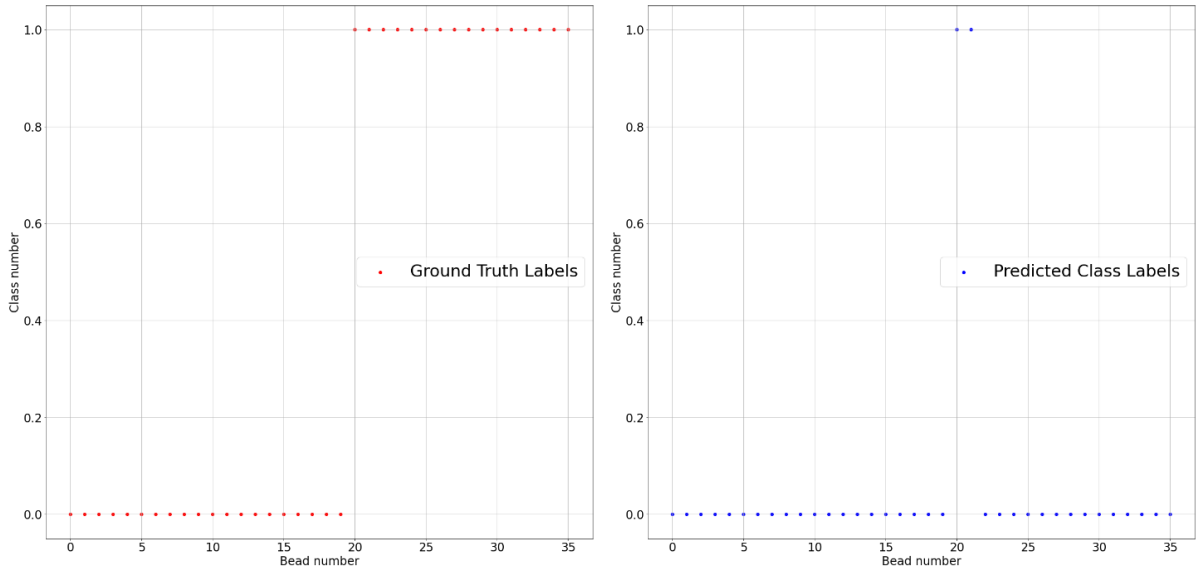


Figure 57: Bar plot of feature importances of all input parameters for Inconel 718

It is evident that the pulse width variance parameters of welding voltage and welding current (V_2 and I_2) are the most important features for classification. Interestingly, it is also seen that the audio features are of comparatively low importance in this classification. The large error bars are important to note, but discussion in the subsequent sections will use the mean values depicted in the bar plot to discuss relative feature importance.

For Invar 36, the minority class is the opposite (*i.e.*, defective beads). Using a similar training-testing split of the minority class as Inconel 718, running the same random forest model gave the predictions in Figure 58.



(a) (b)

Figure 58a-b: (a) Ground truth, (b) Predicted class labels for Invar 36 beads

Figure 58 clearly shows that the predictions on the defect class (label = 1) are poor. Only 2 beads in the defective data class have been identified correctly. The numerical representation of this result is shown in Table 14.

Table 14: Evaluation metrics for Random Forest classification of Invar 36 beads

Evaluation Metric	*Score*
Accuracy	0.611 (61.1%)
Precision	1.000 (100.0%)
Recall	0.125 (12.5%)
False Positive Rate (FPR)	0.000 (0.0%)
F1	0.222 (22.2%)
MCC	0.271 (27.1%)

The results 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. It is also interesting to note how the accuracy, precision and false positive rates are fairly good, which shows the danger of using an insufficient number of metrics to evaluate a model.

The calculation of feature importances leads to the bar plot shown in Figure 59. Compared to Figure 57, the discrepancy between importances seems smaller. The voltage count variance (V_1) is seen to be the most important parameter, followed by both current parameters (I_1 and I_2). The audio parameters are also seen to have increased in importance in comparison to Figure 57.

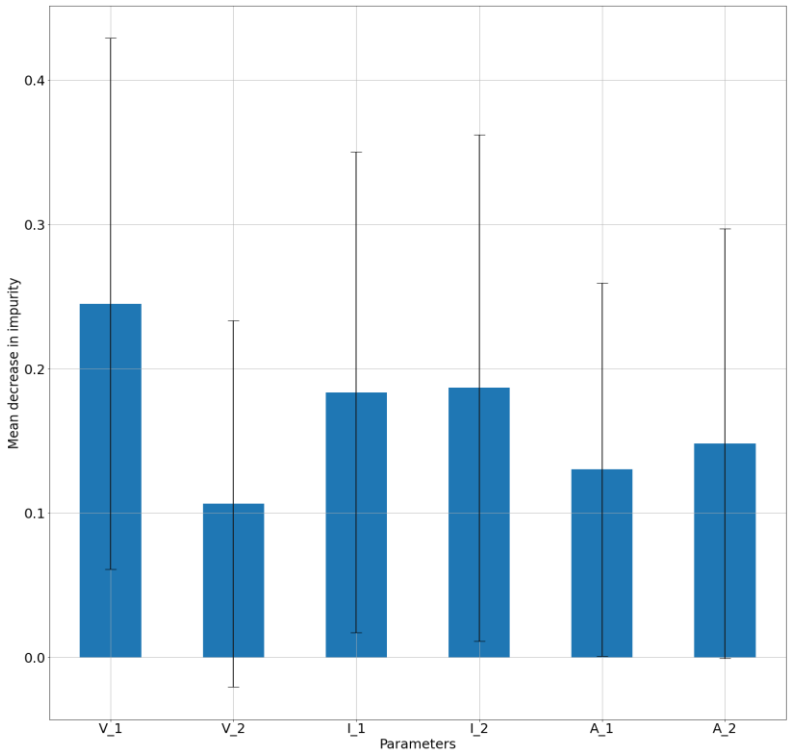


Figure 59: Bar plot of feature importances of all input parameters for Invar 36

4.3.1.3 K-Means Clustering Evaluation

The next approach was unsupervised K-Means Clustering, which is one of the most well-known unsupervised clustering methods as mentioned in Section 3.1. The objective was to let the model decide boundaries for classification of clean and defective beads, and then compare the outcome to the known labels. Due to the lack of requirement of training, there was no splitting of data done here to delineate training and testing sets.

As detailed in Section 3.2, an elbow point should be identified first. Python offers this functionality using Yellowbrick, and this was used to analyze the dataset of Inconel 718 beads [94]. The elbow point plot obtained has been shown in Figure 60.

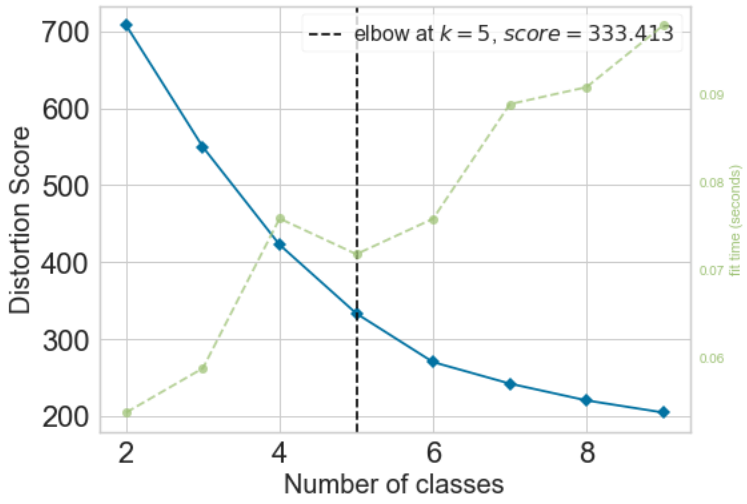
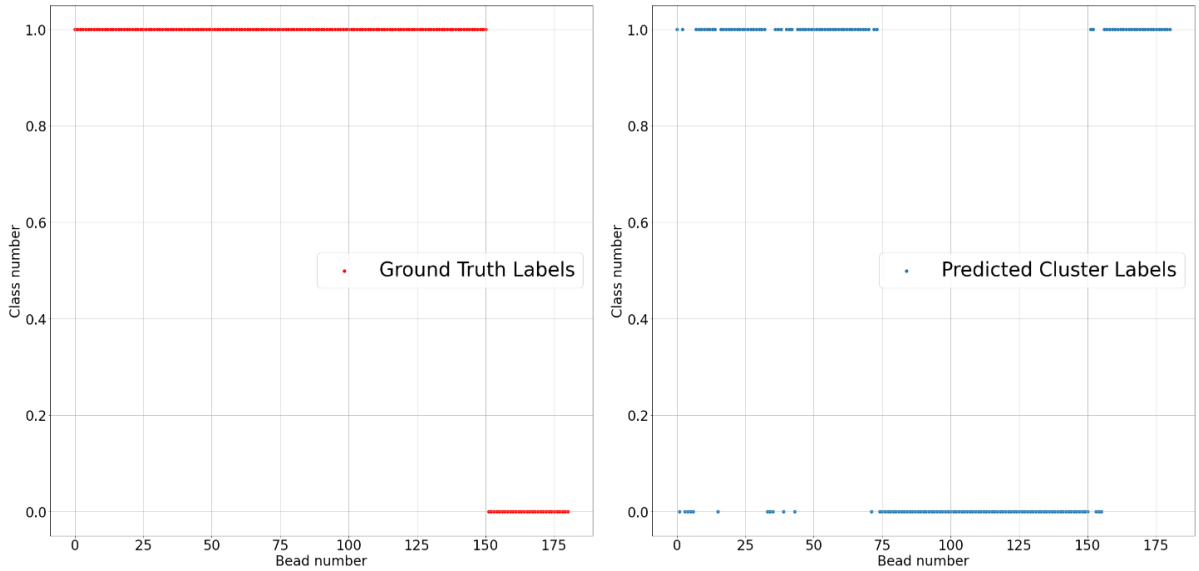


Figure 60: Elbow point plot for K Means Clustering for Inconel 718

The plot indicates that the threshold number of clusters for the data is 5, which seems unphysical since under any paradigm, it is not expected to obtain five different classes. It is important to note, also, that in this case the elbow is not sharp, indicating that the model has problems differentiating between clusters. For the analysis of this dataset, the hyperparameters of the model used are shown in Table 15. Conducting a test with 2 clusters, a prediction is obtained, which is shown side-by-side with the actual labels for 181 beads of Inconel 718 in Figure 61.

Table 15: Hyperparameters of K-Means clustering model used

Hyperparameter	*Value*
Number of clusters (n_cluster)	2
Number of cluster centroid initializations (n_init)	25
Tolerance for convergence (tol)	0.0001
Number of iterations (max_iter)	300



(a) (b)

Figure 61a-b: (a) Ground truth, (b) Predicted cluster labels for Inconel 718 beads

The predicted cluster labels can be flipped, since the numbering that the model assigns to clusters is simply a semantic issue. Thus in order to evaluate the prediction, one can compare both the actual prediction and the flipped prediction (all labels flipped from 0 to 1 and vice versa) and select the better prediction. With this methodology, the evaluation scores discussed earlier have been calculated and tabulated in Table 16.

Table 16: Evaluation metrics for K-Means classification of Inconel 718 beads

Evaluation Metric	*Score*
Accuracy	0.641 (64.1%)
Precision	0.967 (96.7%)
Recall	0.589 (58.9%)
False Positive Rate (FPR)	0.100 (10.0%)
F1	0.733 (73.3%)
MCC	0.364 (36.4%)

Generally for clustering problems, ROC-AUC curves/metrics are not calculated since such models are not probabilistic in nature. The calculation would effectively return a single point instead of a curve. Some authors have put forward methodologies to construct ROC curves for clustering models using point-shifting between clusters, but such approaches are considered to be out of the scope of this research [89].

The results for K-Means clustering indicate a fairly average result, with a decent accuracy of around 64% and a Matthew's correlation coefficient of about 0.36, which indicates a weak positive correlation (since it is closer to 0 than 1). The precision score is extremely high, indicating that the false negative rate of the model is very low (which is very good from a conservative safety point of view). In addition, it is seen that the false positive rate is around 10%, which could be on the borderline of being usable for the application being examined here.

Since K-Means clustering does not assign feature importances implicitly, a different methodology was implemented to calculate this. The cluster centroids (which have the same dimensions as the number of input parameters) were retrieved from the fitted model, and the sum of squared distances (SSDs) of each point from its respective centroid was calculated along each dimension (which represents each parameter individually). The idea was that a greater extent of clustering along a particular dimension (*i.e.*, smaller squared distance between datapoints and centroid) would imply greater importance of that particular dimension (parameter) in deciding the centroid location. This would give an approximate idea of feature importance. This approach can be run multiple times in order to construct confidence intervals. Using this methodology for 1000 runs, the SSDs with error bars are presented in Figure 62.

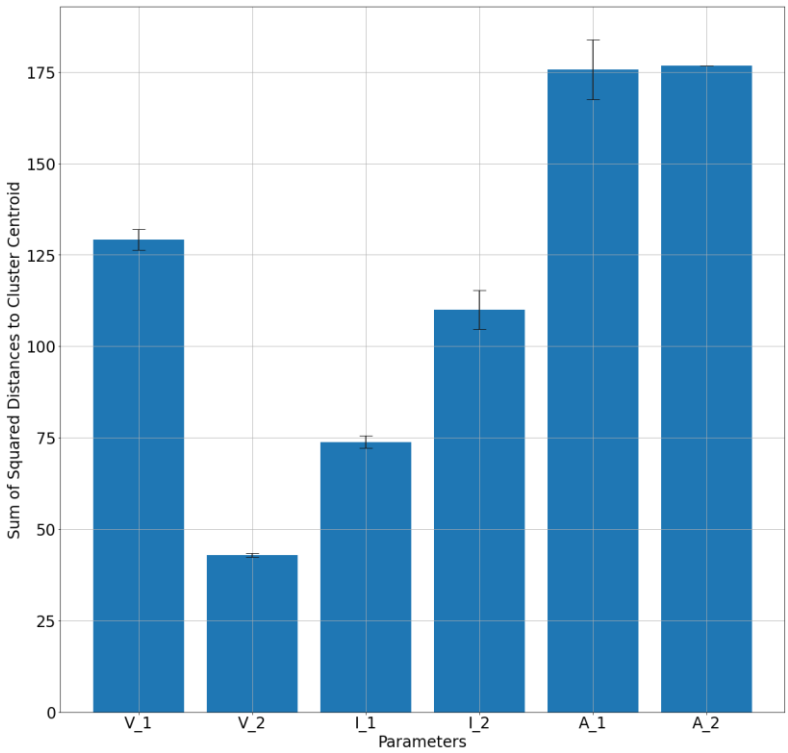


Figure 62: Sum of squared distances of Inconel 718 datapoints from respective clusters

From Figure 62, it is visible that V_2 (voltage pulse width variance) is the most important parameter for K-Means clustering of Inconel 718 beads. Interestingly, the audio parameters are shown to have the least importance. The possible reasons for this observation will be discussed in later sections.

In general, from this data analysis of Inconel 718, K-Means clustering is likely a poor algorithm for defect detection, since the correlation between prediction and ground truth is weak. This confirms the hypothesis from the elbow plot where a clear elbow was not seen, indicating the likelihood of poor analytical ability.

The next material analyzed is Invar 36, and the elbow plot for the dataset is shown in Figure 63. Once again a clear elbow is not seen, indicating that K-Means will yield a poor analysis. Using the same K-Means model detailed in Table 15 on the data of Invar 36 beads, the metrics obtained are shown in Table 17.

