Article

Prediction of Bead Geometry Using a Two-Stage SVM–ANN Algorithm for Automated Tungsten Inert Gas (TIG) Welds

Rohit Kshirsagar [1,*], Steve Jones [2], Jonathan Lawrence [1] and Jim Tabor [3]

[1] Institute for Advanced Manufacturing and Engineering, Coventry University, Coventry CV6 5LZ, UK; ac5588@coventry.ac.uk

[2] Nuclear Advanced Manufacturing Research Centre, University of Sheffield, Rotherham S60 5WG, UK; steven.jones@sheffield.ac.uk

[3] sigma Maths and Stats Support Centre, Coventry University, Coventry CV1 5FB, UK; mtx041@coventry.ac.uk

* Correspondence: kshirsar@uni.coventry.ac.uk

check for
updates

Abstract: Prediction of weld bead geometry is critical for any welding process, since several mechanical properties of the weldment depend on this. Researchers have used artificial neural networks (ANNs) to predict the bead geometry based on the input parameters for a welding process; however, the number of hidden layers used in these ANNs are limited to one due to the small amount of data usually available through experiments. This results in a reduction in the accuracy of prediction. Such ANNs are also incapable of capturing sudden changes in the input–output trends; for example, where a wide range of heat inputs results in flat crown (zero crown height), but any further reduction in the current sharply increases the crown height. In this study, it was found that above mentioned issues can be resolved on using a two-stage algorithm consisting of support vector machine (SVM) and an ANN. The two-stage SVM–ANN algorithm significantly improved the accuracy of prediction and could be used as a replacement for the multiple hidden layer ANN, without requiring additional data for training. The improvement in prediction was evident near regions of sudden changes in the input–output correlation and can lead to a better prediction of mechanical properties.

Keywords: bead geometry prediction; support vector machines; artificial neural networks; data classification

1. Introduction

Controlling the weld bead geometry is critical for any welding process since it can greatly influence several mechanical properties, such as the yield strength (σ_y), tensile strength (TS), elongation before failure (ε), stress concentration factor (K_t), and consequently the fatigue life of the weldment. Prediction and optimization of the bead geometry is becoming increasingly important in many industrial applications, as it can save a significant amount of time and material involved in unnecessary trials. Geometry related imperfections on the weld bead include crown concavity, excess crown, incomplete penetration, excess penetration, and incorrect weld toe [1]. Many industries have developed their own specifications to limit the size of these imperfections depending on the application, beyond which they are classified as defects. Some international standards, such as ISO 5817 [1], also provide guidelines to classify such imperfections as defects. ISO 5817 states three different quality levels—B, C, and D—to allow a wide range of applications to be considered under it. Level B corresponds to the highest quality requirement on finished welds. In the present research, since no specific application of the welds is considered, the geometrical features are assessed against the ISO 5817 Level B standard, which is explained in subsequent sections.

Conventionally, computational models developed to predict bead geometry used techniques such as regression analysis and response surface methodology [2]. Schneider et al. [3] used the Taguchi method for optimizing the parameters of hybrid welds made using the tungsten inert gas (TIG) and metal inert gas (MIG) processes through 27 different experiments. Some other researchers who have used regression models to predict weld properties include [4], who used linear regression for modelling a submerged arc welding (SAW) process, [5], who used regression analysis and response surface methodology for optimising the parameters of the SAW process, and [6], who modelled the MIG process using second-order regression analysis. However, in recent years, artificial intelligence and machine learning has led to the development of advanced models using artificial neural networks (ANNs) that have a better capability to approximate non-linear processes. These algorithms typically use a large amount of data to learn the input–output correlation. Several researchers have attempted the prediction of weld features using such algorithms. Dutta and Pratihar [7] used regression analysis as well as ANN to predict the crown width (CW), crown height (CH), back width (BW), and back height (BH) of the welds obtained through the TIG welding process, based on the welding speed (S), wire feed rate (R), cleaning, root gap, and current. They found that the ANN-based approaches give better results than the regression-based approaches. In a similar investigation, [8] also concluded that although second-order regression analysis as well as ANNs can predict the bead geometry of the welds within a considerable accuracy, ANNs show a better performance. ANNs can not only predict the bead geometry, but also the mechanical properties of the welds. Okuyuku et al. [9] used an ANN to predict the TS, σ_y, ε, and hardness of friction stir welded joints on aluminium plates, based on the weld speed and the tool rotation speed. They found an excellent agreement between the experimentally obtained and predicted data. Similarly, Vitek et al. [10] developed an ANN model (Oak-Ridge Ferrite Number, ORFN) to predict the amount of retained δ-ferrite in austenitic stainless-steel welds, depending on the base material chemistry and the cooling rate of the weld pool. This model outperformed all the other internationally accepted models that had been developed before then. It has been reported that the root mean square error (E_{rms}) in the prediction of ferrite number reduced from 5.84 for WRC-1992 model to 3.88 for the ORFN model. Several other researchers have used ANNs to predict weld features, including [11], who used an ANN for real time arc-welding defect detection and classification; [12], who used ANN along with genetic algorithm to optimise laser welding process parameters for super austenitic stainless steels; and [13], who used an ANN to predict the joint strength of pulsed MIG welding, based on the arc signal. Additionally, [14] compared the performance of an ANN to that of regression analysis to model the heat source parameters in a TIG welding process. They reported that the ANN model outperforms the regression model. Convolutional neural networks (CNNs) were used by [15] to predict the penetration of the weld seam using time-frequency imaging of the arc sound. They achieved 98.2% accuracy in the predictions using the CNN, which outperformed most of the previously developed models.

Apart from ANNs, other machine learning algorithms, such as support vector machines (SVMs), have also been used by researchers to predict the weld features. Hang et al. [16] used an SVM to predict the backside weld bead shape for online weld quality control. They reported that such algorithms can be effectively used to estimate the weld quality. Rong et al. [17] attempted to use machine learning techniques to predict the bead penetration from the weld pool surface. This could be done using multi-layer perceptron (MLP) ANN, however, due to the limited amount of data available, they proposed using support vector regression (SVR) instead. The SVR model showed an improved accuracy of prediction over the ANN in their study.

From all the above mentioned research, it is evident that machine learning models can be effectively used to predict many of the weld features. However, the parameter ranges considered during these studies, especially the ones that involve prediction of bead geometry, are very limited, owing to the nature of these models. Usually, the number of data points available to develop the computational models in the case of welding are limited to just over one hundred, due to the large amount of time and effort required to prepare the metallographic samples for obtaining the data. This limits the number

of hidden layers and the number of neurons in these hidden layers that can be used to predict the weld features [18]. Most of the models mentioned above make use of single hidden layer ANNs. Such single hidden layer ANNs have a one-to-one correspondence from input to output, meaning that any change in an influential input parameter will be reflected in the output of the network. For example, it is well known that the welding current significantly influences the geometrical features of the weld bead in the TIG welding process. Its effect is that the CH of the bead decreases with an increase in the welding current if all the other parameters are kept constant. Beyond a certain cut-off value, the CH becomes practically zero for a wide range of current values before it starts becoming concave. It is difficult to train a single hidden layer ANN to predict a zero CH for a wide range of input currents, while predicting it positive for the other values. Similarly, the back width (BW) and back angle (BA) are zero for all the weld profiles that have incomplete penetration, and sharply become positive beyond a certain heat input to the weld. Such sharp changes in the effect of input parameters on the outputs of the process lead to a high E_{rms} when a single hidden layer ANN is used. This problem can be addressed by training ANNs with multiple hidden layers, but as mentioned previously, such networks require a larger amount of data due to the non-linearity in the input–output correlation. Another consideration before the application of an ANN to any process is the number of outputs that are required to be predicted. In the case of TIG welds, up to seven outputs, which include the CH, CW, crown angle (CA), weld penetration (PEN), BW, BA, and weld cross-sectional area (WA), may be required to define the bead geometry. With the amount of data available, obtaining this large number of outputs from a smaller number of inputs makes training of the ANNs difficult. Adding a greater number of neurons in the hidden layer may overcome this issue to a certain extent, but can also lead to overfitting on some of the features.