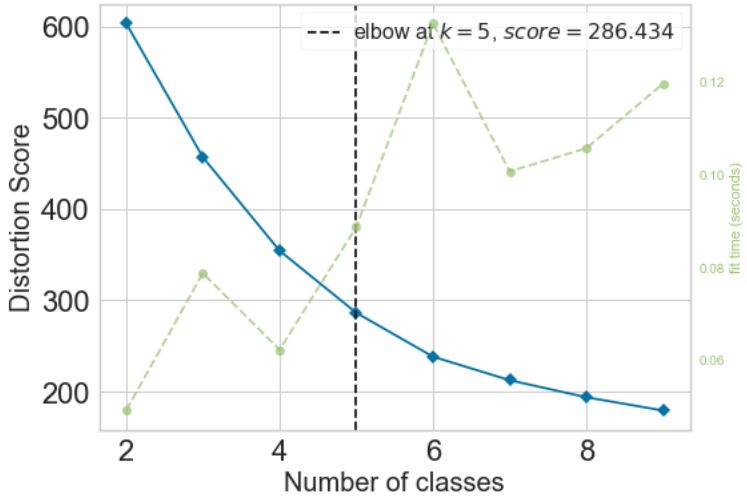


Figure 63: Elbow point plot for K Means Clustering for Invar 36

Table 17: Evaluation metrics for K-Means classification of Invar 36 beads

Evaluation Metric	*Score*
Accuracy	0.607 (60.7%)
Precision	0.291 (29.1%)
Recall	0.953 (95.3%)
False Positive Rate (FPR)	0.463 (46.3%)
F1	0.447 (44.7%)
MCC	0.369 (36.9%)

While the recall score is excellent (indicating good ability to detect positives in the dataset), most of the other scores are average-poor, with the MCC showing a weak positive correlation. The low precision shows that the quality of positive (defect) prediction is poor. The false positive rate is almost 50%, which is too high for any practical applications.

The relative feature importances are calculated using the SSD method as explained in the Inconel 718 section and shown in Figure 64. The figure shows that I_2 (current pulse width variance) is the most important parameter for clustering Invar 36 beads. While audio parameters are still of lower importance, they are now comparable to the voltage parameters, implying that the relative importance of audio parameters is higher. This will be further discussed in later sections.

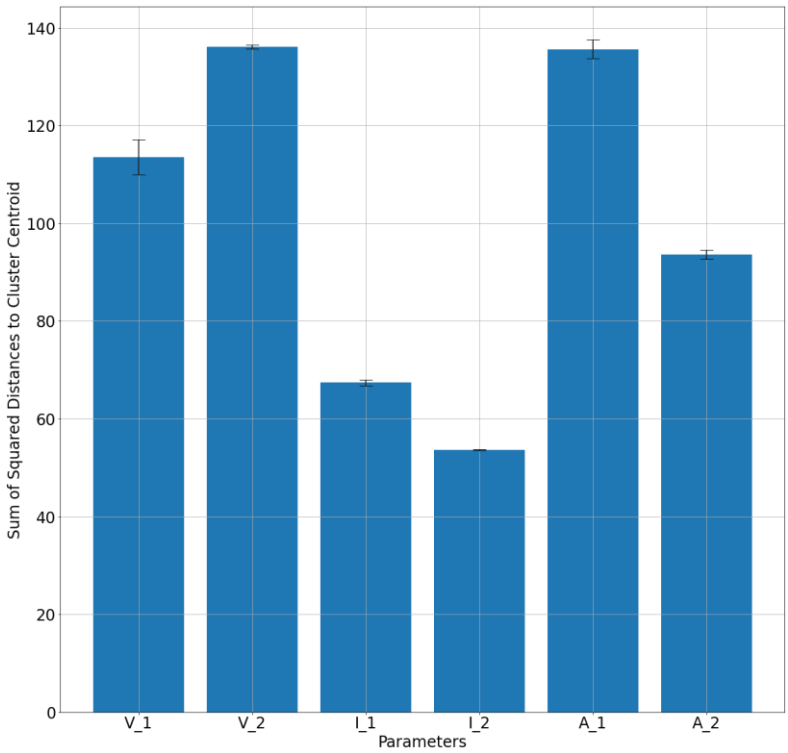


Figure 64: Sum of squared distances of Invar 36 datapoints from respective clusters

The final material analyzed is Inconel 625. The situation is a little different since all the beads deposited were defect free. The extreme class imbalance in the dataset meant that the performance of the model would be fairly inconclusive to verify its effectiveness. In spite of this, an elbow plot is constructed for this dataset and shown in Figure 65.

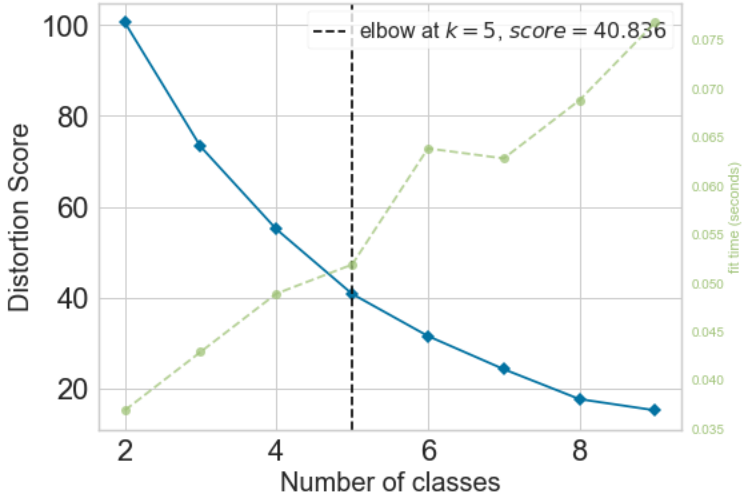


Figure 65: Elbow point plot for K Means Clustering for Inconel 625

In the interest of thoroughness, the model in Table 15 was run on the entire dataset using only current and voltage data (the number of beads for which audio data was available was far too low), and performance metrics were obtained as shown in Table 18. This set of results can serve as an

example of accuracy being misleading when evaluating a model. The model accuracy is extremely high, and yet the performance of the model itself is poor. It also shows how class balance is important and necessary to properly train a model.

Table 18: Evaluation metrics for K-Means classification of Inconel 625 beads

Evaluation Metric	*Score*
Accuracy	0.967 (96.7%)
Precision	0.000 (0.0%)
Recall	Undefined (no defect positives in ground truth)
False Positive Rate (FPR)	0.033 (3.3%)
F1	0.000 (0.0%)
MCC	Undefined (no defect positives in ground truth)

4.3.1.4 Artificial Neural Network Evaluation

Neural networks have high flexibility in their structure (in the form of a large number of hyperparameters) and are suitable for big data, which can become a useful characteristic when enough data is collected over a longer period of time [90].

Multilayer perceptrons (MLPs) were chosen for this research work since the nature of the data did not necessitate the usage of RNNs (recurrent neural networks) or CNNs (convolutional neural networks). The model architecture used in the current work is shown in Figure 66.

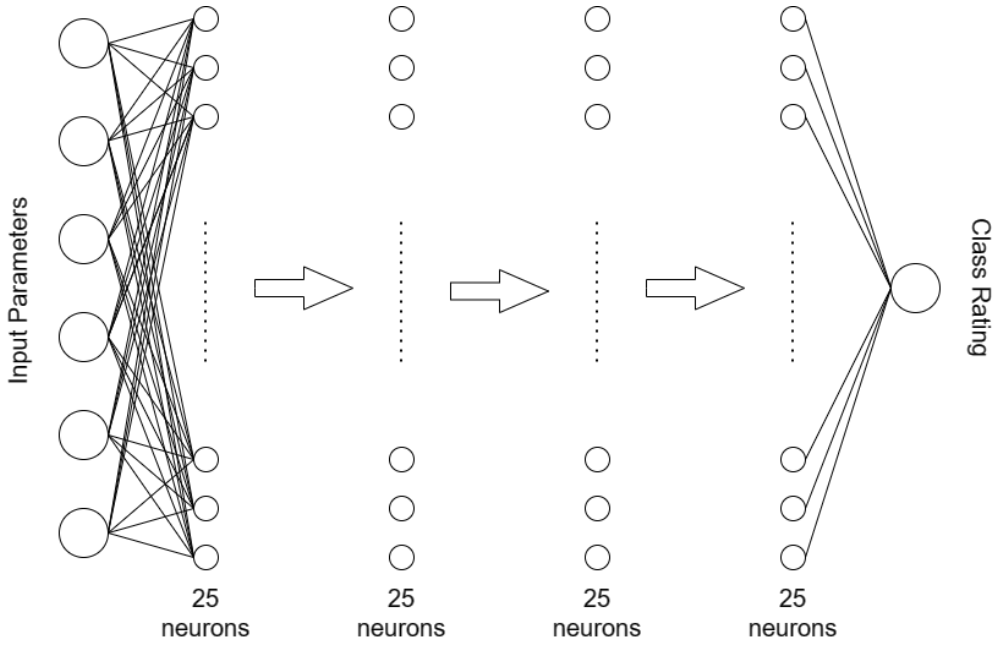


Figure 66: MLP architecture used for bead data analysis

The number of layers and neurons were selected after some trial runs using random subsets of the available bead data in order to prevent overfitting. The validation loss (the loss measured on

the validation set) was used to detect overfitting, and the complexity of the model was appropriately changed (either changing the number of layers or the number of neurons in each layer). The training and validation loss curves obtained for all three materials are shown in Appendix B. Models that were too simple led to underfitting, which gave highly inaccurate predictions since the learned function was not complex enough to capture the full relation between input parameters and output labels. Throughout the rest of this thesis, this model will be interchangeably referred to as either an MLP or an ANN model (since MLPs are essentially a subset of ANNs).

The hyperparameters of the model have been summarized in Table 19. While the earlier clustering and random forest approaches were of classification nature, the MLP was configured to run as a regression. The reasoning behind this was that modeling the nature 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.

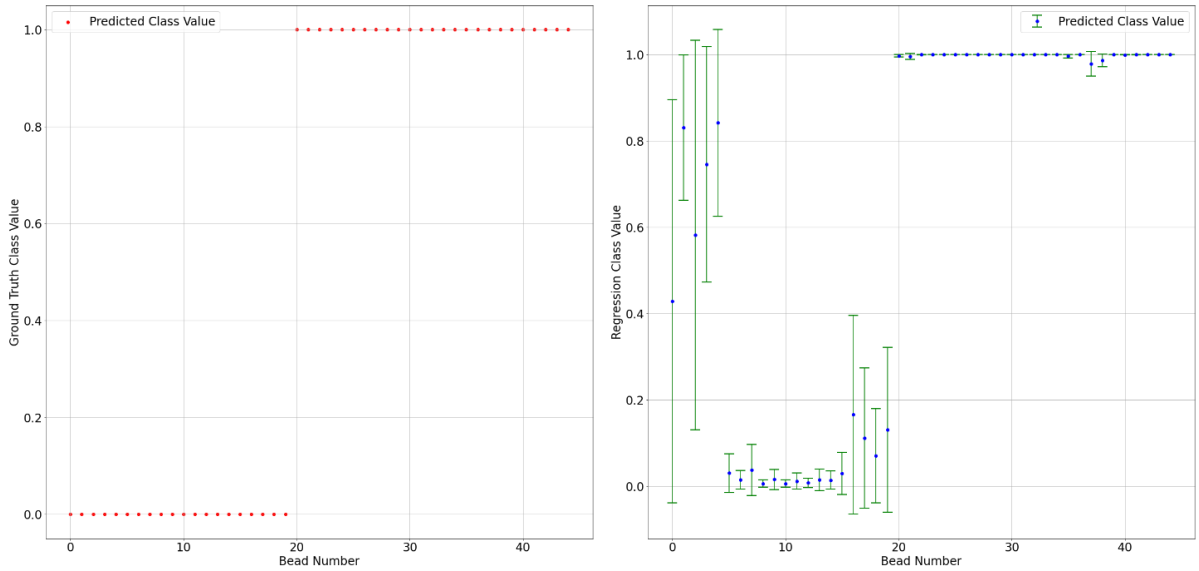
Table 19: Hyperparameters of the MLP model used

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

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 would be to apply a clustering method to the set of predicted labels (like K-Means clustering). Within this paradigm, evaluation would be done based on error metrics defined for classification. The advantage of this approach would be that parameter boundaries can be defined for defects easily by looking at the set of input parameters in each cluster. Both approaches will be examined in this section.

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 replacement (*i.e.*, some elements in the training set could be repeated), and the model is run on all these individual training sets to get multiple predictions [92]. The mean and variance of these predictions are used to make confidence intervals (akin to a brute-force approach) [93]. For the ANN analysis, bootstrapping was used to generate confidence intervals for statistical analysis.

Similar to earlier sections, Inconel 718 beads are examined first. The obtained predictions on the testing set with 95% confidence intervals along with ground truths are shown in Figure 67.



(a) (b)

Figure 67a-b: (a) Ground truth, (b) Predicted class values for testing Inconel 718 beads

As seen in Figure 67, many of the beads of the minority class (class label = 0) have predicted values that are close to the ground truth, with a few points being far. The confidence intervals also seem to indicate that the points far from ground truth have significant prediction variance, whereas the highly accurate points have very low prediction variance. The good performance seen with regard to the majority class (class label = 1) is expected, considering that larger number of majority class datapoints available for training.

The average error metrics (over all bootstrapped testing sets) are shown in Table 20. Evaluating the metrics against the class label range of 0 to 1, it is visible that the error values are fairly small, indicating decent performance. The RMSE is higher than the MAE due to the squaring of errors, leading to higher penalties for outliers. In the context of this thesis, outliers are important to observe because they indicate beads that are not easily classified as either clean or defective (*i.e.*, most likely to be classified as false positives/negatives). Thus it is of higher importance to detect the presence of outliers, meaning that RMSE will be the metric of importance when comparing the results of this analysis to those of other materials.

Table 20: Evaluation metrics for ANN regression of Inconel 718 beads

Evaluation Metric	*Score*
MSE	0.086 (95% CI {0.084, 0.089})
RMSE	0.286 (95% CI {0.282, 0.290})
MAE	0.092 (95% CI {0.089, 0.095})

The next step would be to apply K-Means clustering to the obtained set of results and examine the quality of clustering. Using the same model parameters as shown in Table 15, the error metrics along with their confidence intervals can be compiled. These metrics are shown in Table 21. The results are very good, with a high MCC indicating a strong positive correlation between prediction and original data. The false positive rate is around 20%, which is a little too high. The high precision and recall indicates good ability to predict positives (defects), which is expected, since it is the majority class.

Table 21: Evaluation metrics for K-Means classification of Inconel 718 bead regression results

Evaluation Metric	*Score*
Accuracy	0.894 (95% CI {0.887, 0.901})
Precision	0.844 (95% CI {0.835, 0.853})
Recall	0.972 (95% CI {0.961, 0.982})
False Positive Rate (FPR)	0.202 (95% CI {0.197, 0.208})
F1	0.903 (95% CI {0.893, 0.912})
MCC	0.796 (95% CI {0.781, 0.810})

It is possible to look at the set of bootstrapped predictions and define a probability of accurate prediction (since the clustering has converted the problem from regression to classification). The probability can be calculated simply as the ratio of number of accurate predictions to the total number of predictions (in this case, 1000).

A scatter plot of probabilities for each testing set bead can be seen in Figure 68. It is visible that the majority of beads have more than 95% probability of accurate prediction with the lowest probability being around 20%. It is also visible that many beads of the minority class also have a high prediction probability, which is good (since good predictions are obtained in spite of limited training data).