Both these issues can be addressed by initially classifying the input parameters into those that lead to fusion profiles having a zero CH but full PEN (Class 0), positive CH and full PEN (Class 1), or positive CH but incomplete PEN (Class 2), using support vector machines (SVMs). For classification, Class 0 inputs will only require CW, PEN, BW, BA, and WA to be predicted, since CH and CA are always zero for this class. Similarly, Class 2 inputs will have CH, CW, CA, PEN, and WA as the outputs of the welding process, whereas all seven outputs are required to be predicted for Class 1 parameters. Using an SVM to initially classify the data can significantly improve the overall accuracy of prediction and increase the range of the input parameters that can be used for training the models. A comparison between the performance of this two-stage SVM–ANN algorithm with the ANN-only algorithm is shown later in this paper. Using SVM ensures that the CH and BW are predicted to be zero when the experimentally obtained welds show no crown or incomplete penetration, respectively. The criticality of accurately predicting the CH as zero for the welds belonging to Class 0 can be emphasised, considering its effect on the mechanical properties of the weldment. It was found from the experiments that the welds with a positive CH elongate roughly 20% more before failure than those that have a zero CH. They also exhibit a higher TS by approximately the same amount. Predicting a positive CH when the experimentally obtained value is zero can lead to an over-estimation of the mechanical properties. Similarly, predicting a positive BW for welds that are only partially penetrated can also lead to over-estimation of the mechanical properties. The use of a two-stage SVM–ANN algorithm for the prediction of geometrical features can effectively eliminate such miscalculations.

The aim of this research is to develop a computational model that combines the capabilities of different machine learning algorithms, such as SVMs and ANNs, in order to improve the accuracy of prediction of the bead geometry using the limited data available. The data required for developing the model was acquired through a conventional design of experiment. As reported by other researchers, the amount of data was not sufficient to develop an MLP ANN. Acquiring additional data was not feasible due to the significant amount of time and cost involved. In such a case, it was found that applying an SVM to initially classify the data into different classes and then using an ANN belonging to those individual classes to predict the bead geometry can act as a replacement for MLP ANN, without requiring additional data for training. This two-stage algorithm was found to have

visible benefits for predicting the bead geometry, especially for those features where sharp changes in the input–output trends were observed (for example the CH).

2. Materials and Methods

The development of the SVM as well as the ANN to classify the inputs and predict the bead geometry required a large amount of experimental data. In this study, an automated TIG welding process was used to weld 1.5 mm thick 304L stainless steel (SS) sheets heterogeneously, using a 308LSi filler wire. Pairs of sheets with dimensions 200 mm × 60 mm were joined together by welding along the 200 mm length. Three different welds, each with a length of approximately 60 mm and separated by a small distance, were made on every pair of sheets, as shown later in Figure 12. The chemical composition of the base material and the filler wire, as provided by the supplier of the sheets, is shown in Table 1, along with the chromium equivalent (Cr_{eq}) and Nickel equivalent (Ni_{eq}) values that can indicate the susceptibility to solidification cracking using the WRC-1992 diagram. All the experiments were performed using a pulsed current waveform. High purity argon (Ar, 99.995% pure) was used as the shielding gas for most of the welds. However, several researchers have identified the beneficial effects of adding nitrogen (N_2) to Ar in the shielding gas, such as [19], who found that welding in a 20% N_2 + Ar atmosphere requires 40% less current than welding in a pure Ar atmosphere to obtain full penetration of the weld bead, and [20], who reported that the addition of N_2 can improve the impact toughness of stainless steels at sub-zero temperatures. Hence, for some of the experiments, the pure Ar shielding gas was replaced with 2.5%, 5%, or 10% mixtures of N_2 + Ar. Irrespective of the composition of the shielding gas, pure Ar was always used as the backing gas. The variable inputs to the welding process include peak welding current (I_p), torch travel speed (S), pulsing frequency (*f*), filler wire feed rate (R), filler wire diameter (D), and the concentration of N_2 in the shielding gas (N). For all the experiments, the arc length was kept constant at 2.5 mm and the background current (I_b) used was 33% of I_p with a 50% duty cycle.

Table 1. Chemical composition of the materials used in the experiments.

Material	C	Cr	Mn	N	Ni	P	S	Si	Cr_{eq}	Ni_{eq}
Base material (1.5 mm thick)	0.026	17.795	2.000	0.100	8.120	0.025	0.001	0.497	17.795	11.032
Filler 308LSi	0.010	20.000	1.800	0.000	10.000	0.015	0.015	0.800	20.000	10.356

Considering the possibility of a non-linear dependence of the outputs on the inputs, a central composite design (CCD) scheme was chosen for the experiments. From the abovementioned input variables, I_p, S, *f*, and R were continuous variables, whereas D and N were considered discrete due to the limitation on the material and equipment available. The minimum and maximum limits of the parameters used in this research are mentioned in Table 2.

Table 2. Minimum and maximum limits of the variable inputs used for the experiments.

Parameter	Minimum Value	Maximum Value
Peak Current (A)	60	120
Travel Speed (mm/s)	1	4
Pulsing frequency (Hz)	1	6
Filler wire feed rate (mm/min)	100	800
Filler wire diameter (mm)	0.8	1
Nitrogen concentration (%)	0	10
Heat Input (J/mm)	111.69	223.38

However, it was found from initials trials that these parameter ranges were too broad to obtain measurable data from all the experiments. For example, it was found that if an I_p of 120 A was used with

an S of 1 mm/s, the heat input was sufficiently high to cause burn-through of the welds, without giving any useful results. To avoid such welds, the experiments were divided into two sets, based on I_p and S. If the parameters were not divided, any of the conventional design of experiments, such as full factorial, half factorial, or central composite, would include the abovementioned combination of the inputs. Dividing into two sets also ensured that nearly equal amounts of data were available for all the classes of the SVM, making it easier to train the model. The parameter ranges for the divided sets are shown in Table 3. Apart from the I_p and S, no division was required based on other variables, since they are not directly linked to the heat input of the welding process.

Table 3. Parameter ranges after division of the experiments into two sets.

Set 1	Minimum	Maximum
Peak Current (A)	60	90
Travel Speed (mm/s)	1	2
Set 2	**Minimum**	**Maximum**
Peak Current (A)	90	120
Travel Speed	3	4

For each set with four continuous variables, the CCD scheme required 31 experiments at every level of the discrete parameters. In order to reduce the total experiments, only some of these were repeated at other levels of the discrete parameters. For example, out of the 31 CCD experiments performed on Set 1 parameters with 1 mm filler wire and pure Ar shielding gas, only 12 were repeated for each of the 2.5%, 5%, and 10% N_2 + Ar mixtures. With these designs, a total of 62 experiments were performed using 1 mm filler wire and pure Ar as the shielding gas, including both the sets. Another 72 experiments were performed at various levels of N_2 in the shielding gas and a few random experiments were repeated using a wire diameter of 0.8 mm. Finally, 14 additional experiments were performed using random input parameters within the ranges mentioned in Table 2, which were used for validating the developed computational models at a later stage. This led to a total of 180 data points to develop the computational models.