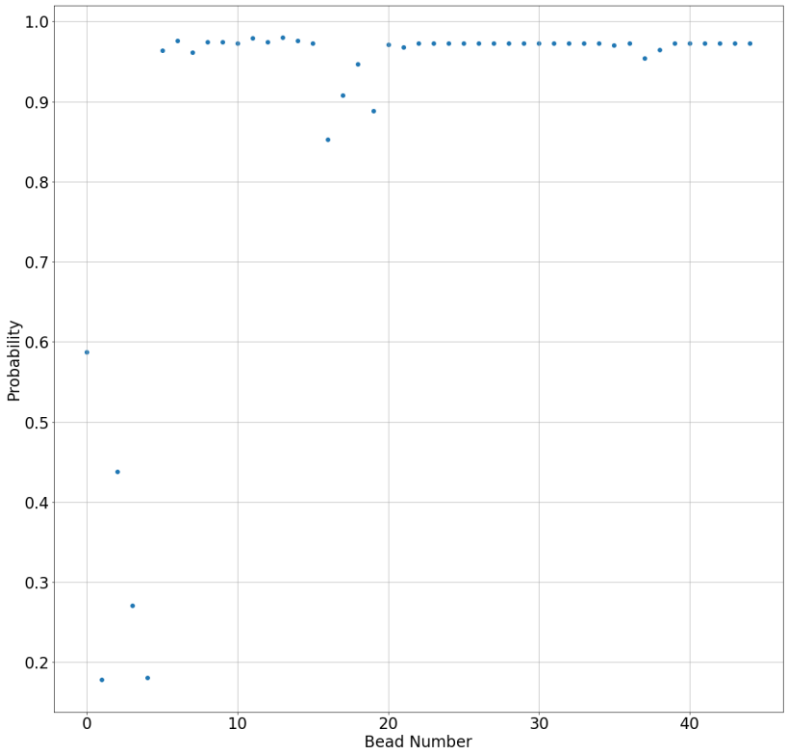


Figure 68: Prediction probabilities of Inconel 718 beads

Now that the results have been compiled, it is meaningful to look at the scaled input parameters and probability of prediction side by side. This has been shown in the plots in Figure 69. The figure shows that there is no clear visible trend in the input parameters that corresponds to the dips in probability.

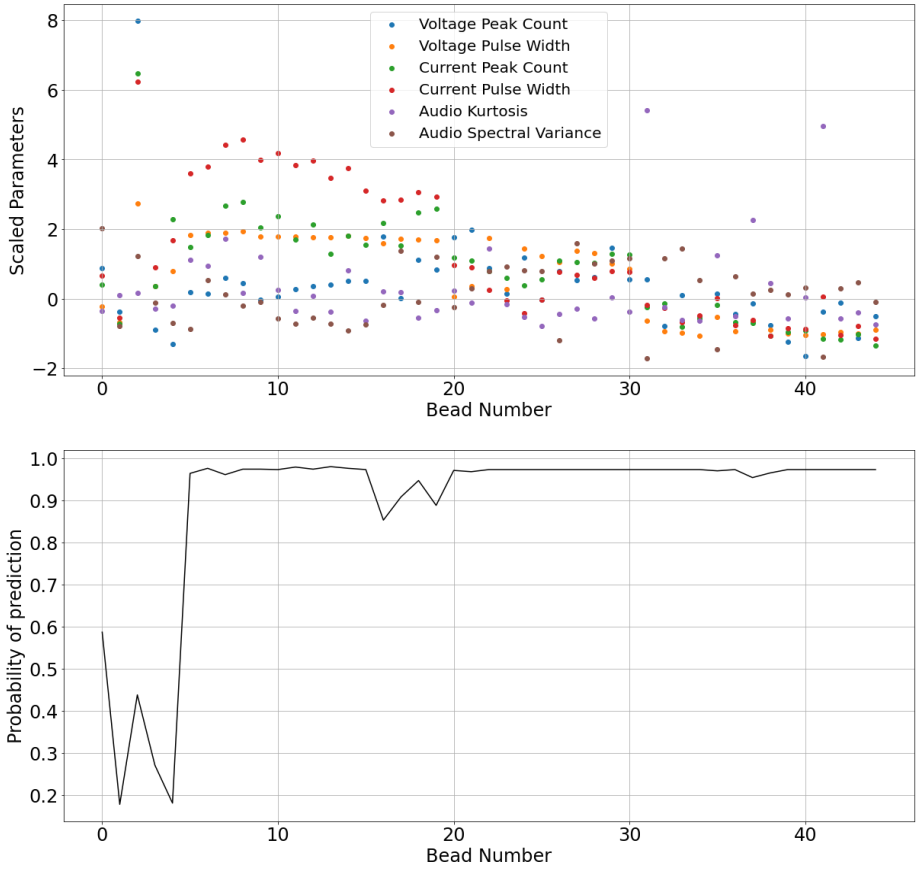


Figure 69: Scaled input parameters and probabilities of accurate prediction for Inconel 718 beads

In order to evaluate the relative importance of each input parameter, SHAP values (Shapley Additive Explanations) are used. SHAP gives a list of feature importances that can indicate the contribution of each feature to the prediction made by a model [101]. Python has an inbuilt module to calculate such SHAP values, which was used to evaluate the ANN. The results are shown in Figure 70.

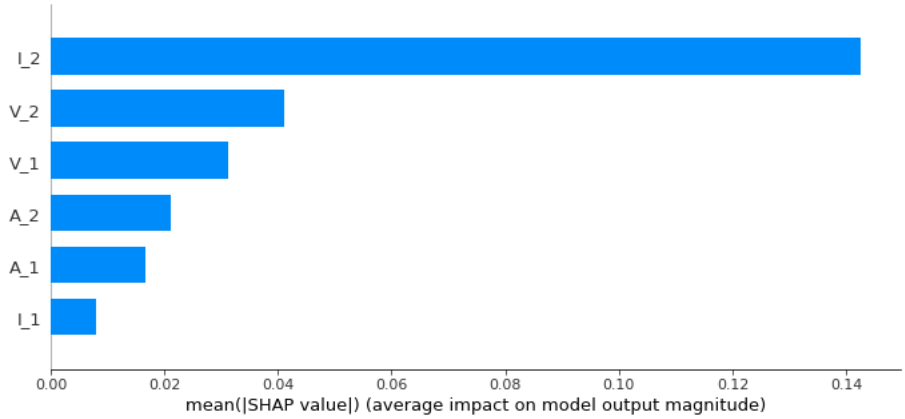
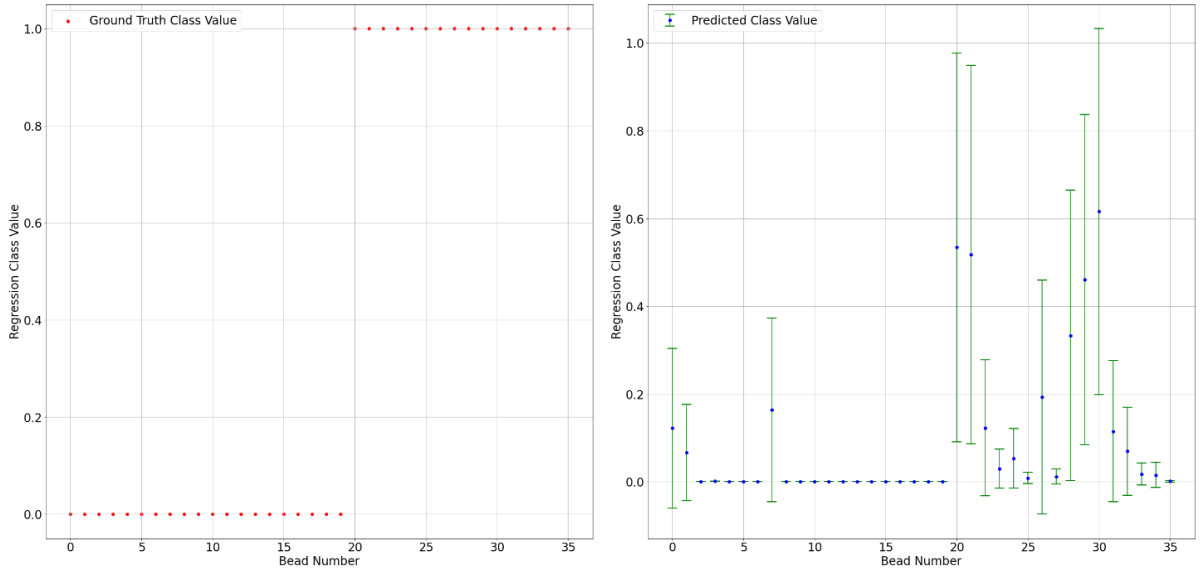


Figure 70: SHAP feature importances for prediction of Inconel 718 using ANN

Figure 70 shows that the most important parameter by far is I_2, which is the welding current peak width variance. Interestingly, the audio parameters are shown to be slightly more important than one of the current parameters, which is unlike the importances obtained from the random forest evaluation in Section 4.3.1.2. Voltage parameters are seen to retain their importance relative to other parameters across both methods (random forests and ANNs).

Subsequently, Invar 36 beads are analyzed. The predictions (with 95% confidence intervals) and ground truth of the testing set are shown in Figure 71. Examining the minority class datapoints (class value = 1), it is apparent that the prediction quality is poor. Not only are most of the predicted class values below 0.5 (which would indicate being closer to a clean bead than a defective bead), but the confidence intervals are also large, indicating high uncertainty. For the majority class beads (class value = 0), however, the prediction appears to be mostly accurate with very small error bars (which is to be expected, since a larger number of majority class datapoints are available for training).



(a) (b)

Figure 71a-b: Ground truth and predicted class values for testing Invar 36 beads

The average of error metrics over the bootstrap testing sets is shown in Table 22. All the metrics are seen to be larger than the corresponding metrics for Inconel 718, indicating comparatively worse performance. The values are also significant when considering a range of 0 to 1, indicating potential for a lot of improvement. RMSE was earlier chosen as the most important metric due to the emphasis on outliers, and it is seen that the value has nearly doubled. The confidence intervals are fairly narrow however, indicating a low extent of uncertainty in the values.

Table 22: Evaluation metrics for ANN regression of Invar 36 beads

Evaluation Metric	*Score*
MSE	0.358 (95% CI {0.355, 0.362})
RMSE	0.597 (95% CI {0.595, 0.600})
MAE	0.368 (95% CI {0.365, 0.371})

Similar to Inconel 718, the K-Means clustering method is applied to the regression results of the testing dataset (using the same parameters from Table 15), and the final classification obtained is

examined against the ground truth labels to look for any improvements. The error metrics for this classification are shown in Table 23.

Table 23: Evaluation metrics for K-Means classification of Invar 36 bead regression results

Evaluation Metric	*Score*
Accuracy	0.611 (95% CI {0.605, 0.617})
Precision	0.846 (95% CI {0.831, 0.860})
Recall	0.271 (95% CI {0.259, 0.282})
False Positive Rate (FPR)	0.117 (95% CI {0.100, 0.135})
F1	0.354 (95% CI {0.346, 0.363})
MCC	0.245 (95% CI {0.230, 0.259})

It is seen that applying clustering to the results leads to average-poor results. The high precision indicates that the positive (defect) detection done by the model is of high quality (*i.e.*, low false negative prediction). The recall score is poor, indicating 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 low, indicating usability from an industrial standpoint. The MCC is closer to 0 than 1, indicating a weak positive correlation between ground truth and predictions.

It is also meaningful to look at the evaluation probabilities of the testing set beads. Using the same procedure as with Inconel 718, the probability plot obtained is shown in Figure 72. There are similarities to the Inconel 718 plot when looking at the lowest probabilities in the figure. It is visible here that the lower limit of probability is very small, being close to 10% probability (for Inconel 718, the lowest probability was 20%). Comparing with Figure 68, it is also visible that the majority class datapoints have high prediction probability, whereas nearly all the minority class datapoints have probabilities between 10% and 60%.

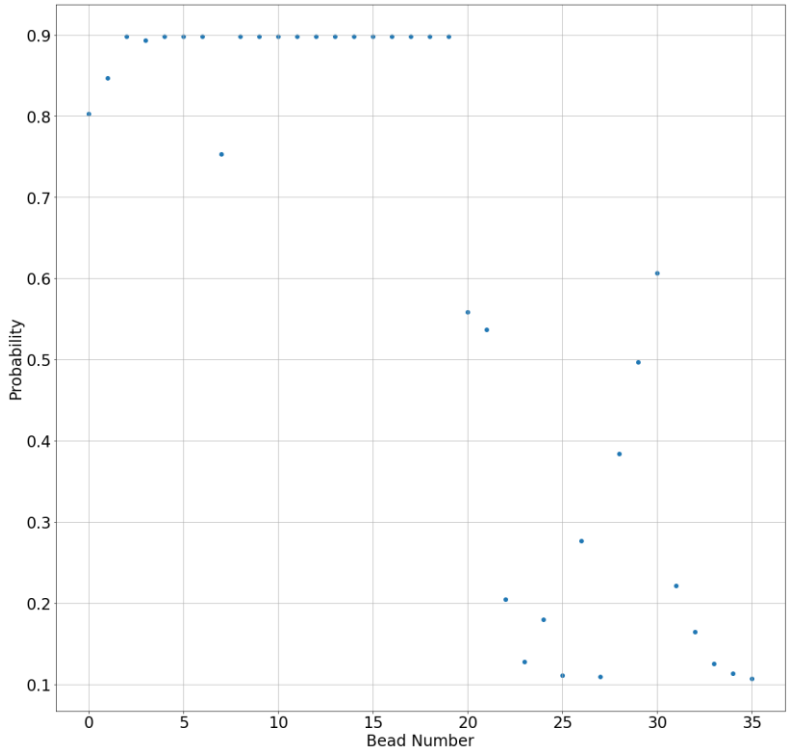


Figure 72: Prediction probabilities of Invar 36 beads

The next step would be to examine the probability side by side with the scaled input parameters to look for any visible indications of probability drops. This is shown in Figure 73. The figure shows that there isn't a clear visual indication in the input parameters.

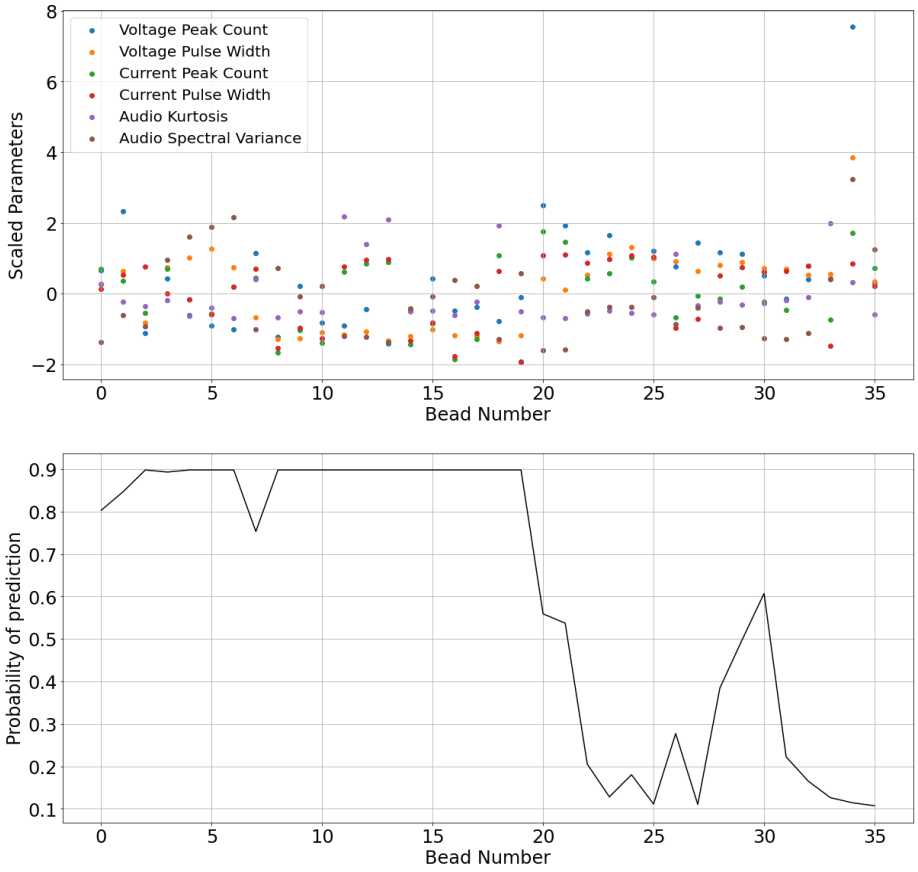


Figure 73: Scaled input parameters and probabilities of accurate prediction for Invar 36 beads

The SHAP values are calculated in order to assess the importance of each parameter (similar to Inconel 718). The results are shown in Figure 74.