On completing the experiments, metallographic samples were extracted and mounted in conductive bakelite moulds, following which grinding and polishing operations were performed for better visibility of the geometrical features. All the geometrical features mentioned previously could be measured using a Leica DM2700 microscope. Figure 1 illustrates the measurements taken to quantify the considered geometrical features of the weld. Some of the welds were found to have severe misalignment between the base sheets, as shown in Figure 2. Any weld with a misalignment of over 0.35 mm (which complies with the ISO 5817 Level B standard) were not assessed, since it can significantly alter the bead geometry and skew the obtained data. It was also ensured that the metallographic samples were taken from locations at least 20 mm away from the start and stop position of the weld. This was because the slow S at these locations is known to influence the bead geometry.
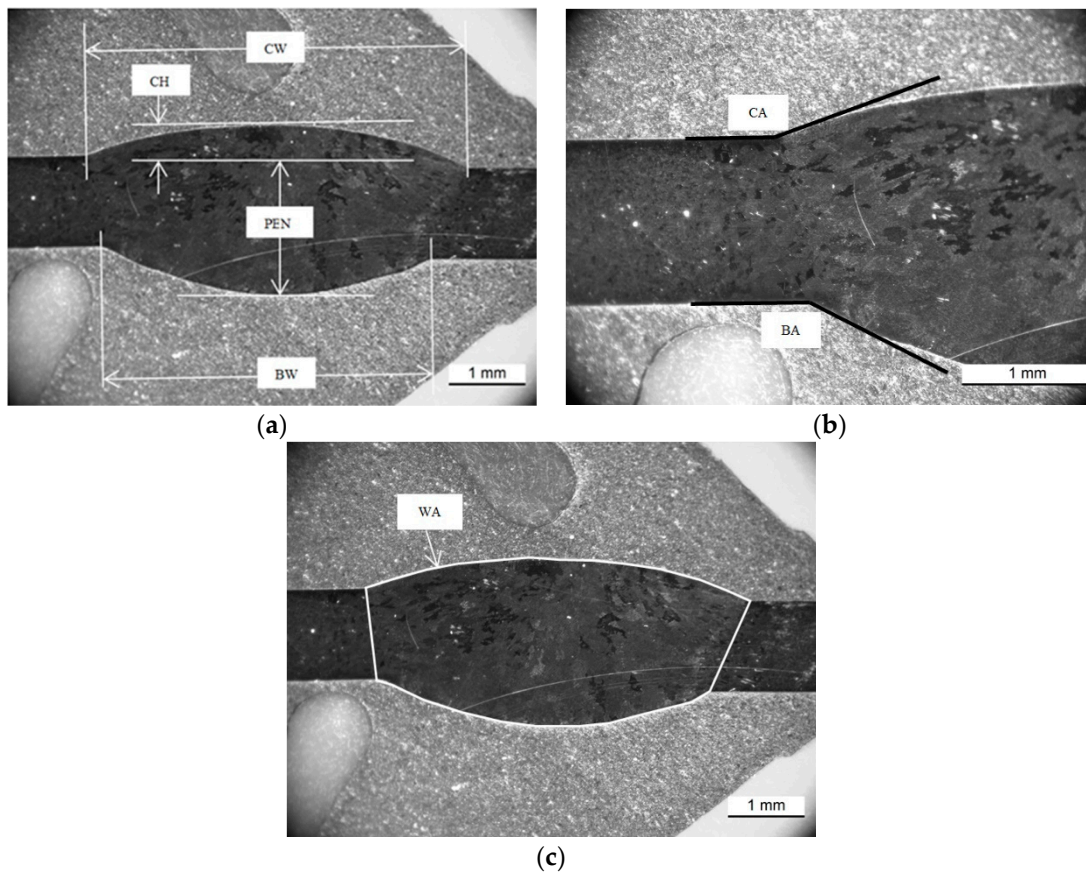
Figure 1. Bead geometry features measured for obtaining the data: (**a**) crown height, crown width, penetration, back width; (**b**) crown angle, back angle; (**c**) weld cross-sectional area. Welding parameters used for this weld: I_p = 90 A; TS = 2 mm/s; f = 1.5 Hz; R = 300 mm/min; D = 1 mm; N = 0%.
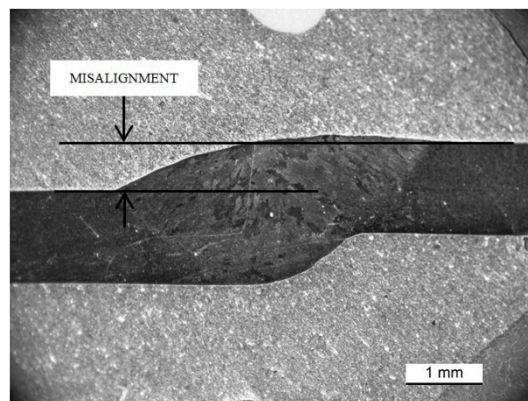
Figure 2. A case of severe misalignment; such welds were not assessed in this study. Welding parameters used for this weld: I_p = 110 A; TS = 3 mm/s; f = 1.5 Hz; R = 400 mm/min; D = 1 mm; N = 0%.

3. Development of Computational Models

As mentioned previously, ANNs could not be directly trained using the experimental data for very low E_{rms} due to the sharp changes observed in the trends of the outputs and the limited data available. This can be illustrated using Figure 3, in which R is plotted on the x-axis against the obtained CH on the y-axis for various I_p values, keeping all other parameters constant. When an I_p of 90 A was used, the obtained CH was practically zero until R reached a value of 100 mm/min. Similarly, when the I_p was increased to 100 A, CH was zero up to an R value of 190 mm/min, and for an I_p of 120 A, it was

zero up to an R value of 340 mm/min. After these values of R, the obtained CH sharply increased, causing such a change that was difficult to capture using a single hidden layer ANN. In such cases, the prediction accuracy can be significantly improved by classifying a set of input parameters into those that will lead to a zero or a positive CH. Although the above example is based on only the I_p, all the other variable inputs can similarly influence the bead profile. This makes it impractical to use a simple "if-else" statement for the classification. SVMs in such cases can be effectively used.
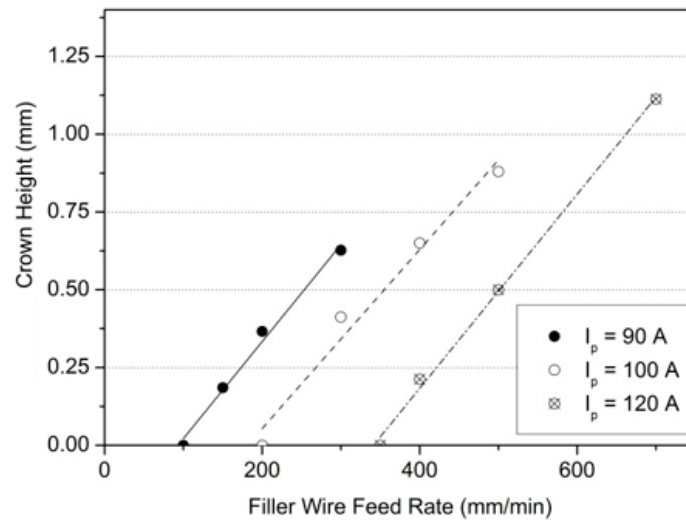

Figure 3. Obtained crown height for different filler wire feed rates for multiple current levels.
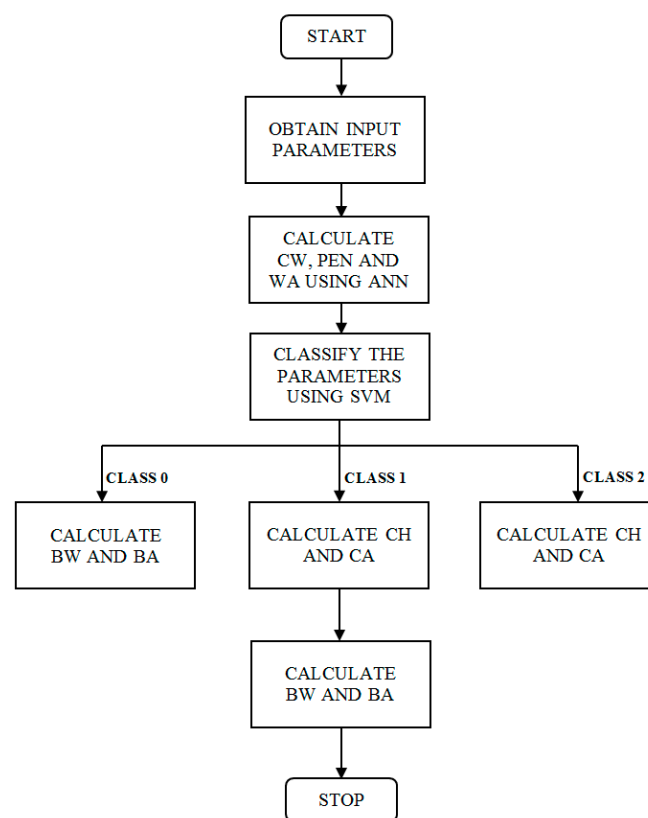

Figure 4. Flowchart of the two-stage SVM–ANN algorithm to predict bead geometry.

Developing computational models using machine learning techniques requires the division of the total data into a training set and a test set. While dividing the data, it was ensured that the test set

lay within the training set extremes, since most of the machine learning techniques underperform in extrapolation, unless the input–output correlation is nearly linear.

Before diving into training and test sets, all the data was normalised between 0 and 1. This helped in giving equal weightage to all the inputs, irrespective of their absolute values. Normalisation also helped in keeping the ANN weights small and avoiding overfitting. The two-stage SVM–ANN algorithm initially classified the input parameters in their respective classes using an SVM and then used the ANN associated with those classes to predict only the necessary geometrical features of the weld bead. The flowchart of this two-stage SVM–ANN algorithm is shown in Figure 4.

The flowchart shows that the CW, PEN, and WA are calculated even before the application of an SVM to classify the input parameters. This is because, irrespective of the class to which the inputs belong, the fusion profile will always have positive values for these three outputs. The other four parameters, CH, CA, BW, and BA, may or may not have positive values, depending on the inputs.

3.1. Development of SVM Models

SVMs are supervised machine learning algorithms that can be used for the classification of data into multiple classes. The principle behind the functioning of an SVM is to establish a hyper plane in an n-dimensional space that can maximize the distance between the two nearest points of different classes. A number of libraries, such as LIBSVM, Weka, and Spider, have been developed by researchers with SVM classification capabilities. Any of these libraries can be effectively used to obtain the hyperplanes. In this research, LIBSVM, developed by [21], was used to build the classification model, since it provided the java interface for SVM development. All the models in this study, including the SVM and ANNs, were developed using java programming language.

SVM classification makes use of kernel functions to determine the similarity between pairs of data points. One of these is the point under consideration, while the other is a specific landmark corresponding to a particular class in an n-dimensional space. These landmarks are obtained from the training data while developing the SVM. Various kernel functions can be used to check the similarity between the point under consideration and the landmark, which include linear, radial basis function (RBF), polynomial, and sigmoid. The most commonly used kernel is the RBF kernel, which makes use of Equation (1), as stated in [21], to estimate the similarity between two data points—x and x'—in an n-dimensional space:

$$k(x, x') \;=\; \exp\!\left(-\gamma \left\| x - x' \right\|^2\right), \tag{1}$$

where $k(x, x')$ is the kernel similarity function, x is the point under consideration, x' is the landmark, $\|x - x'\|^2$ is the Euclidean distance between x and x', and γ is the free parameter.

The free parameter γ plays an important role in determining the similarity between two points, as demonstrated in Figure 5. Assume that similarity between a point (x) and a landmark (zero in this case) is required to be estimated. On the x-axis of this figure is the value of the parameter x, and on the y-axis is the value of the similarity function. A value of 1 for the similarity function indicates complete similarity, since this is obtained only when the Euclidean distance between the two points $\|x - x'\|^2$ is zero. Any other value lower than 1 indicates less similarity. In the case when γ is small (0.2), the similarity function depreciates gradually, increasing the 'reach' of the landmark. This means that values which are slightly away from the landmark may also be considered similar. Conversely, if the value of γ is large, the similarity function depreciates too sharply, which may cause data nearly similar to the landmark be classified as dissimilar. Consequently, an optimum value of γ needs to be identified while training the SVM model.
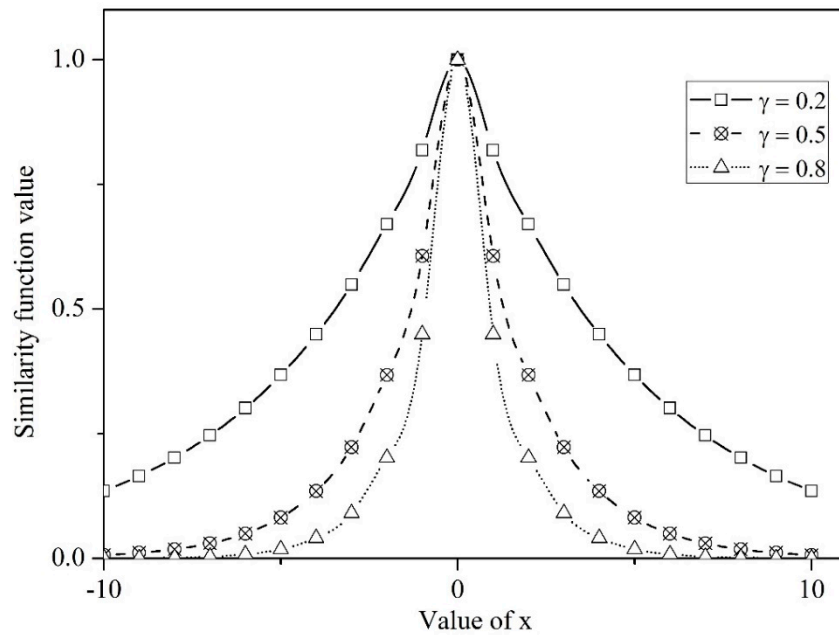
Figure 5. Effect of the free parameter γ on the estimation of similarity.

Another important parameter that must be considered while developing the SVM model is the cost parameter (C), which penalises any misclassification by the SVM. The effect of this cost parameter is illustrated in Figure 6. In this figure, a boundary obtained through training the SVM separates Class 1 and Class 2 parameters. If the value of C is small, some of the outliers are ignored, such as the one shown in Figure 6a. However, when the value of C is large, the model will try to classify every data point in its respective class, which may lead to overfitting. Consequently, the value of C also needs to be optimised in the training phase.
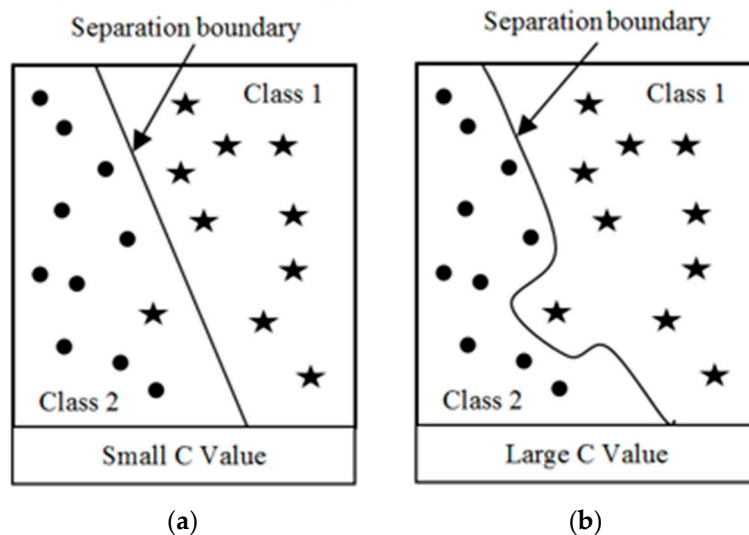

(**a**)   (**b**)

Figure 6. Effect of cost parameter C on classification of data. (**a**) Small C value (**b**) Large C value.

This study considered three different classes based on the fusion profile of the bead: Class 0 for parameters leading to zero CH but full penetration, Class 1 for parameters leading to positive CH as well as complete penetration, and Class 2 for parameters leading to positive CH but incomplete penetration.