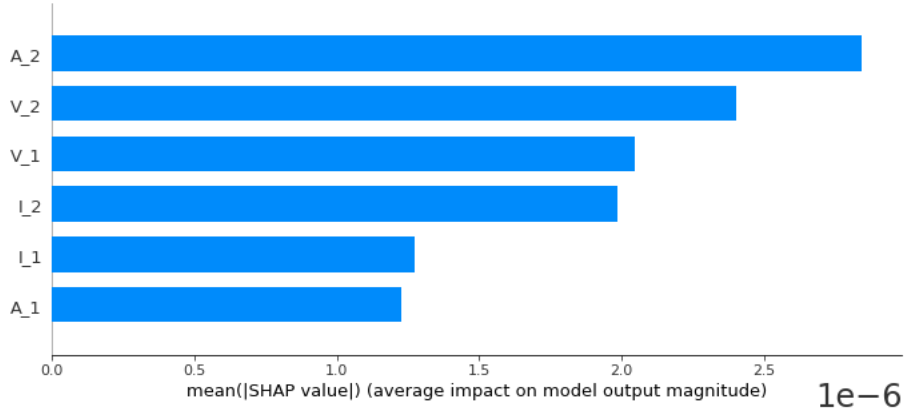


Figure 74: SHAP feature importances for prediction of Invar 36 using ANN

Contrary to the results from random forests in Section 4.3.1.2, the most important feature here is A_2, which is the audio spectral variance. This is followed by both voltage parameters, both current parameters, and finally A_1, which is audio kurtosis. Voltage parameters retain high importance (relative to other input parameters) in both random forests and ANN models. It is interesting to note that the absolute values on the x-axis (average impact on model output magnitude) are much smaller than that of Inconel 718 (10^5 times smaller). The exact reason for this is not known, but since relative importance is what is important, it is not of too much concern.

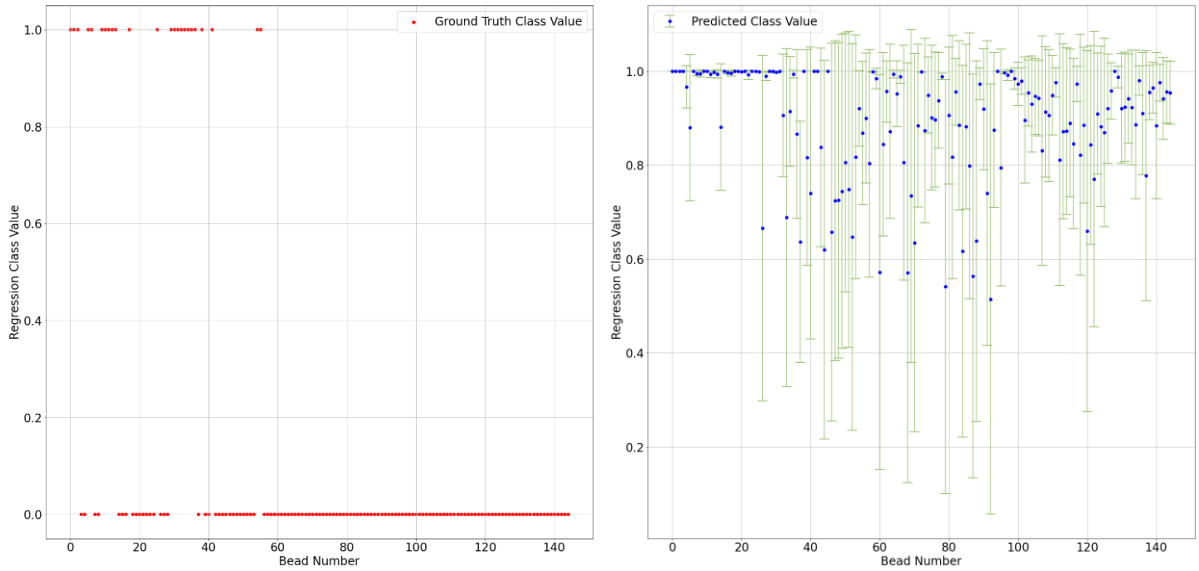
For Inconel 625, analysis using audio data in the abovementioned manner did not seem useful due to the low number of data points available (22), while follow up research work could involve further testing with more designed experiments that introduces more data with defective beads.

4.3.2 Cross-Material Evaluation

Section 4.3.1 detailed some single-material evaluations that gave an idea of how accurate random forests, K-Means classifiers, and multilayer perceptrons can be at predicting different kinds of defects. An interesting avenue of inquiry, however, would be the possibility of doing this kind of analysis on a cross-material scale (*i.e.*, across different materials). More specifically, 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). There are different levels of depth possible for such an investigation, ranging from using different materials that show similar kinds of defects, to using different materials that show different kinds of defects.

In this thesis, ANN models trained on Inconel 718 and Invar 36 (separately) have been tested against the other two materials (Inconel 718/Invar 36 and Inconel 625). Both regression and classification steps were completed and the resulting predictions were compared to ground truth to extract error metrics. Training on Inconel 625 was omitted due to the lack of sufficient datapoints that included an audio component.

First, an ANN model trained on Inconel 718 using the hyperparameters in Table 19 is tested on Inconel 625 and Invar 36. In both cases, bootstrapping was done (250 samples) in order to construct confidence intervals. The regression results for Invar 36 are shown in Figure 75. Subsequently, clustering is applied on the regression results, leading to the evaluation metrics shown in Table 24.



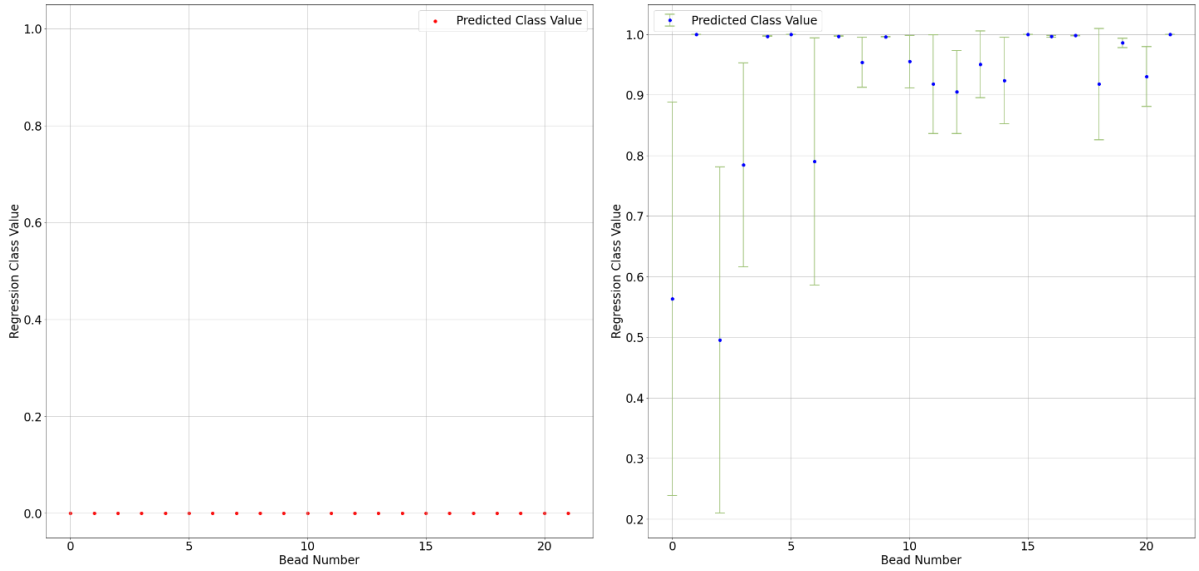
(a) (b)

Figure 75a-b: (a) Ground truth, (b) Predicted class values for Invar 36 beads with Inconel 718 data

Table 24: Evaluation metrics for ANN prediction of Invar 36 beads with Inconel 718 data

Evaluation Metric	*Score*
Accuracy	0.258 (95% CI {0.245, 0.272})
Precision	0.181 (95% CI {0.178, 0.184})
Recall	0.962 (95% CI {0.954, 0.969})
False Positive Rate (FPR)	0.881 (95% CI {0.865, 0.898})
F1	0.303 (95% CI {0.299, 0.307})
MCC	0.086 (95% CI {0.076, 0.097})

The regression results for Inconel 625 are shown in Figure 76. Clustering the regression results leads to the evaluation metrics shown in Table 25.



(a) (b)

Figure 76a-b: (a) Ground truth, (b) Predicted class values for Inconel 625 beads with Inconel 718 data

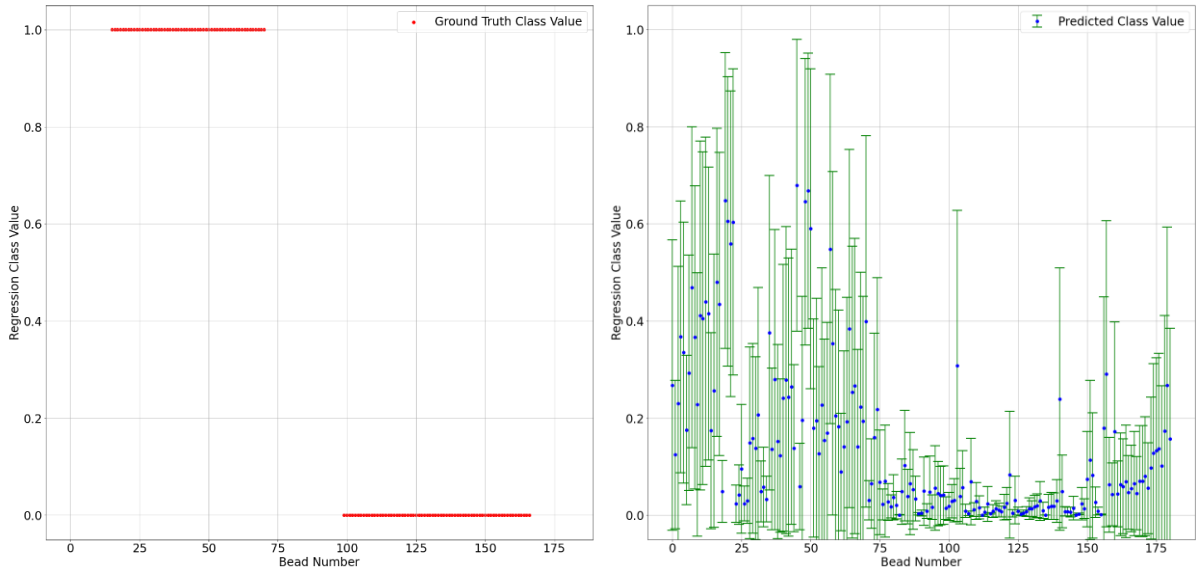
Table 25: Evaluation metrics for ANN prediction of Inconel 625 beads with Inconel 718 data

Evaluation Metric	*Score*
Accuracy	0.078 (95% CI {0.069, 0.087})
Precision	0.000 (95% CI {0.000, 0.000})
Recall	N/A
False Positive Rate (FPR)	0.922 (95% CI {0.913, 0.931})
F1	0.000 (95% CI {0.000, 0.000})
MCC	N/A

When looking at the performance on Invar 36, it is visible that the metrics are very poor. At this point however, it is important to note that not only are the two materials used in training and testing different, but the defect classes being analyzed are completely different too (overflow vs. hot cracking). The MCC is very close to 0, indicating a nearly non-existent positive correlation between prediction and ground truth. The recall and false positive rates are both very high, indicating that the model considers most examples to be positive (defective). The reasoning for this kind of prediction is discussed later on in the chapter. The acquisition of more datapoints for each material could lead to better results, since the model would be better equipped to find the relationship between input parameters and class labels.

Looking at Inconel 625, a completely different picture is painted. The lack of defective datapoints leads to two undefined metrics (similar to what is seen in Table 18) and other extremely poor results. The accuracy is nearly 0 and the false positive rate is nearly 100%. It is important to note, however, that the acquisition of a sufficient amount of Inconel 625 data with defects deliberately induced can open up this avenue for further investigation.

Next, a model trained on Invar 36 is tested on Inconel 718 and Inconel 625. The regression results for Inconel 718 are shown in Figure 77. Clustering is applied on the regression results, giving classification error metrics shown in Table 26.



(a) (b)

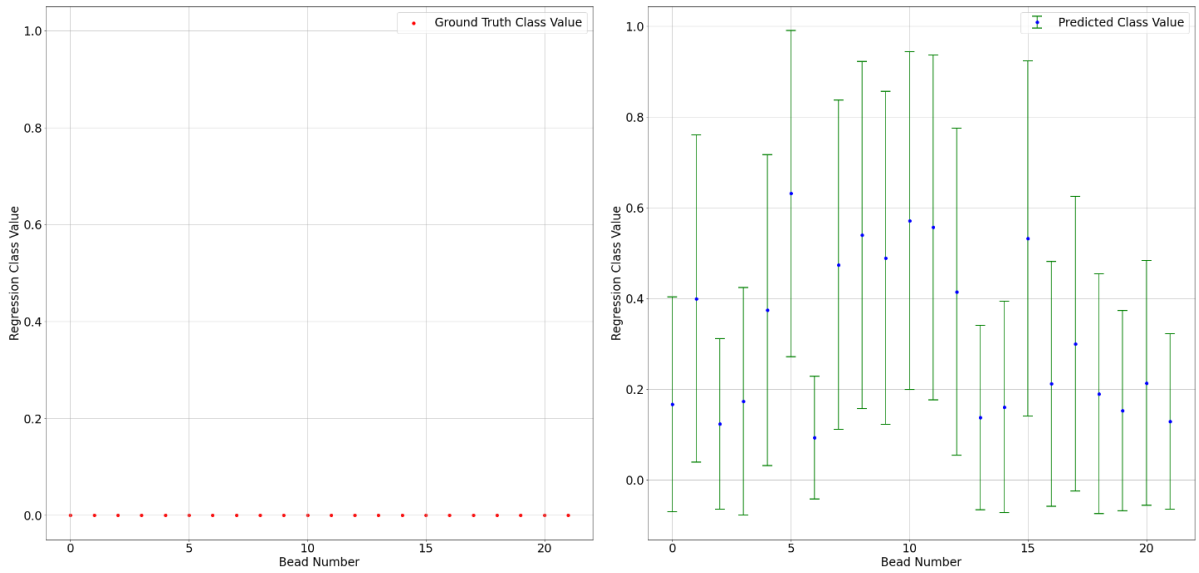
Figure 77a-b: (a) Ground truth, (b) Predicted class values for Inconel 718 beads with Invar 36 data

Table 26: Evaluation metrics for ANN prediction of Inconel 718 beads with Invar 36 data

Evaluation Metric	*Score*
Accuracy	0.271 (95% CI {0.260, 0.281})
Precision	0.906 (95% CI {0.884, 0.928})
Recall	0.144 (95% CI {0.131, 0.158})
False Positive Rate (FPR)	0.093 (95% CI {0.071, 0.116})
F1	0.234 (95% CI {0.216, 0.252})
MCC	0.064 (95% CI {0.043, 0.084})

The results of Inconel 718 are poor. The false positive rate is low (less than 10%), which is encouraging. The precision is very high, indicating that the model is good at selecting positives of high quality (which is corroborated by the low false positive rate). The recall is low, however, 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. The possibility exists that more data could give the model more information to work with, leading to better results.

The regression results for Inconel 625 are shown in Figure 78. Applying clustering on the regression results gives Table 27.