Out of all the data points obtained through the experiments, 30 data points (roughly 15%) were retained for testing the model, while the others were used for training the SVM. The parameters γ

and C were estimated using the trial and error method, by varying C between 10^{-4} and 10^4 at intervals of 500 and varying γ between 0.1 and 1 at intervals of 0.1. The accuracy of classification of training and test data was considered to select the best SVM model.
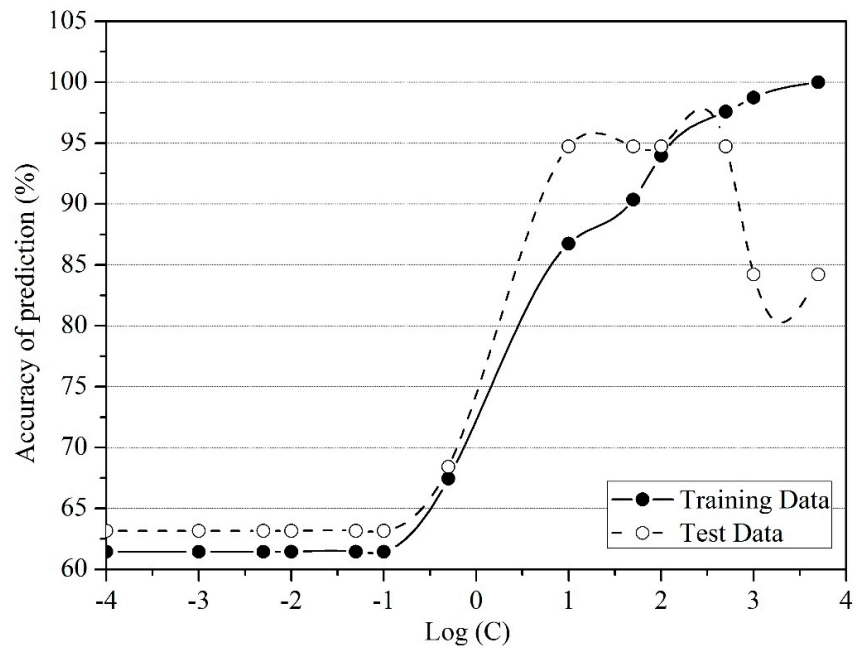

Figure 7. Dependence of the accuracy of classification on the cost parameter C.

The effect of C on the accuracy of classification is illustrated in Figure 7 on a logarithmic scale. The accuracy goes on increasing for the training and test data until the value of the parameter C is 500 (2.7 on logarithmic scale), beyond which the accuracy on the test data decreases, indicating over-fitting. Consequently, the value of the parameter C was chosen to be 500. A similar method was used to estimate the value of γ, as shown in Figure 8.
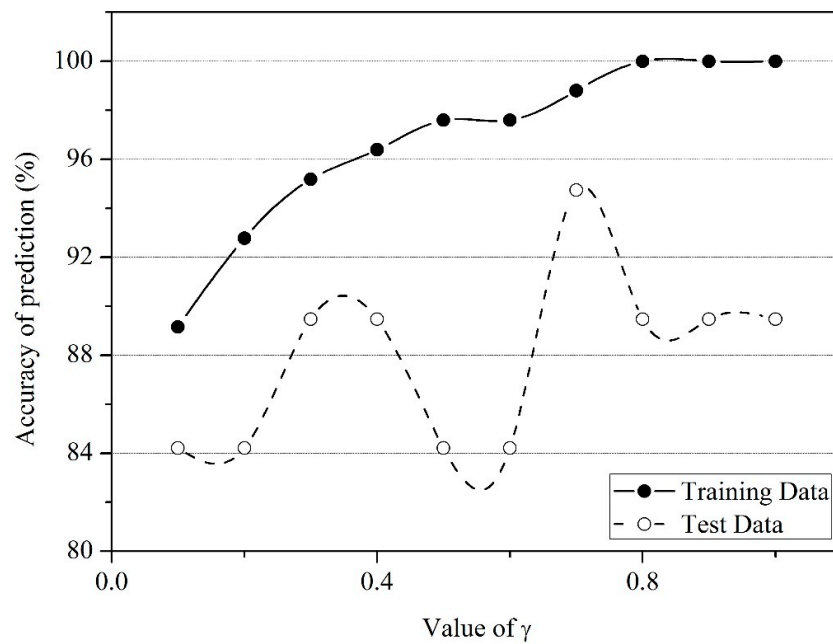

Figure 8. Optimisation of the parameter γ at C = 500.

It can be seen from Figure 8 that when the value of γ is roughly 0.75, the best accuracy for both the training and test data is obtained. With this combination of C = 500 and γ = 0.75, the accuracy on the training data was 98.8% and on the test data was 96.67%. On the test data, the SVM correctly classified 29 out of the 30 data inputs. The small amount of error in this model was from classifying Class 0 parameters as Class 1 parameters on the training as well as the test data, which makes the prediction conservative. This accuracy of classification was acceptable, as attempts to further improve would lead to overfitting.

3.2. Development of ANN Models

ANNs applied to engineering studies effectively consist of neurons in different layers that are connected through connection weights. They typically consist of an input layer, one or more hidden layers, and an output layer. The aim of training a neural network is to establish appropriate connection weights such that the error obtained in predicting the outputs is minimised. Of those many algorithms available to train neural networks, the most commonly used is back-propagation (BP). The details of this algorithm are well documented in several research papers and hence are not discussed in this paper. The readers are requested to refer to other papers, including [7–15], for the details of this.

As for the development of SVM, all the acquired data were divided into a training set and test set; the test dataset comprising roughly 15% of the total available data. In this case too, it was ensured that the test set was contained within the training set to avoid extrapolation. The data was normalised between 0 and 1 in order to give equal weightage to all the parameters. In this study, three different networks were required to be developed, one (ANN_A) to predict CW, PEN, and WA common for all the three SVM classes, the second (ANN_0) to predict the BW and BA required for Class 0 parameters, and the third (ANN_2) to predict the CH and CA for Class 2 parameters. The geometrical features of the welds belonging to Class 1 can be estimated using all three networks (ANN_A, ANN_0, and ANN_2). For developing ANN_A, all the acquired data were used, irrespective of the class to which the inputs belonged, since it predicted features that were common to all the classes. However, for developing ANN_0, only the welds that were fully penetrated, i.e., belonging to Class 0 and Class 1, were used. Likewise, for developing ANN_2, only the data belonging only to Class 1 and Class 2 was considered. A convergence study was carried out to estimate the number of neurons required in the hidden layer of the network for each of the network.

The total E_{rms} for the network was calculated using Equation (2), as used by several researchers [22]:

$$E_{rms} = \frac{1}{p} * \sqrt{\sum_{j=0}^{p} \sum_{i=0}^{n} \left(T_{ji} - O_{ji}\right)^2 / 2,} \tag{2}$$

where T_{ji} is the target value of ith output of the jth pattern, O_{ji} is the output value of the ith output of the jth pattern, n is the number of outputs, and p is the number of patterns (trials).

3.2.1. Development of ANN_A

This network consisted of six input (I_p, S, f, R, D, and N) and three output neurons (CW, PEN, and WA). From the convergence study, it was estimated that seven neurons in the hidden layer were sufficient to obtain the targeted E_{rms} (<0.005 targeted for this network). The network consisted of a bias neuron in the input and hidden layers. A learning rate of 0.5 along with a sigmoid transfer function was used for obtaining the outputs. The training was terminated when either the E_{rms} on the output reached below 0.005, or the number of epochs reached a maximum of 2000. In the development of this network, the total E_{rms} for all the outputs dropped to 0.00497 after 988 epochs. On the test data, the E_{rms} was 0.00373 confirming that the network is neither undertrained nor overfitted.

3.2.2. Development of ANN_0

This network also consisted of the same six input neurons, but predicted only two outputs (BW, BA). It required 10 neurons in the hidden layer. The same learning rate, transfer function, and termination criteria as for ANN_A were used for the development of ANN_0. The total E_{rms} dropped to 0.00481 after 1584 epochs while training this network. On the test data, the E_{rms} obtained was 0.00426.