(a) (b)

Figure 78a-b: (a) Ground truth, (b) Predicted class values for Inconel 625 beads with Invar 36 data

Table 27: Evaluation metrics for ANN prediction of Inconel 625 beads with Invar 36 data

Evaluation Metric	*Score*
Accuracy	0.681 (95% CI {0.643, 0.720})
Precision	0.000 (95% CI {0.000, 0.000})
Recall	N/A
False Positive Rate (FPR)	0.319 (95% CI {0.280, 0.357})
F1	0.000 (95% CI {0.000, 0.000})
MCC	N/A

Interestingly, the accuracy score for the prediction of Inconel 625 is somewhat good (almost 70%). The precision is 0 and MCC and recall are undefined due to the lack of positives (defects) in the ground truth. The false positive rate is still too high (almost 32%) for practical applications. The low number of datapoints being predicted (22) means that it is difficult to draw conclusions from the data, but the fact that the accuracy was decently high means there is potential for good performance (with more datapoints being involved in both training and testing).

From all the predictions done in this section in Figure 75, Figure 76, Figure 77 and Figure 78, it is visible that the error bars are large across the entire prediction (as opposed to the single material predictions where the error bars were only large in the minority class predictions). This is to be expected, considering the fact that there is much lower commonality between the training and testing sets in a cross-material prediction as compared to a single-material prediction.

From Section 4.3.1 it is seen that random forests performed well on single material predictions as compared to ANNs. This means it is useful to examine the possibility of doing a similar cross-material analysis using random forests. Following the same methodology from the single material analysis, the first set of results is obtained by training on Inconel 718 and testing on Invar 36 and Inconel 625. The results are shown in Table 28 and Table 29.

Table 28: Evaluation metrics for Random Forest classification of Invar 36 beads with Inconel 718 data

Evaluation Metric	*Score*
Accuracy	0.172 (17.2%)
Precision	0.162 (16.2%)
Recall	0.958 (95.8%)
False Positive Rate (FPR)	0.983 (98.3%)
F1	0.277 (27.7%)
MCC	-0.066 (-6.6%)

Table 29: Evaluation metrics for Random Forest classification of Inconel 625 beads with Inconel 718 data

Evaluation Metric	*Score*
Accuracy	0.000 (0.0%)
Precision	0.000 (0.0%)
Recall	N/A
False Positive Rate (FPR)	1.000 (100.0%)
F1	0.000 (0.0%)
MCC	N/A

It is seen from Table 28 and Table 29 that even with the random forests approach, the results are poor. For Inconel 718, the accuracy is found to be very low (close to 20%), and the MCC is negative and close to 0, indicating a very weak negative correlation (it may even be considered as effectively having no correlation between predictions and ground truth due to how close it is to 0). The recall score is very 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 very high (above 98%), indicating the tendency of the model to label most samples it sees as positive. For Inconel 625, the lack of defective data causes poor results as expected. The false positive rate is 100%, indicating a perfectly inaccurate prediction.

The next set of results is obtained by training on Invar 36 and testing on Inconel 718 and Inconel 625. The results are shown in Table 30 and Table 31.

Table 30: Evaluation metrics for Random Forest classification of Inconel 718 beads with Invar 36 data

Evaluation Metric	*Score*
Accuracy	0.199 (19.9%)
Precision	0.636 (63.6%)
Recall	0.093 (9.3%)
False Positive Rate (FPR)	0.267 (26.7%)
F1	0.162 (16.2%)
MCC	-0.198 (-19.8%)

Table 31: Evaluation metrics for Random Forest classification of Inconel 625 beads with Invar 36 data

Evaluation Metric	*Score*
Accuracy	0.545 (54.5%)
Precision	0.000 (0.0%)
Recall	N/A
False Positive Rate (FPR)	0.455 (45.5%)
F1	0.000 (0.0%)
MCC	N/A

Table 30 and Table 31 indicate that even with switching to Invar 36 for training the model, the results remain poor. For Inconel 718, the accuracy is seen to have slightly improved over Table 28 (an increase of about 3%), and the precision has improved significantly (an increase of almost 50%). The recall score has become much worse, indicating that the model has become too stringent with finding positives. This is also reflected in the false positive rate, which has reduced to around 27% (which is still too high for practical use, but much less than what is seen in Table 28). The MCC is negative and higher in magnitude than in Table 28, indicating a stronger negative correlation between predictions and ground truth (the value is objectively still indicative of a weak correlation due to a magnitude < 0.5). As seen earlier, the results for Inconel 625 are inconclusive/poor due to the lack of defective datapoints.

4.4 Overview of Nickel Alloy Analysis

Based on the obtained results from Sections 4.3.1 and 4.3.2, it is clear that some relationships could be established between the different materials that were analyzed. While the prediction qualities were poor in some scenarios, meaningful conclusions could still be drawn based on them. The inferences from the above results are presented below.

4.4.1 Relationships between Defect Occurrence and Input Parameters

The collections of welding voltage and welding current (which were the input parameters for all the ML models apart from audio-based parameters) corresponding to clean and defective beads for each material have been separately presented in Table 32, Table 33 and Table 34. The confidence intervals are presented as percentages of the mean to help with visualizing their size. The values of the scaled input parameters themselves have been included in Appendix C along with the confidence intervals.

Table 32: Input parameters for Inconel 718 beads

Parameters	V_{avg}	I_{avg}
Mean (clean)	20.60	155.43
Mean (defective)	18.86	155.80
95% CI (clean)	3.45%	5.09%
95% CI (defective)	1.33%	1.54%

Table 33: Input parameters for Invar 36 beads

Parameters	V_{avg}	I_{avg}
Mean (clean)	16.12	137.50
Mean (defective)	16.87	147.76
95% CI (clean)	1.47%	2.24%
95% CI (defective)	7.50%	10.52%

Table 34: Input parameters for Inconel 625 beads

Parameters	V_{avg}	I_{avg}
Mean (clean)	20.30	185.65
Mean (defective)	-	-
95% CI (clean)	1.28%	1.00%
95% CI (defective)	-	-

From Table 32, Table 33 and Table 34 it is possible to calculate the average input power for the clean and defective beads of each material. The experiments conducted involved a relatively small range of traverse speeds, and thus an algebraic average of traverse speeds were taken for the defective and clean bead sets of each material. Using the average 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 average power and AEL values of clean and defective beads for all materials are shown in Table 35.

Table 35: Average input power and AEL values of all materials

Material	*Clean*		*Defective*	
	Power (W)	*Arc energy/length (J/mm)*	*Power (W)*	*Arc energy/length (J/mm)*
Inconel 718	3201.88	291.08	2938.61	322.57
Invar 36	2216.95	308.84	2492.30	536.46
Inconel 625	3769.24	339.62	-	-

Comparing the AEL values of clean and defective beads, it is seen from Table 35 that both Inconel 718 and Invar 36 have correspondence between defect occurrence 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. Thus, increased heat input would exacerbate the issue and lead to increased probability of defects. For Inconel 718, literature suggests that increased heat input can lead to cracking due increased segregation of elements, leading to the formation of brittle Laves phases, which helps corroborate the observations made here [105, 106].

From Section 4.1 it is shown 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 shown in Table 35 will be used to try to explain the trends in error metrics shown in Section 4.3.2.

4.4.2 Predictions between Inconel 718 and Invar 36

Consider an ML model trained on Inconel 718 data. This model would likely learn the fact that welding voltage and welding current parameters that lead to high arc power typically correspond to clean beads, whereas the parameters that correspond to lower input power typically correspond to defective beads (the exact threshold is difficult to identify, but the average powers of both bead sets are known from Table 35). It is important to note that since traverse speed was not passed as a parameter, the model should not have information about AEL. If the power range of Inconel 718 is compared to the power range of Invar 36, it is immediately visible 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 Table 24 and Table 28, 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 done on Inconel 718 in Sections 4.3.1.2, 4.3.1.3 and 4.3.1.4. The plots in Figure 57, Figure 62 and Figure 70 clearly show the low feature importance of audio parameters as compared to welding voltage and welding current.

The reason for this low importance 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 as 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 35, it is clear that such a model would likely learn the fact that welding voltage and welding current parameters corresponding to high input power correspond to defective beads, and the parameters corresponding to low input power correspond 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 35). 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 Table 26 and Table 30. 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). This seems 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 about the AEL due to the lack of a welding speed parameter in the input, if the AEL is taken as 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 as 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 Sections 4.3.1.2 and 4.3.1.4 (supervised single material analysis) 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 refer to the feature importances of the single-material analysis done on Invar 36 in Sections 4.3.1.2, 4.3.1.3 and 4.3.1.4. The plots in Figure 59, Figure 64 and Figure 74 show that the relative importance of audio features is slightly increased in comparison to Inconel 718.

This may be understood when considering the nature of defect 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 can cause significant variation in the instantaneous contact tip distance of the welding torch from 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 as 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 metric of the precision and recall since it is the harmonic mean of both metrics (as mentioned in Table 10). The scores are shown in Table 36.

Table 36: Precision, recall and F1 scores of cross-material predictions

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 as 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 about 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 about 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.

4.4.3 Predictions on Inconel 625

The cross-predictions on Inconel 625 were not examined in detail earlier due to the low number of datapoints and lack of defective datapoints. On the one hand, going off of the trend of AEL seen in Table 35, 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 as compared to the defective beads. This expected result is confirmed when looking at Table 25, Table 27, Table 29 and Table 31 in Section 4.3.2. 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 37.

It is clear from the table that both ANN and RF models trained on Inconel 718 tend to excessively predict positives (defects), whereas the models trained on Invar 36 tend to predict a reduced number of positives. This is exactly in line with the hypothesis that training on Inconel 718 would lead to overprediction of positives whereas training on Invar 36 would lead to the opposite, which further verifies the line of reasoning pursued to explain the connection between sensor responses and defect prediction.

Table 37: 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

4.5 Limitations of the ML Approach and Future Work

The primary limitations seen in the analysis of ML predictions in this chapter stem from the lack of datapoints and class composition imbalance. The problem of datapoints is universal for ML models since the availability of more data enables a more comprehensive "learning" of relations between input parameters and output labels. Haykin [102] stated that an ANN requires at least 10 times the number of network weights in order to get a meaningful prediction. Looking at the ANN in Section 4.3.1.4, the number of weights is

$$W = 6*25 + 3*(25^2) + 25 = 2050$$

which implies that at least 20,500 datapoints would be needed for a good prediction. This is far higher than the number of datapoints that was actually used, which was on the order of 10^2, implying that the predictions have to be taken with a grain of salt.

Random forest models also require a moderate-large amount of data for good performance, as shown in research works reported in literature [103, 104]. A relation to the number of model parameters similar to ANNs was not found, but the general number of datapoints used in the aforementioned studies were at least an order of magnitude larger than the datasets used in this study.

The issue of class imbalance can be mitigated to some extent using bootstrapping (as done in this research work), but the lack of unique datapoints can still pose problems when attempting to properly understand the relations between welding input parameters and defect presence. A shortage of clean bead data is comparatively easy to resolve if ideal conditions for the particular material are known (this was the case for Inconel 718, but time constraints made it difficult to obtain sufficient amounts of clean bead data). A shortage of defective data can be somewhat trickier due to the variety of circumstances that can lead to such defects. Overrepresenting one particular cause of a defect by repeating a particular set of conditions can cause the model to generate wrong conclusions and misdiagnose alternative process aberrations that lead to the same defect.

Another limitation is the number of sensors involved in the monitoring process. In this study, welding voltage, welding current and audio signals were recorded, but additional signals like spectroscopic data, high speed camera images, etc. can provide a wider range of information about the welding arc and substrate. For instance, spectroscopic data can give some information about the arc temperature and elemental composition in the vicinity of the melt pool, thereby giving better information about thermal input into the bead, as well as the possible presence of contaminants in the melt pool. There is also the possibility of extracting more parameters from the available recorded signals to possibly capture variations that are missed by the current set of parameters.

4.6 Applicability to Other Classes of Alloys

This research study specifically focused on nickel alloys. When considering other classes of alloys like carbon steels or titanium alloys, it remains to be tested whether such cross-material analyses can give some meaningful results.

It is likely that topographic defects (similar to bead overflow) would be easier to detect as compared to internal defects (like hot cracking) due to the fact that surface changes in a bead have

a significant impact on the stability of the arc, which is a response that is independent of the chemistry of the deposited material (interestingly, this is the opposite of what was hypothesized in Section 4.2.2). This makes it more likely for a model trained on other materials to classify such an event as an anomaly.

In order to better expand the applicability of ML models to various alloy classes for internal defects, it would be needed to conduct further research on obtaining more parameters from recorded sensor data that can be made independent of the material itself. Barring this, the best approach would be to use incremental learning to create a continuously learning model that expands its repertoire of information as and when it encounters a new material being deposited, akin to the work of Li *et al.* [43].

5 Conclusions and Recommendations

Through the course of this research work, many interesting connections were found between various sensor responses and defects in additively manufactured pieces. Physically consistent reasonings were deduced for most observations, but some still need further work. The primary conclusions that can be made from this work are as follows.

1. In both Invar 36 and Inconel 718, increased arc energy per unit length is found to be correlated with the presence of defects. This is consistent with both literature and the physics of defect formation since bead overflow (in Invar 36) is indeed caused by excessive heating of the melt pool and hot cracking in Inconel 718 can be caused by Laves phase formation due to increased heat input.
2. Unsupervised K-Means clustering is found to be poor at conducting predictions on unlabeled welding parameter data. Out of the two supervised ML models used for predictions on the defect status of WAAM prints, the random forest model is found to have better results as compared to the ANN.
3. The relative importance of audio-based parameters (kurtosis and spectral variance) are found to be low when training/testing on Inconel 718, whereas it increases when training/testing on Invar 36. This is sensible when considering the nature of defect found in each material, since Inconel 718 shows hot cracking, which is an internal defect that has minimal effects on the electric arc (which is the primary source of sound). Invar 36 shows bead overflow, which affects the contact tip distance and can thus affect the stability of the arc, which has a prominent audio signature.
4. 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 shows similarities to the arc energy/length values, which would require knowledge of the welding speed. Even when considering the possibility of audio features causing this effect, their relative importance does not seem significant enough to explain this.
5. Cross-material predictions between Inconel 718 and Invar 36 are found to heavily depend on the nature of defect in either material (in connection with point 3). Training with Inconel 718 is found to lead to underprediction of defects in Invar 36 and Inconel 625 whereas training with Invar 36 is found to lead to overprediction of defects in Inconel 718 and Inconel 625.

6. Class imbalance is found to have a significant impact on the prediction capabilities of both supervised models. In both random forest and ANN models with a reduced amount of minority class datapoints in training (1/3rd), the testing performance of Inconel 718 was found to be better than Invar 36. This could mean that a minority in clean beads is less impactful on performance reduction as compared to a minority in defective beads. The effect of type of class imbalance on cross-predictions seems to be minimal, with the change in MCC being very small.
7. The usage of welding current, welding voltage and audio signals is found to give 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.

From the above conclusions, it is clear that the research questions outlined in Section 1.4 have been answered to a fair extent. The results indicate that the potential for a cross-material defect prediction model indeed exists, and resolving the problems with the current approach can help approach that goal. To that end, it is useful to outline future research work that can extend upon this work and help produce a more comprehensive model.

1. Collecting more data for the three materials in the study (Inconel 718, Invar 36 and Inconel 625) in a manner that creates a balanced dataset (comparable numbers of clean and defective samples) is the first step needed to potentially get better results.
2. Expanding the material domain to include more nickel alloys like Nimonic superalloys and copper-nickel alloys can enhance the versatility of predictions. An appropriate amount of data, however, will be necessary to ensure good quality predictions. The next step in this direction would be to expand to other alloy classes like iron or titanium alloys.
8. Expanding the defect domain to include more diverse defects is also useful for increasing versatility (in conjunction with point 2). Depending on the various causative factors for these defects, additional sensors may be necessary (as mentioned in point 7 of the conclusions). Future work could 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. One example would be hydrogen porosity, which is primarily linked to contamination of the welding wire/substrate. Such a defect would be better tracked by spectroscopic signals which would detect hydrogen peaks.
3. Investigating incremental learning models akin to the study of Li *et al.* [43] can be very helpful for practical applications since this enables continuous learning and updation of the model. The information obtained from the current analysis can be passed to the model as a starting point in order to help it process the incoming real-time data better. A variant of this, known as class incremental learning, can be used as well since this type of modeling allows for the introduction of new class labels in real-time [107]. Additive manufacturing involves all sorts of unpredictable factors which could cause defects, and thus a model that is open to classifying new types of defects without requiring prior information will be apt.