3.2.3. Development of ANN_2

ANN_2 was similar to ANN_0 in terms of the number of neurons in the input, hidden, and output layers, but predicted CH and CA instead of BW and BA. All the other parameters, including the learning rate and termination criteria, were the same as those used for developing ANN_0. The E_{rms} dropped to 0.00492 after 1195 epochs on the training data. E_{rms} on the test data was found to be 0.00433.

4. Results

Analysis of variance (ANOVA) was carried out using Minitab 2018 software in order to check the existence of any influence of the input variables on the outputs of the process. The p-value—which indicates the probability that the null hypothesis (no influence on the outputs) is true, if lower than 0.05 (95% confidence)—indicates an influence of the input parameter on the output of the weld. The results from ANOVA are summarised in Table 4. It can be concluded from the results that f has minimum influence on the outputs of the welds. This is also reflected quantitatively in the connection weight approach, which is illustrated later in this section to study the relative influence of the variable inputs on the outputs.

Table 5 summarises the results from when the two-stage SVM–ANN algorithm was applied to those 14 welds made for the validation of the model. It also mentions the ANNs that were triggered to predict the bead geometry features (A for ANN_A, 0 for ANN_0, 2 for ANN_2). Table 6 shows the experimentally obtained values for the same welds as in Table 5 and the absolute error in the measurement of the geometrical features. The small absolute errors in Table 6 show that the predictions by the computational models were consistently accurate, except for a few parameters that may be the outliers during the welding process.

Table 4. ANOVA results (p-value) in order to understand the existing influence of the inputs on the outputs of the welds. The results indicate that f has minimum influence on the outputs of the welds.

	I_p	S	f	R	D	N
CH	0.000	0.000	0.414	0.007	0.000	0.000
CW	0.000	0.000	0.685	0.000	0.8	0.000
CA	0.000	0.000	0.834	0.000	0.000	0.409
PEN	0.000	0.000	0.35	0.000	0.014	0.000
BW	0.000	0.000	0.288	0.000	0.26	0.000
BA	0.000	0.006	0.928	0.000	0.004	0.009
WA	0.000	0.000	0.733	0.165	0.002	0.000

If an SVM was not used for the classification of the parameters before applying the ANN, the CH and CA for welds W1 to W4 would have some non-zero value, thus depreciating the prediction accuracy of the model. Similarly, for welds W12 to W14, the values of BW and BA would be non-zero. Thus, using an SVM–ANN hybrid system gives the model the ability to filter certain geometrical features that are not required to be predicted, making the predictions more accurate over a wider range on inputs.

**Table 5.** Bead geometry predictions using the hybrid SVM–ANN algorithm.

Weld ID	I_p	S	f	R	D	N_2	SVM Class	ANNs Used	CH	CW	CA	PEN	BW	BA	WA
W1	90	2	1	300	0.8	10	0	A, 0	–	6.78	–	1.98	6.66	10.69	11.51
W2	120	4	1	200	1	0	0	A, 0	–	5.88	–	1.91	3.86	9.20	8.56
W3	100	3	3	200	1	0	0	A, 0	–	5.79	–	1.83	4.08	14.86	8.98
W4	90	2	1.5	400	1	5	0	A, 0	–	6.16	–	2.09	5.40	22.08	12.44
W5	100	3	6	400	0.8	10	0	A, 0	–	6.03	–	1.86	5.09	13.10	9.76
W6	70	2	1.5	200	1	2.5	1	A, 0, 2	0.46	4.45	21.22	1.73	2.65	12.80	6.90
W7	80	2	1.5	300	1	0	1	A, 0, 2	0.54	5.12	18.64	1.76	3.40	14.94	8.98
W8	110	3	1.5	400	1	0	1	A, 0, 2	0.42	5.78	13.13	1.87	4.23	16.63	10.21
W9	90	1	1	300	1	0	1	A, 0, 2	0.39	7.30	9.61	2.17	6.66	19.54	15.06
W10	90	2	5.5	400	1	0	1	A, 0, 2	0.59	5.62	26.00	1.80	3.88	17.75	10.40
W11	80	2	1	300	1	0	1	A, 0, 2	0.54	5.13	17.75	1.77	3.50	14.61	9.03
W12	70	2	1.5	300	1	0	2	A, 2	0.82	4.17	33.53	0.73	–	–	3.54
W13	90	4	1.5	40	1	0	2	A, 2	0.76	3.93	30.23	0.91	–	–	4.97
W14	70	2	4	50	1	0	2	A, 2	1.25	3.90	66.35	1.12	–	–	6.02

Table 6. Experimentally obtained results for welds done using the same parameters as in Table 5 and the corresponding errors in predicting the geometrical outputs.
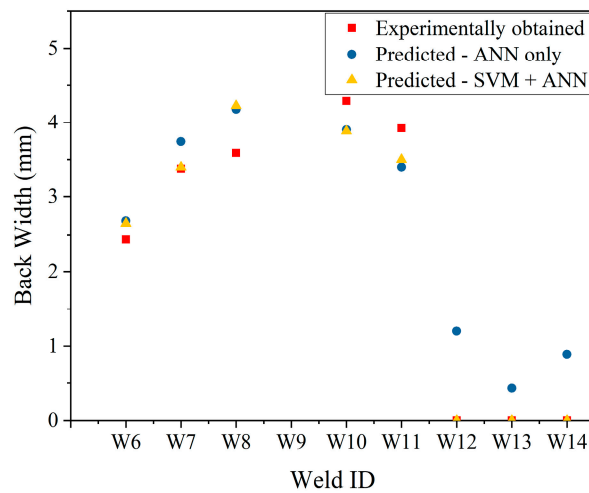
	Experimentally Obtained Results							Error in Estimation Using the Hybrid Algorithm						
Weld ID	CH	CW	CA	PEN	BW	BA	WA	CH	CW	CA	PEN	BW	BA	WA
W1	–	6.64	–	1.94	6.19	12.23	11.62	–	0.13	–	0.03	0.46	1.54	0.11
W2	–	6.06	–	1.95	3.84	10.07	8.831	–	0.18	–	0.04	0.01	0.87	0.27
W3	–	6.08	–	1.79	4.31	11.38	9.3	–	0.29	–	0.04	0.23	3.47	0.32
W4	–	6.40	–	2.03	5.20	17.88	10.78	–	0.24	–	0.05	0.19	4.2	1.65
W5	–	5.85	–	1.86	4.80	13.77	9.12	–	0.17	–	0.00	0.28	0.67	0.63
W6	0.46	4.33	17.16	1.72	2.44	15.95	7.12	0.00	0.12	4.058	0.00	0.20	3.15	0.22
W7	0.60	5.26	22.29	1.66	3.38	15.84	9.10	0.06	0.14	3.657	0.09	0.01	0.90	0.12
W8	0.35	5.81	13.09	1.82	3.59	21.25	9.66	0.06	0.03	0.032	0.04	0.63	4.62	0.54
W9	0.32	7.10	11.13	2.23	6.46	21.29	15.2	0.07	0.19	1.528	0.06	0.2	1.75	0.17
W10	0.54	5.81	23.67	1.81	4.29	16.97	11.63	0.04	0.19	2.329	0.01	0.41	0.77	1.23
W11	0.50	5.34	15.83	1.88	3.92	11.18	9.417	0.03	0.21	1.918	0.11	0.42	3.42	0.38
W12	0.82	4.30	31.38	1.41	–	–	6.691	0.00	0.13	2.142	0.02	–	–	0.15
W13	0.64	3.69	27.23	0.71	–	–	3.257	0.11	0.23	2.994	0.19	–	–	1.71
W14	1.30	3.71	68.86	1.07	–	–	5.824	0.05	0.18	2.51	0.04	–	–	0.19

For comparison purposes, an ANN-only model was developed to predict the bead geometry of the welds. The ANN structure was optimised using the similar procedure that was used for the optimisation of the previously mentioned ANNs. This network consisted of six neurons in the input layer, 10 in the hidden layer, and seven in the output layer. The same termination criteria as for the previous ANNs was also used for this ANN. It was found that after 2000 iterations, the training terminated with an E_{rms} of 0.00755 on the training data and led to an E_{rms} of 0.00814 on the test data. Further training the ANN would lead to overfitting. Figure 9 compares the experimentally obtained CHs with those predicted by the SVM–ANN model and ANN-only model for the welds W1 through W11 from Table 5. It can be seen that for welds W1 to W5, the SVM–ANN model predicted a zero value for the CH, as experimentally obtained. However, the ANN-only model predicted small positive values for the CH. Similarly, Figure 10 shows a comparison of the experimentally obtained BWs with those predicted by the ANN-only and the SVM–ANN model. For welds W12 to W14, the ANN-only model predicted small positive values even when the welds were not fully penetrated.

As seen from Figures 9 and 10, the error in prediction is significantly reduced on using the two-stage SVM–ANN algorithm as compared to the ANN-only algorithm. The small differences in the values obtained experimentally and those predicted using the computational model could be attributed to several factors, including the generally low repeatability of welding processes, certain inevitable variations in the experimental setup, and the accuracy of parameter control. Additionally, the trained computational model had a small error in prediction which, if attempted to eliminate, would lead to overfitting. This error can reflect in the predictions in Figures 9 and 10.
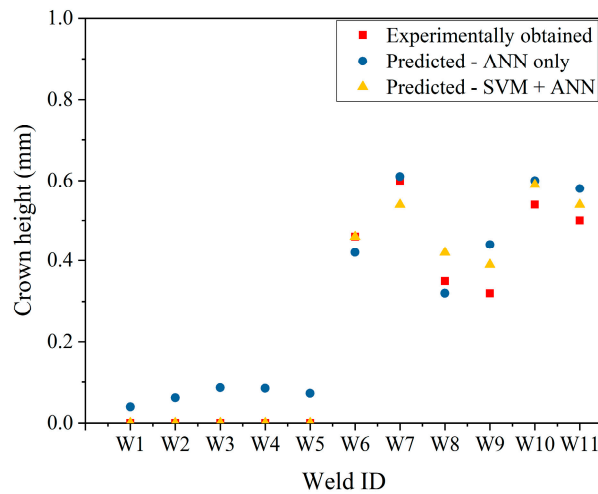
Figure 9. Comparison of the experimentally obtained and predicted CH using the ANN-only and SVM–ANN models for welds W1 to W11 in Table 4.


Figure 10. Comparison of the experimentally obtained and predicted BW using the ANN-only and SVM–ANN models for welds W6 to W14 in Table 4.

Although these CH and BW values predicted by ANN-only model are small, this error in prediction can have significant consequences on the estimation of other mechanical properties of the weldment based on the bead geometry. For example, it can be seen from Figure 11 that the ε largely depends on the CH of the welds. The sub size specimens mentioned in ASTM E8 were used for tensile testing the welded joints. A mechanical extensometer was used to measure the elongation of the weld zone. The position of the samples relative to the welded sheet is shown in Figure 12. As seen from Figure 11, if the CH is zero, the ε is at least 10% lower than that obtained when it has a positive value. Similarly, the TS for such welds is also lower than for those having a positive CH. These small positive values of CH predicted by the ANN-only model would lead to an overestimation of the mechanical properties. Additionally, it can lead to an incorrect estimation of the K_t, and consequently the fatigue life of the welds. This analysis can emphasise the importance of using a two-stage SVM–ANN model for prediction of the bead geometry of the welds.
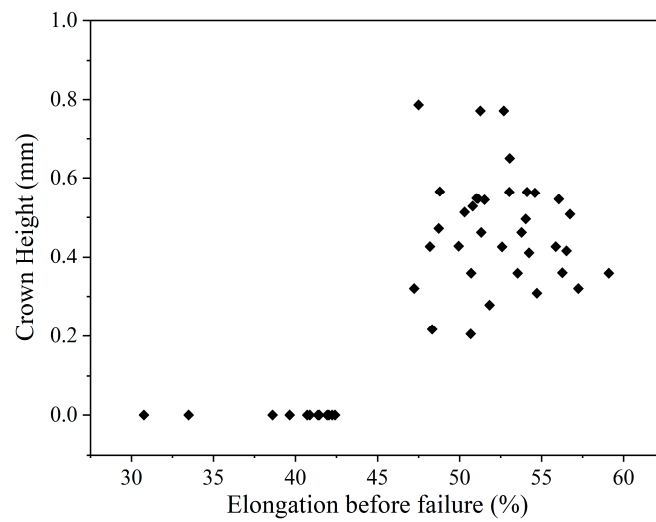
Figure 11. Elongation before failure (ε) plotted against the CH. ε is significantly lower when the CH has a zero value.
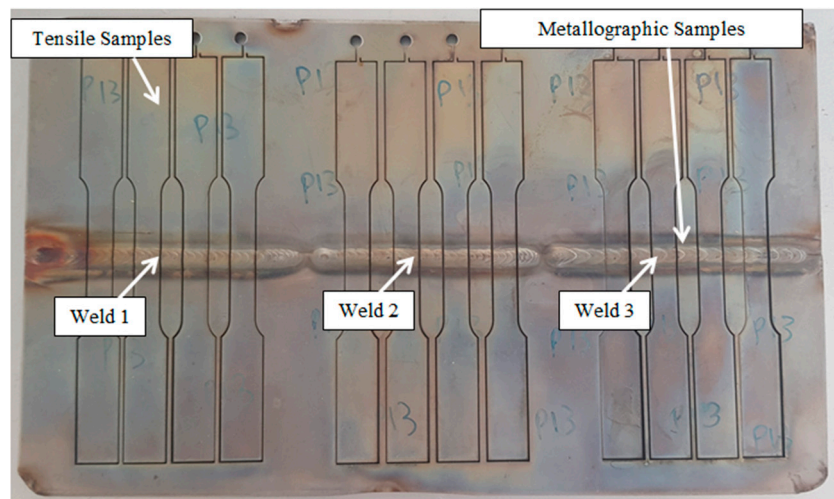
Figure 12. Location of the tensile samples relative to the welded sheets. Four samples were extracted from every weld and tested in tension. Values in Figure 11 are an average of the four tests.

Using the individual ANNs developed for every class of the input parameters, the relative influence of every variable input on the geometrical features can be obtained. This can be done using the connection weight approach, as discussed in [23]. They have reported that the connection weight approach is a better indicator of the relative influence than other approaches such as the Garson's algorithm. Figure 13 shows the relative influence of six input variables on the seven outputs, calculated using the weight of individual ANNs developed for predicting those outputs.

(**a**)

(**b**)