The application of machine learning in additive manufacturing is a vital step that takes us one step closer towards complete automation of the process. The research work done through the course of this thesis is a contribution towards investigating the predictive capabilities of ML models across different materials, which is something that has not been analyzed in current literature. Further work done along the lines of what has been outlined above can result in a complete system which learns from new prints in real-time and gives a comprehensive view of various defect probabilities.

Thus ends this thesis.

References

[1] X. Chen, F. Kong, Y. Fu, X. Zhao, R. Li, G. Wang, and H. Zhang, "A review on wire-arc additive manufacturing: Typical defects, detection approaches, and multisensor data fusion-based model," *The International Journal of Advanced Manufacturing Technology*, vol. 117, no. 3-4, pp. 707–727, 2021.

[2] T. Zhang and C.-T. Liu, "Design of titanium alloys by Additive Manufacturing: A critical review," *Advanced Powder Materials*, vol. 1, no. 1, p. 100014, 2022.

[3] Y. Li, C. Su, and J. Zhu, "Comprehensive Review of Wire Arc Additive Manufacturing: Hardware System, physical process, monitoring, property characterization, application and future prospects," *Results in Engineering*, vol. 13, p. 100330, 2022.

[4] E. M. Palmero and A. Bollero, "3D and 4D printing of functional and smart composite materials," *Encyclopedia of Materials: Composites*, pp. 402–419, 2021.

[5] D.-G. Ahn, "Directed energy deposition (DED) process: State of the art," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 8, no. 2, pp. 703–742, 2021.

[6] A. Dass and A. Moridi, "State of the art in directed energy deposition: From additive manufacturing to materials design," *Coatings*, vol. 9, no. 7, p. 418, 2019.

[7] L. Nele, G. Mattera, and M. Vozza, "Deep neural networks for defects detection in gas metal arc welding," *Applied Sciences*, vol. 12, no. 7, p. 3615, 2022.

[8] M. Rohe, B. N. Stoll, J. Hildebrand, J. Reimann, and J. P. Bergmann, "Detecting process anomalies in the GMAW process by acoustic sensing with a Convolutional Neural Network (CNN) for Classification," *Journal of Manufacturing and Materials Processing*, vol. 5, no. 4, p. 135, 2021.

[9] Z. Zhang, G. Wen, and S. Chen, "Weld Image Deep Learning-based on-line defects detection using convolutional neural networks for Al Alloy in robotic arc welding," *Journal of Manufacturing Processes*, vol. 45, pp. 208–216, 2019.

[10] P. K. Nalajam and R. Varadarajan, "A hybrid deep learning model for layer-wise melt pool temperature forecasting in wire-arc additive manufacturing process," *IEEE Access*, vol. 9, pp. 100652–100664, 2021.

[11] J. Wu, C. Huang, Z. Li, R. Li, G. Wang, and H. Zhang, “An in situ surface defect detection method based on improved you only look once algorithm for wire and arc additive manufacturing,” *Rapid Prototyping Journal*, 2022.

[12] B. Wu, Z. Pan, D. Ding, D. Cuiuri, H. Li, J. Xu, and J. Norrish, “A review of the wire arc additive manufacturing of metals: Properties, defects and Quality Improvement,” *Journal of Manufacturing Processes*, vol. 35, pp. 127–139, 2018.

[13] T. Hauser, R. T. Reisch, P. P. Breese, B. S. Lutz, M. Pantano, Y. Nalam, K. Bela, T. Kamps, J. Volpp, and A. F. H. Kaplan, “Porosity in wire arc additive manufacturing of aluminium alloys,” *Additive Manufacturing*, vol. 41, p. 101993, 2021.

[14] E. M. Ryan, T. J. Sabin, J. F. Watts, and M. J. Whiting, “The influence of build parameters and wire batch on porosity of wire and arc additive manufactured aluminium alloy 2319,” *Journal of Materials Processing Technology*, vol. 262, pp. 577–584, 2018.

[15] J. Bai, H. L. Ding, J. L. Gu, X. S. Wang, and H. Qiu, “Porosity evolution in additively manufactured aluminium alloy during high temperature exposure,” *IOP Conference Series: Materials Science and Engineering*, vol. 167, p. 012045, 2017.

[16] D. Jafari, T. H. J. Vaneker, and I. Gibson, “Wire and arc additive manufacturing: Opportunities and challenges to control the quality and accuracy of manufactured parts,” *Materials & Design*, vol. 202, p. 109471, 2021.

[17] N. Béraud, A. Chergui, M. Limousin, F. Villeneuve, and F. Vignat, “An indicator of porosity through simulation of melt pool volume in aluminum wire arc additive manufacturing,” *Mechanics & Industry*, vol. 23, p. 1, 2022.

[18] P. N. Anyalebechi, “Hydrogen solubility in liquid and solid pure aluminum—critical review of measurement methodologies and reported values,” *Materials Sciences and Applications*, vol. 13, no. 04, pp. 158–212, 2022.

[19] B. Cong, Z. Qi, B. Qi, H. Sun, G. Zhao, and J. Ding, “A comparative study of additively manufactured thin wall and block structure with Al-6.3%cu alloy using Cold Metal Transfer process,” *Applied Sciences*, vol. 7, no. 3, p. 275, 2017.

[20] A. Ramalho, T. G. Santos, B. Bevans, Z. Smoqi, P. Rao, and J. P. Oliveira, “Effect of contaminations on the acoustic emissions during wire and arc additive manufacturing of 316L Stainless Steel,” *Additive Manufacturing*, vol. 51, p. 102585, 2022.

[21] T. Era and T. Ueyama, “Spatter reduction in GMAW by current waveform control,” *Welding International*, vol. 21, no. 7, pp. 496–501, 2007.

[22] R. Scharf-Wildenhain, A. Haelsig, J. Hensel, K. Wandtke, D. Schroepfer, A. Kromm, and T. Kannengiesser, “Influence of heat control on properties and residual stresses of additive-welded high-strength steel components,” *Metals*, vol. 12, no. 6, p. 951, 2022.

[23] S. Thapliyal, “Challenges associated with the wire arc additive manufacturing (WAAM) of aluminum alloys,” *Materials Research Express*, vol. 6, no. 11, p. 112006, 2019.

[24] T. Klein, L. Reiter, and M. Schnall, "Wire-arc additive manufacturing of Al-Zn5.5-mg-Cu (ML7075): Shifting paradigms of additive manufacture-ability," *Materials Letters*, vol. 313, p. 131841, 2022.

[25] S. Kou, "A criterion for cracking during solidification," *Acta Materialia*, vol. 88, pp. 366–374, 2015.

[26] C. E. Seow, J. Zhang, H. E. Coules, G. Wu, C. Jones, J. Ding, and S. Williams, "Effect of crack-like defects on the fracture behaviour of wire + arc additively manufactured nickel-base alloy 718," *Additive Manufacturing*, vol. 36, p. 101578, 2020.

[27] A. Horgar, H. Fostervoll, B. Nyhus, X. Ren, M. Eriksson, and O. M. Akselsen, "Additive manufacturing using WAAM with AA5183 wire," *Journal of Materials Processing Technology*, vol. 259, pp. 68–74, 2018.

[28] S. Kou, "Solidification and liquation cracking issues in welding," *JOM*, vol. 55, no. 6, pp. 37–42, 2003.

[29] C. Shen, Z. Pan, D. Cuiuri, D. Ding, and H. Li, "Influences of deposition current and interpass temperature to the FE3AL-based iron aluminide fabricated using wire-arc additive manufacturing process," *The International Journal of Advanced Manufacturing Technology*, vol. 88, no. 5-8, pp. 2009–2018, 2016.

[30] Y. Javadi, N. E. Sweeney, E. Mohseni, C. N. MacLeod, D. Lines, M. Vasilev, Z. Qiu, C. Mineo, S. G. Pierce, and A. Gachagan, "Investigating the effect of residual stress on hydrogen cracking in multi-pass robotic welding through process compatible non-destructive testing," *Journal of Manufacturing Processes*, vol. 63, pp. 80–87, 2021.

[31] T. Hauser, R. T. Reisch, T. Kamps, A. F. Kaplan, and J. Volpp, "Acoustic emissions in directed energy deposition processes," *The International Journal of Advanced Manufacturing Technology*, vol. 119, no. 5-6, pp. 3517–3532, 2022.

[32] I. Shareef and C. Martin, "Effect of process parameters on weld spatter in robotic welding," *Procedia Manufacturing*, vol. 48, pp. 358–371, 2020.

[33] S. Bhattacharya, K. Pal, and S. K. Pal, "Multi-sensor based prediction of metal deposition in pulsed gas metal arc welding using various soft computing models," *Applied Soft Computing*, vol. 12, no. 1, pp. 498–505, 2012.

[34] W. Qiu, "Kurtosis: A new tool for noise analysis," *Acoustics Today*, vol. 16, no. 4, p. 39, 2020. doi:10.1121/at.2020.16.4.39

[35] S. C. Alfaro and E. H. Cayo, "Sensoring fusion data from the optic and acoustic emissions of electric arcs in the GMAW-s process for welding quality assessment," *Sensors*, vol. 12, no. 6, pp. 6953–6966, 2012.

[36] A. Vandone, S. Baraldo, and A. Valente, "Multisensor data fusion for Additive Manufacturing Process Control," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3279–3284, 2018.

[37] F. Xu, V. Dhokia, P. Colegrove, A. McAndrew, S. Williams, A. Henstridge, and S. T. Newman, "Realisation of a multi-sensor framework for process monitoring of the wire arc additive

manufacturing in producing ti-6al-4v parts," *International Journal of Computer Integrated Manufacturing*, vol. 31, no. 8, pp. 785–798, 2018.

[38] R.-J. Hsieh, J. Chou, and C.-H. Ho, "Unsupervised online anomaly detection on multivariate sensing time series data for Smart Manufacturing," *2019 IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA)*, 2019.

[39] R. Reisch, T. Hauser, B. Lutz, M. Pantano, T. Kamps, and A. Knoll, "Distance-based multivariate anomaly detection in wire arc additive manufacturing," *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2020.

[40] R. T. Reisch, L. Janisch, J. Tresselt, T. Kamps, and A. Knoll, "Prescriptive Analytics - A Smart Manufacturing System for First-Time-Right Printing in Wire Arc Additive Manufacturing using a Digital Twin ," *Prescriptive Analytics - A Smart Manufacturing System for First-Time-Right Printing in Wire Arc Additive Manufacturing using a Digital Twin*, 2022. [Online]. Available: https://mediatum.ub.tum.de/doc/1662103/1662103.pdf. [Accessed: 25-Feb-2023].

[41] Y. Cui, Y. Shi, T. Zhu, and S. Cui, "Welding penetration recognition based on arc sound and electrical signals in K-Tig Welding," *Measurement*, vol. 163, p. 107966, 2020.

[42] A. M. Pringle, S. Oberloier, A. L. Petsiuk, P. G. Sanders, and J. M. Pearce, "Open source arc analyzer: Multi-sensor monitoring of wire arc additive manufacturing," *HardwareX*, vol. 8, 2020.

[43] Y. Li, J. Polden, Z. Pan, J. Cui, C. Xia, F. He, H. Mu, H. Li, and L. Wang, "A defect detection system for wire arc additive manufacturing using incremental learning," *Journal of Industrial Information Integration*, vol. 27, p. 100291, 2022.

[44] X. Xiao, C. Waddell, C. Hamilton, and H. Xiao, "Quality Prediction and control in wire arc additive manufacturing via Novel Machine Learning Framework," *Micromachines*, vol. 13, no. 1, p. 137, 2022.

[45] Y. K. Kwak, T. Lehmann, M. Tavakoli, and A. Qureshi, *Sensor-Based In-situ Process Control of Robotic Wire Arc Additive Manufacturing Integrated with Reinforcement Learning*, 2021. [Online]. Available: http://www.ece.ualberta.ca/~tbs/pmwiki/pdf/Kwak-HI-AM-2021.pdf. [Accessed: 25-Feb-2023].

[46] J. Qin, Y. Wang, J. Ding, and S. Williams, "Optimal droplet transfer mode maintenance for wire + arc additive manufacturing (WAAM) based on Deep Learning," *Journal of Intelligent Manufacturing*, vol. 33, no. 7, pp. 2179–2191, 2022.

[47] J. Qin, F. Hu, Y. Liu, P. Witherell, C. C. L. Wang, D. W. Rosen, T. W. Simpson, Y. Lu, and Q. Tang, "Research and application of machine learning for Additive Manufacturing," *Additive Manufacturing*, vol. 52, p. 102691, 2022.

[48] H. Ko, P. Witherell, Y. Lu, S. Kim, and D. W. Rosen, "Machine learning and knowledge graph based design rule construction for additive manufacturing," *Additive Manufacturing*, vol. 37, p. 101620, 2021.

[49] F. He, L. Yuan, H. Mu, M. Ros, D. Ding, Z. Pan, and H. Li, "Research and application of artificial intelligence techniques for Wire Arc Additive Manufacturing: A state-of-the-art review," *Robotics and Computer-Integrated Manufacturing*, vol. 82, p. 102525, 2023.

[50] D. Ding, F. He, L. Yuan, Z. Pan, L. Wang, and M. Ros, "The first step towards intelligent wire arc additive manufacturing: An automatic bead modelling system using machine learning through Industrial Information Integration," *Journal of Industrial Information Integration*, vol. 23, p. 100218, 2021.

[51] Kobe Steel, Ltd., *Weld Imperfections and Preventative Measures*. Kita-Shinagawa, Tokyo: Kobe Steel, Ltd., 2015.

[52] An Elementary Introduction to Kalman Filtering, Y Pei, S Biswas, DS Fussell, K Pingali, 27/06/2019

[53] "The Arc Welding Robot System TAWERS," TAWERS Series, https://industrial.panasonic.com/content/data/WS/PDF/201801_TAWERS_E.pdf (accessed 2023).

[54] D. O'Shaughnessy, in *Speech communication: Human and machine*, Piscataway, NJ: Addison-Wesley, 1987, p. 150

[55] INCONEL alloy 718, https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-718.pdf (accessed 2023).

[56] W. S. James, S. Ganguly, and G. Pardal, "High temperature performance of wire-arc additive manufactured Inconel 718," *Scientific Reports*, vol. 13, no. 1, 2023. doi:10.1038/s41598-023-29026-9

[57] INVAR 36, https://salomons-metalen.nl/datasheets/Invar_36.pdf (accessed 2023). UNS Number K93601/K93603

[58] "The Nobel Prize in Physics 1920," The Nobel Prize, https://www.nobelprize.org/prizes/physics/1920/summary/ (accessed 2023).

[59] H. L. Eiselstein and D. J. Tillack, "The Invention and Definition of Alloy 625," *Superalloys 718, 625 and Various Derivatives (1991)*, 1991. doi:10.7449/1991/superalloys_1991_1_14

[60] INCONEL alloy 625, https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-625.pdf (accessed 2023).

[61] R. A. Patterson, R. B. Nemec, and R. D. Reiswig, "Discontinuities Formed in Inconel GTA Welds," *Welding Research Supplement*, Jan. 1987.

[62] N. Li, M. Shepperd, and Y. Guo, "A systematic review of unsupervised learning techniques for software defect prediction," *Information and Software Technology*, vol. 122, p. 106287, 2020. doi:10.1016/j.infsof.2020.106287

[63] D. Marutho, S. Hendra Handaka, E. Wijaya, and Muljono, "The determination of cluster number at K-mean using elbow method and purity evaluation on Headline news," *2018 International Seminar on Application for Technology of Information and Communication*, 2018. doi:10.1109/isemantic.2018.8549751

[64] E. Umargono, J. E. Suseno, and V. G. S. K., "K-means clustering optimization using the elbow method and early centroid determination based-on mean and Median," *Proceedings of the*

International Conferences on Information System and Technology, 2019. doi:10.5220/0009908402340240

[65] H. Humaira and R. Rasyidah, "Determining the appropriate cluster number using elbow method for K-means algorithm," *Proceedings of the Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia*, 2020. doi:10.4108/eai.24-1-2018.2292388

[66] C. Shi *et al.*, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, 2021. doi:10.1186/s13638-021-01910-w

[67] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference for Learning Representations, San Diego, 2015*, Jan. 2017. doi:https://doi.org/10.48550/arXiv.1412.6980

[68] C. Desai, "Comparative Analysis of Optimizers in Deep Neural Networks," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 10, Oct. 2020.

[69] Y. Tian, Y. Zhang, and H. Zhang, "Recent advances in stochastic gradient descent in deep learning," *Mathematics*, vol. 11, no. 3, p. 682, 2023. doi:10.3390/math11030682

[70] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv*, Jun. 2017. doi:https://doi.org/10.48550/arXiv.1609.04747

[71] Z. Lu, W. Xia, S. Arora, and E. Hazan, "Adaptive Gradient Methods with Local Guarantees," *arXiv*, Jan. 2023. doi:https://doi.org/10.48550/arXiv.2203.01400

[72] W. Hastomo, A. S. Bayangkari Karno, N. Kalbuana, A. Meiriki, and Sutarno, "Characteristic parameters of epoch deep learning to predict COVID-19 data in Indonesia," *Journal of Physics: Conference Series*, vol. 1933, no. 1, p. 012050, 2021. doi:10.1088/1742-6596/1933/1/012050

[73] S. Afaq and S. Rao, "Significance Of Epochs On Training A Neural Network," *International Journal of Scientific & Technology Research*, vol. 9, no. 6, Jun. 2020.

[74] P. M. Radiuk, "Impact of training set batch size on the performance of convolutional neural networks for diverse datasets," *Information Technology and Management Science*, vol. 20, no. 1, 2017. doi:10.1515/itms-2017-0003

[75] M. Bae, "Optimizing the hyper-parameters of Multi-layer Perceptron with greedy search," *American Journal of Computer Science and Technology*, vol. 4, no. 4, p. 90, 2021. doi:10.11648/j.ajcst.20210404.11

[76] S. L. Smith, P-J. Kindermans, C. Ying, and Q. V. Le, "Don't Decay the Learning Rate, Increase the Batch Size," *International Conference on Learning Representations 2018*, Feb. 2018. doi:https://doi.org/10.48550/arXiv.1711.00489

[77] D. N. Luu, W. Zhou, and S. M. Nai, "Mitigation of liquation cracking in selective laser melted Inconel 718 through optimization of layer thickness and laser energy density," *Journal of Materials Processing Technology*, vol. 299, p. 117374, 2022. doi:10.1016/j.jmatprotec.2021.117374

[78] X. Ye, X. Hua, M. Wang, and S. Lou, "Controlling hot cracking in Ni-based Inconel-718 superalloy cast sheets during tungsten inert gas welding," *Journal of Materials Processing Technology*, vol. 222, pp. 381–390, 2015. doi:10.1016/j.jmatprotec.2015.03.031

[79] T. Li, Z. Wang, S. Hu, Z. Yang, and Y. Wang, "Hot cracking during the fabrication of Inconel 625/stainless steel 308 L functionally graded material by dual-wire arc additive manufacturing," *Journal of Manufacturing Processes*, vol. 82, pp. 461–473, 2022. doi:10.1016/j.jmapro.2022.08.018

[80] S. J. Hales, C. S. Domack, and K. M. Taminger, "Electron Beam Freeform Fabrication of Dissimilar Materials: Cracking in Inconel® 625 Deposited on GRCop-84, " NASA Scientific and Technical Information Program, Hampton, Virginia, tech., 2020

[81] *2017 ASME Boiler & Pressure Vessel Code Section VIII Division 1*. New York, NY: American Society of Mechanical Engineers, 2017.

[82] N. Shamsaei *et al.*, *Structural Integrity of Additive Manufactured Parts*. West Conshohocken, PA: ASTM International, 2020.

[83] N. Elmrabit, F. Zhou, F. Li, and H. Zhou, "Evaluation of Machine Learning Algorithms for Anomaly Detection," *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 2020. doi:10.1109/cybersecurity49315.2020.9138871

[84] C. Meng, L. Trinh, N. Xu, J. Enouen, and Y. Liu, "Interpretability and fairness evaluation of deep learning models on Mimic-IV Dataset," *Scientific Reports*, vol. 12, no. 1, 2022. doi:10.1038/s41598-022-11012-2

[85] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics*, vol. 10, no. 5, p. 593, 2021. doi:10.3390/electronics10050593

[86] T. Antoniou and M. Mamdani, "Evaluation of Machine Learning Solutions in medicine," *Canadian Medical Association Journal*, vol. 193, no. 36, 2021. doi:10.1503/cmaj.210036

[87] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and Information Systems*, vol. 51, no. 2, pp. 339–367, 2016. doi:10.1007/s10115-016-0987-z

[88] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, 2020. doi:10.1186/s12864-019-6413-7

[89] G. N. Lee and H. Fujita, "K-means clustering for classifying unlabelled MRI data," *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)*, 2007. doi:10.1109/dicta.2007.4426781

[90] M. Rithani, R. P. Kumar, and S. Doss, "A review on Big Data based on Deep Neural Network approaches," *Artificial Intelligence Review*, 2023. doi:10.1007/s10462-023-10512-5

[91] A-N. Spiess and N. Neumeyer, "An evaluation of R^2 as an inadequate measure for nonlinear models in Pharmacological and Biochemical Research: A Monte Carlo Approach," *BMC Pharmacology*, vol. 10, no. 1, 2010. doi:10.1186/1471-2210-10-6

[92] A. M. Zoubir and D. R. Iskander, "Bootstrap Methods and Applications," *IEEE - Signal Processing Magazine*, vol. 24, no. 4, pp. 10–19, Jul. 2007. doi:10.1109/MSP.2007.4286560

[93] A. R. Nandeshwar, "Models for calculating confidence intervals for Neural Networks," thesis, West Virginia University, Morgantown, 2006

[94] B. Bengfort and R. Bilbro, "Yellowbrick." Aug. 9, 2017

[95] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in Convolutional Neural Networks," *Neural Networks*, vol. 106, pp. 249–259, 2018. doi:10.1016/j.neunet.2018.07.011

[96] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in Big Data," *Journal of Big Data*, vol. 5, no. 1, 2018. doi:10.1186/s40537-018-0151-6

[97] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, 2019. doi:10.1186/s40537-019-0192-5

[98] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi:10.1023/a:1010933404324

[99] L. E. Raileanu and K. Stoffel, "Theoretical comparison between the Gini index and Information Gain Criteria," *Annals of Mathematics and Artificial Intelligence*, vol. 41, no. 1, pp. 77–93, 2004. doi:10.1023/b:amai.0000018580.96245.c6

[100] A. Botchkarev, "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology," *arXiv*, Sep. 2018. doi:https://doi.org/10.48550/arXiv.1809.03006

[101] S. M. Lundberg and S-I. Lee, "A Unified Approach to Interpreting Model Predictions," *31st Conference on Neural Information Processing Systems 2017*, pp. 4768–4777, Dec. 2017.

[102] S. S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2009.

[103] L. Ward *et al.*, "A machine learning approach for engineering bulk metallic glass alloys," *Acta Materialia*, vol. 159, pp. 102–111, 2018. doi:10.1016/j.actamat.2018.08.002

[104] S. Feng, H. Zhou, and H. Dong, "Using deep neural network with small dataset to predict material defects," *Materials & Design*, vol. 162, pp. 300–310, 2019. doi:10.1016/j.matdes.2018.11.060

[105] T. Artaza, T. Bhujangrao, A. Suárez, F. Veiga, and A. Lamikiz, "Influence of heat input on the formation of Laves phases and hot cracking in plasma arc welding (PAW) additive manufacturing of Inconel 718," *Metals*, vol. 10, no. 6, p. 771, 2020. doi:10.3390/met10060771

[106] C. Radhakrishna and K. P. Rao, "The formation and control of Laves phase in superalloy 718 welds," *Journal of Materials Science*, vol. 32, no. 8, pp. 1977–1984, 1997. doi:10.1023/a:1018541915113

[107] D-W. Zhou *et al.*, “Deep Class-Incremental Learning: A Survey,” *arXiv*, Feb. 2023. doi:https://doi.org/10.48550/arXiv.2302.03648

[108] Z. Lin *et al.*, “Deposition of Stellite 6 alloy on steel substrates using wire and arc additive manufacturing,” *The International Journal of Advanced Manufacturing Technology*, vol. 111, no. 1–2, pp. 411–426, 2020. doi:10.1007/s00170-020-06116-w

[109] M. Consonni, D. Howse, C. F. Wee, and C. Schneider, “Production of joints welded with realistic defects,” *Welding International*, vol. 28, no. 7, pp. 535–546, 2013. doi:10.1080/09507116.2012.753263

[110] Steel 1.0570 / S355J2G3, https://xometry.eu/wp-content/uploads/2021/03/Steel-1.0570.pdf (accessed 2023).

Appendix A

This appendix gives the experimental parameters used for the experiments conducted during the course of this thesis in Tables A1, A2 and A3. The tables only include details for the beads used in the analysis. Other tests were conducted which were not considered for the analysis due to lack of audio data (prior to installation of the microphone), which are not shown here.

Table A1: Experimental parameters for Inconel 718 beads

Test Number	*Welding Mode*	*Test Type*	*Bead Type*	*Voltage (V)*	*Current (A)*	*Wire Feed Speed (m/min)*	*Travel Speed (m/min)*
1	Superactive	Single bead	-	24	-	4.80	0.66
2	Superactive	Single bead	-	22.5	-	3.25	0.50
3	Superactive	Single bead	-	24	-	4.80	0.66
4	Superactive	Single bead	-	24	-	3.30	0.50
5	Superactive	Single bead	-	23	-	3.10	0.50
6	Superactive	Single bead	-	22.5	-	3.25	0.50
7	Superactive	Single bead	-	22.5	-	3.25	0.50
8	Superactive	Single bead	-	22.5	-	3.25	0.50
9	Superactive	Single bead	-	22.5	-	3.25	0.50
10	Superactive	Single bead	-	24	-	4.80	0.66
11	Superactive	Single bead	-	24	-	4.50	0.66
12	Superactive	Single bead	-	24	-	4.20	0.66
13	Superactive	Single bead	-	24	-	4.20	0.66

14	Superactive	Single bead	-	24	-	4.20	0.66
15	Superactive	Single bead	-	24	-	4.20	0.66
16	Superactive	Bead wall	Preheat	22.6	-	4.20	0.66
			Wall	22.5	-	3.25	0.50
17	Superactive	Bead wall	Preheat	22.6	-	4.20	0.66
			Wall	24	-	4.80	0.66
18	Superactive	Bead wall	Preheat	22.6	-	4.20	0.66
			Wall	22.5	-	3.25	0.50
19	Superactive	Bead wall	Preheat	22.6	-	4.20	0.66
			Wall	22.5	-	3.25	0.50
20	Superactive	Bead Block	-	22.5	-	2.80	0.60
21	Superactive	Bead Block	-	22.5	-	2.80	0.60

Table A2: Experimental parameters for Invar 36 beads

Test Number	*Welding Mode*	*Test Type*	*Voltage (V)*	*Current (A)*	*Wire Feed Speed (m/min)*	*Travel Speed (m/min)*
1	Superactive	Bead wall	20.8	-	8.00	0.15
2	Superactive	Bead wall	15.6	-	8.00	0.25
3	Superactive	Bead wall	15.6	-	8.00	0.25
4	Superactive	Bead wall	15.6	-	8.00	0.25
5	Superactive	Bead wall	15.6	-	8.00	0.25
6	Superactive	Single bead	10.8	-	3.00	0.25
7	Superactive	Single bead	11.2	-	3.50	0.25
8	Superactive	Single bead	14.8	-	3.00	0.25
9	Superactive	Bead wall	14.8	-	3.00	0.25
10	Superactive	Bead wall	14.8	-	3.00	0.25
11	Superactive	Single bead	21.6	-	5.00	0.75
12	Superactive	Single bead	18.6	-	3.50	0.65
13	Superactive	Single bead	18.6	-	3.00	0.55
14	Superactive	Single bead	18.6	-	2.75	0.50
15	Superactive	Single bead	20.5	-	2.75	0.50
16	Superactive	Single bead	21.6	-	5.50	0.50

17	Superactive	Single bead	18.6	-	2.75	0.48
18	Superactive	Bead wall	19.0	-	2.50	0.52
19	Superactive	Bead wall	19.0	-	2.50	0.52
20	Superactive	Bead wall	19.0	-	2.50	0.52
21	Superactive	Bead wall	19.0	-	2.50	0.52

Table A3: Experimental parameters for Inconel 625 beads

Test Number	*Welding Mode*	*Test Type*	*Voltage (V)*	*Current (A)*	*Wire Feed Speed (m/min)*	*Travel Speed (m/min)*
1	Superactive	Single bead	27.5	-	6.25	0.75
2	Superactive	Single bead	22.9	-	5.25	0.70
3	Superactive	Single bead	21.8	-	4.80	0.66
4	Superactive	Bead layer	24.0	-	4.80	0.66
5	Superactive	Bead wall	24.0	-	4.80	0.66

Appendix B

This appendix contains a few loss curves from the ANNs for each material presented in Figure B1, Figure B2 and Figure B3. These curves can be examined to determine the extent of overfitting/underfitting. Both Inconel 718 and Inconel 625 show fairly uneventful curves with an expected decrease as the epochs pass. Invar 36, however, shows a more tumultuous plot with a validation loss curve that shows many localized peaks. A general increasing trend would suggest overfitting, but that does not seem to be the case here. Looking at the shape of the curve, however, it is clear that fitting the behaviour of Invar 36 is not as straightforward as that of Inconel-class alloys. It would be interesting to take more data in a class-balanced fashion and try the same fitting to see if the validation loss improves.

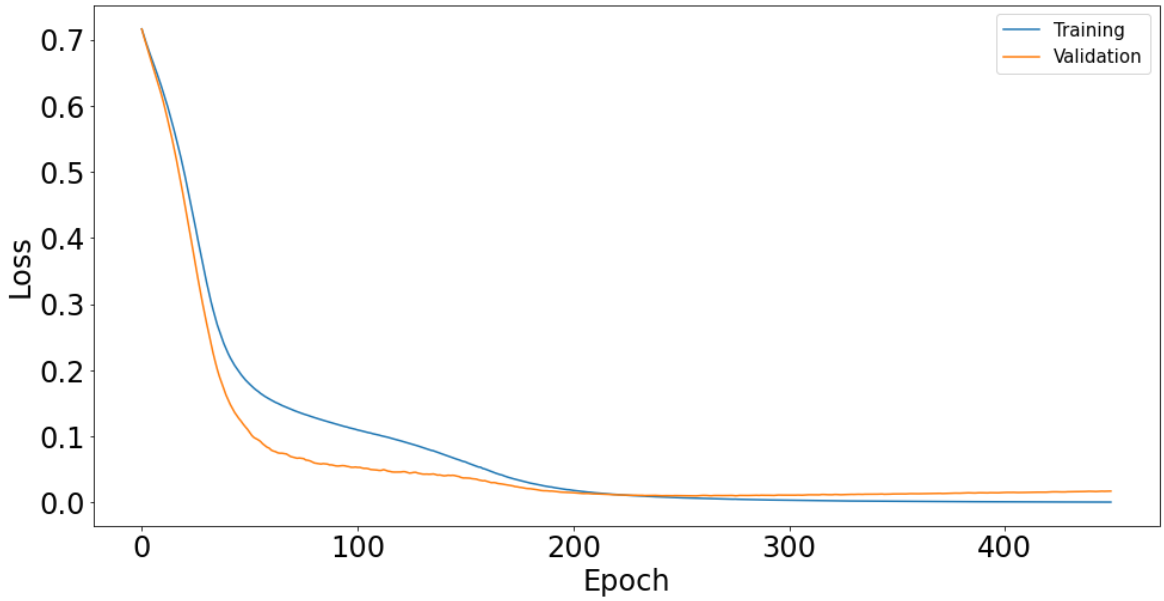


Figure B1: Training loss and validation loss curves of ANN trained on Inconel 718 data

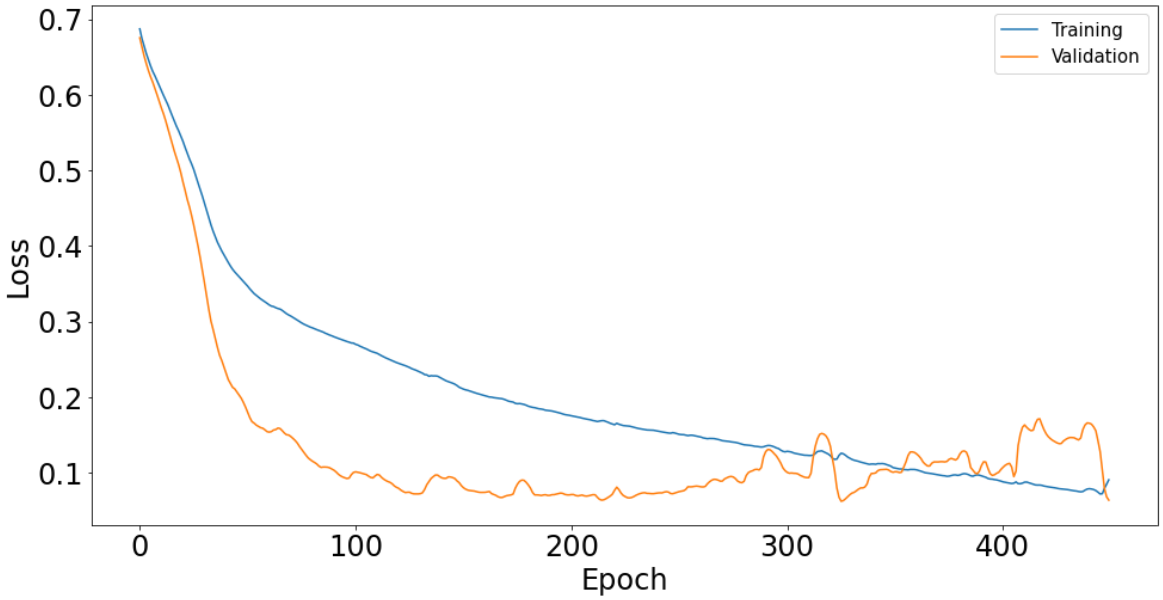


Figure B2: Training loss and validation loss curves of ANN trained on Invar 36

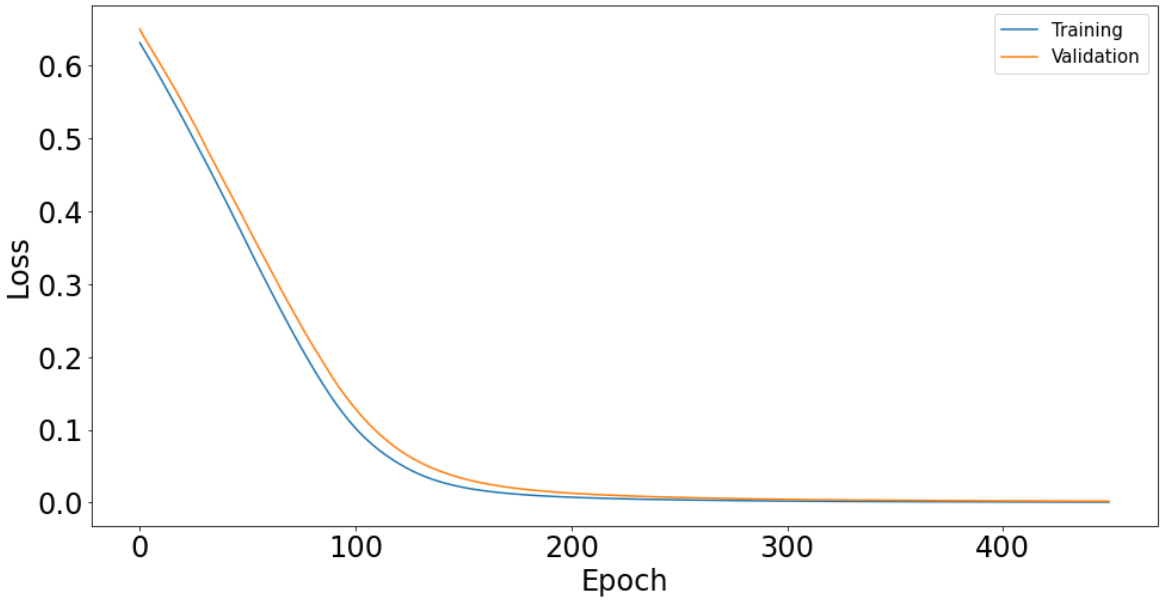


Figure B3: Training loss and validation loss curves of ANN trained on Inconel 625

Appendix C

This appendix compiles the average values of scaled input parameters for clean and defective beads of each material along with the 95% confidence intervals in Tables C1, C2 and C3. The intervals are presented as percentages of the mean in order to better demonstrate the size of the interval. Since Inconel 625 did not have any defective beads, the corresponding table only includes data of clean beads.

Table C1: Scaled input parameters of clean and defective beads of Inconel 718

Parameters	V_1	V_2	I_1	I_2	A_1	A_2
Mean (clean)	4.3124	10.7225	5.4616	9.4607	199.0250	199.0509
Mean (defective)	4.1373	8.1114	4.2137	5.0265	170.2186	217.9748
95% CI (clean)	4.1367%	6.0539%	6.7278%	9.6558%	21.6573%	12.6842%
95% CI (defective)	1.1528%	3.6032%	2.5600%	3.2314%	19.9531%	6.0714%

Table C2: Scaled input parameters of clean and defective beads of Invar 36

Parameters	V_1	V_2	I_1	I_2	A_1	A_2
Mean (clean)	3.0994	8.0356	3.7175	4.7308	271.8471	131.8315
Mean (defective)	4.2451	10.0829	4.4732	5.6741	218.3055	97.1015
95% CI (clean)	4.6000%	5.3616%	5.7599%	5.4065%	22.9620%	9.1833%
95% CI (defective)	11.7467%	7.8948%	9.3869%	7.6141%	47.1079%	29.7043%

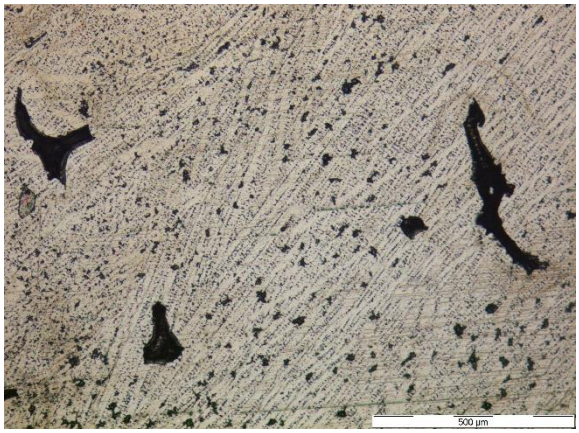
Table C3: Scaled input parameters of clean beads of Inconel 625

Parameters	V_1	V_2	I_1	I_2	A_1	A_2
Mean (clean)	4.4015	9.3678	5.0663	6.7336	100.2629	234.1729
95% CI (clean)	2.3203%	2.4407%	1.0389%	0.8416%	25.9189%	12.9449%

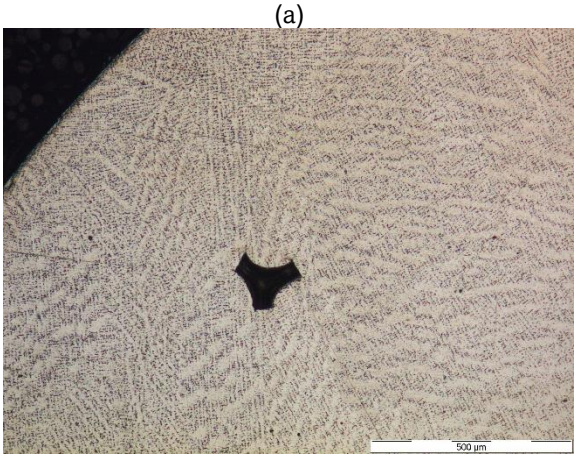
Appendix D

This appendix includes micrographs taken during the metallographic analysis of samples printed at RAMLAB. Groups of defective samples of each material have been compiled and presented to give an idea of the kind of defects observed during the course of this research work. The first collection of micrographs is of Inconel 718, shown in Figure D1. The primary defects observed are lack of fusion and hot cracking. The next collection of micrographs is of Invar 36, shown in Figure D2. The primary defect observed is porosity, but only in a few samples. In addition, some sample pictures of Invar 36 have also been presented to depict the levels of bead overflow encountered in the course of experiments. These are shown in Figure D3. Finally, the large samples made for all three materials are shown in Figure D4. The large sample of Invar 36 was used for the measurements of tensile properties in Table 8.

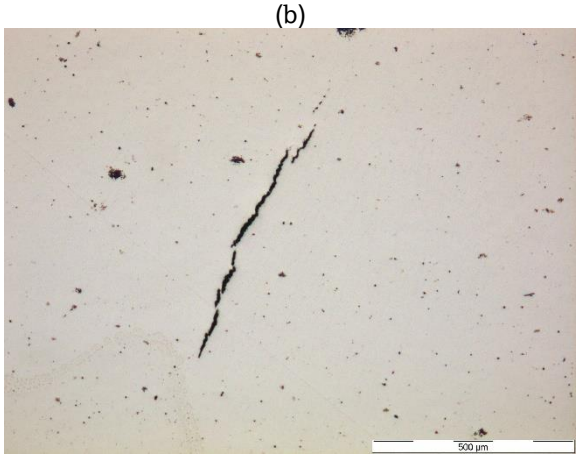
(a)



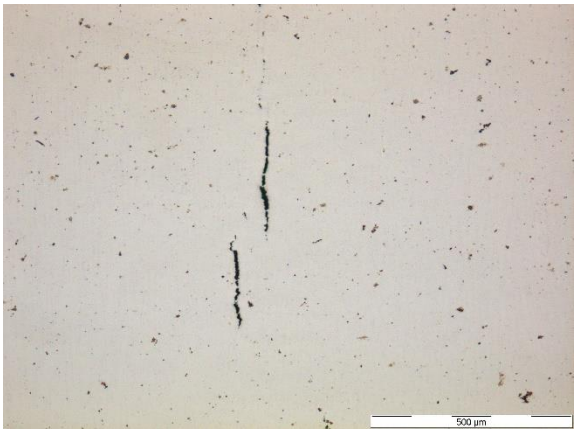
(b)



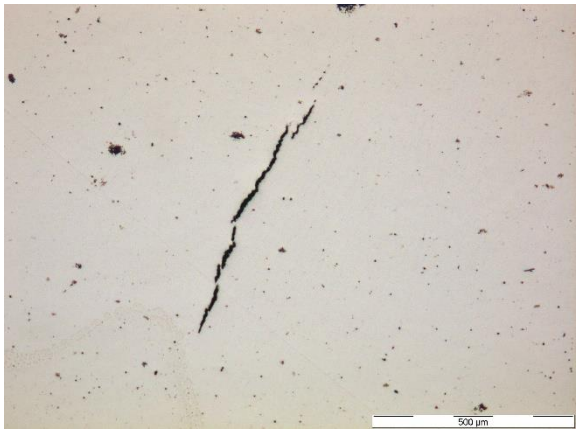
(c)



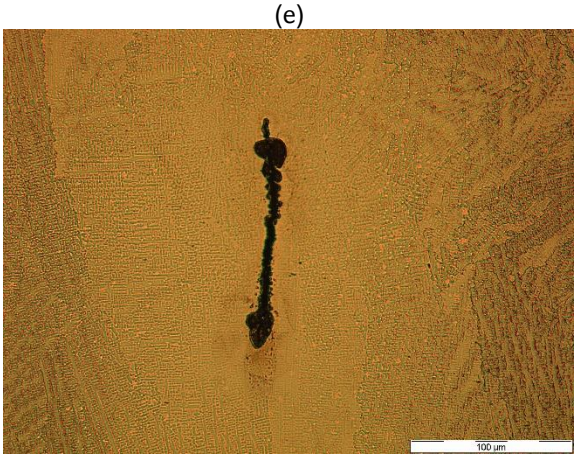
(d)



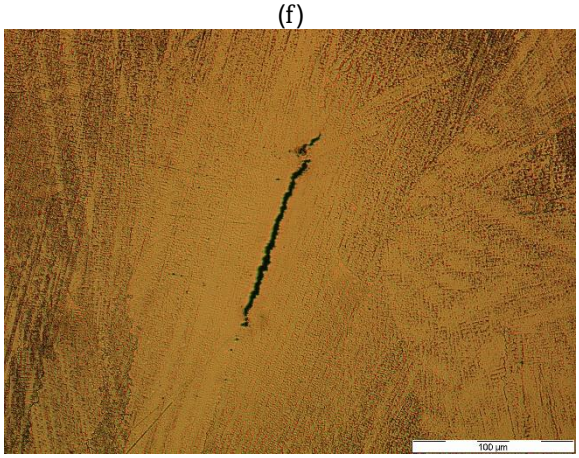
(e)



(f)



(g)

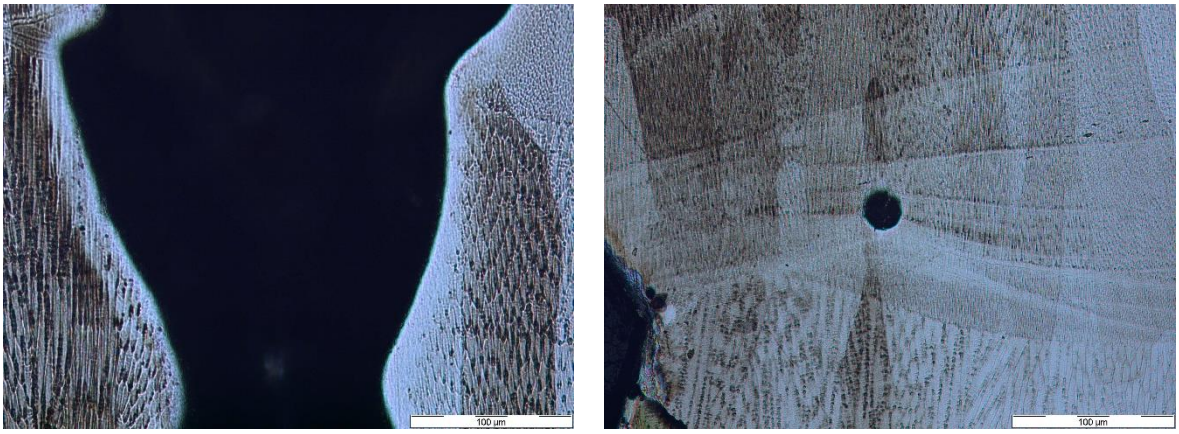


(h)

Figure D1a-h: Micrographs of defects taken of various Inconel 718 samples – (a)-(c) lack-of-fusion defects, (d)-(f) solidification cracks, (g)-(h) liquation cracks

(a)

(b)



(c) (d)

Figure D2a-d: Micrographs of defects taken of various Invar 36 samples – (a) porosity near substrate, (b) porosity in bulk, (c) large porosity in bulk, (d) porosity in bulk

(a) (b) (c)

Figure D3a-c: Sample pictures of Invar 36 showing bead overflow

(a) (b) (c)

Figure D4a-c: Large samples of (a) Invar 36, (b) Inconel 718, (c) Inconel 625