(**c**)
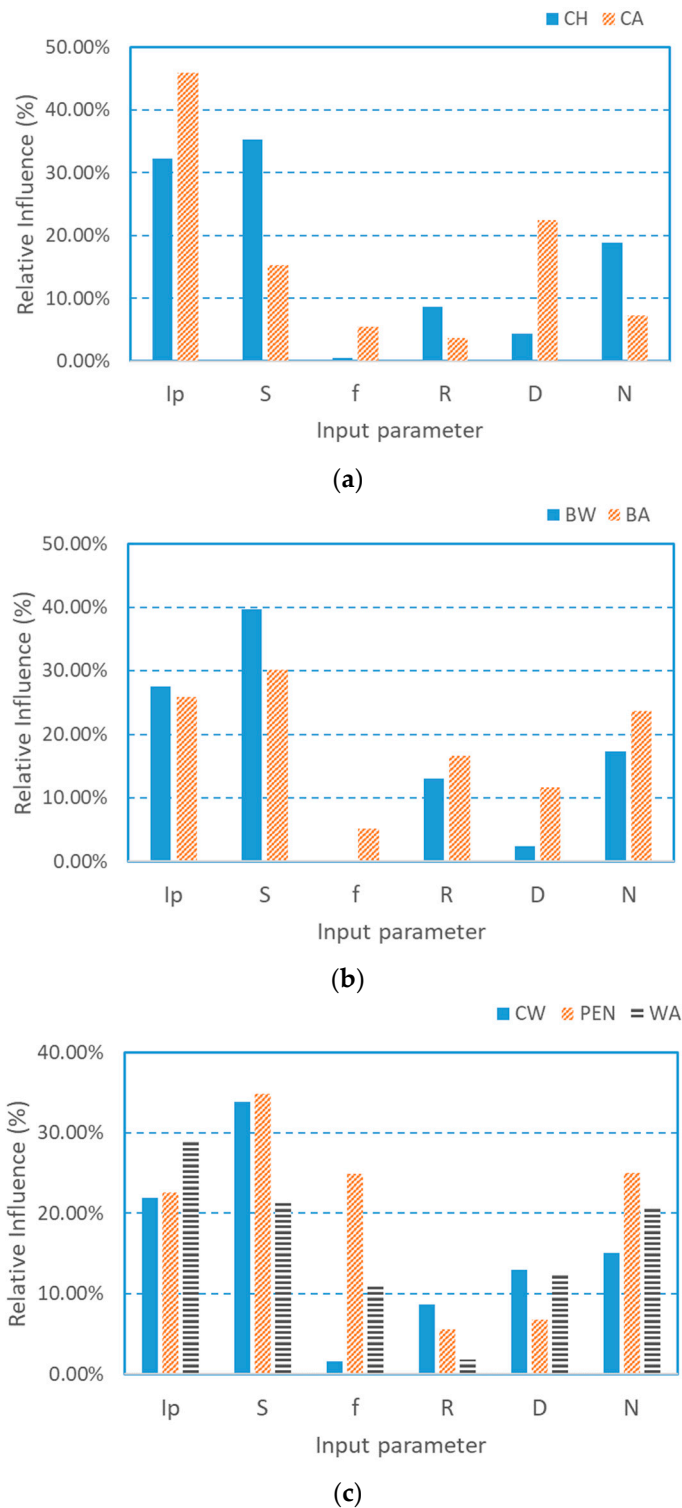
Figure 13. Relative influence of the input parameters in the outputs of the computational model obtained using connection weight approach. (**a**) Influence on CH and CA (ANN_0); (**b**) Influence on BW and BA (ANN_2); (**c**) Influence on CW, PEN and WA (ANN_A).

From Figure 13, it is evident that I_p and S are the most influential inputs for any geometrical feature. Within the range considered in this study, *f* is the least influential. N is more influential than the filler wire related parameters, suggesting that the geometry can be better controlled by changing the shielding gas composition rather than changing the R or the D.

5. Conclusions

From this research and the results produced in this paper, it can be concluded that,

1. Artificial neural networks (ANNs) are extremely efficient in predicting the geometrical features of the weldments and have good capability to approximate non-linear processes. A large number of outputs can be predicted from a set of inputs, however, some pre-processors may be required for the ANNs to function effectively.
2. The performance of a single hidden layer ANN may deteriorate if the trends in the output change sharply. Such situations may occur in the welding process and make the training difficult.
3. Support vector machines (SVMs) can be effectively used as pre-processors to the ANN in cases where the changes in outputs are sharp in order to firstly classify the data into various classes, following which, different ANNs can be applied to predict the geometrical features of the welds belonging to those individual classes.
4. SVMs not only help in significantly improving the accuracy of prediction, but also help in covering a wider range of input parameters.
5. The tensile strength and elongation before failure of the weld depend largely on the bead geometry. Using a two-stage SVM–ANN algorithm can avoid overestimation of the mechanical properties, which is critical for any application.

Author Contributions: Conceptualization, S.J. and R.K.; Methodology, R.K., S.J.; Validation, J.L., J.T.; Formal analysis, R.K. and J.T.; Investigation; Writing—original draft preparation, R.K.; Writing—review and editing, S.J., J.T. and J.L.

References

1. WEE/-/1. *BS EN ISO 5817:2014*; British Standards Institution (BSI): London, UK, 2019.
2. Gunaraj, V.; Murugn, N. Application of response surface methodology for predicting weld bead quality in submerged arc welding of pipes. *J. Mater. Process. Technol.* **1999**, *88*, 266–275. [CrossRef]
3. Schneider, C.; Lisboa, C.; Silva, R.; Lermen, R. Optimizing the Parameters of TIG-MIG/MAG Hybrid Welding on the Geometry of Bead Welding Using the Taguchi Method. *J. Manuf. Mater. Process.* **2017**, *1*, 14. [CrossRef]
4. Yang, L.J.; Bibby, M.J.; Chandel, R. Linear regression equations for modeling the submerged-arc welding process. *J. Mater. Process. Technol.* **1993**, *39*, 33–42. [CrossRef]
5. Vedrtnam, A.; Singh, G.; Kumar, A. Optimizing submerged arc welding using response surface methodology, regression analysis, and genetic algorithm. *Def. Technol.* **2018**, *14*, 204–212. [CrossRef]
6. Imam Fauzi, E.R.; Samad, Z.; Jamil, C.; Nor, N.M.; Boon, G.P. Parametric Modeling of Metal Inert Gas (MIG) Welding Process using Second-Order Regression Model Analysis. *J. Adv. Manuf. Technol.* **2018**, *12*, 367–382.
7. Dutta, P.; Pratihar, D.K. Modeling of TIG welding process using conventional regression analysis and neural network-based approaches. *J. Mater. Process. Technol.* **2007**, *184*, 56–68. [CrossRef]
8. Xiong, J.; Zhang, G.; Hu, J.; Wu, L. Bead geometry prediction for robotic GMAW-based rapid manufacturing through a neural network and a second-order regression analysis. *J. Intell. Manuf.* **2014**, *25*, 157–163. [CrossRef]
9. Okuyuku, H.; Kurt, A.; Arcaklioglu, E. Artificial neural network application to the friction stir welding of aluminum plates. *Mater. Des.* **2007**, *28*, 78–84. [CrossRef]
10. Vitek, J.M.; David, S.A.; Hinman, C.R. Improved Ferrite Number Prediction Model that Accounts for Cooling Rate Effect. *Weld. J.* **2003**, *82*, 10-S.
11. Mirapeix, J.; García-Allende, P.B.; Cobo, A.; Conde, O.M.; López-Higuera, J.M. Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks. *NDT E Int.* **2007**, *40*, 315–323. [CrossRef]

12. Sathiya, P.; Panneerselvam, K.; Abdul Jaleel, M.K. Optimization of laser welding process parameters for super austenitic stainless steel using artificial neural networks and genetic algorithm. *Mater. Des.* **2012**, *36*, 490–498. [CrossRef]

13. Pal, S.; Pal, S.K.; Samantaray, A.K. Artificial neural network modeling of weld joint strength prediction of a pulsed metal inert gas welding process using arc signals. *J. Mater. Process. Technol.* **2008**, *202*, 464–474. [CrossRef]

14. Taffaroj, M.M.; Kolahan, F. A comparative study on the performance of artificial neural networks and regression models in modeling the heat source model parameters in GTA welding. *Fus. Eng. Des.* **2018**, *131*, 111–118. [CrossRef]

15. Ren, W.J.; Wen, G.R.; Liu, S.J.; Yang, Z.; Xu, B.; Zhang, Z.F. Seam Penetration Recognition for GTAW Using Convolutional Neural Network Based on Time-Frequency Image of Arc Sound. In Proceedings of the 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Turin, Italy, 4–7 September 2018; pp. 853–860.

16. Dong, H.; Huff, H.A.; Cong, M.; Zhang, Y.M. Backside Weld Bead Shape Modeling Using Support Vector Machine. In Proceedings of the 2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Honolulu, HI, USA, 31 July–4 August 2017; pp. 277–282.

17. Liang, Y.; Yu, R.; Luo, Y.; Zhang, Y.M. Machine learning of weld joint penetration from weld pool surface using support vector regression. *J. Manuf. Process.* **2019**, *41*, 23–28. [CrossRef]

18. Chollet, F. *Deep Learning with Python and Keras*; Manning Publications Co.: Shelter Island, NY, USA, 2018.

19. Zorc, B. Automatic TIG welding of austenitic stainless steels in nitrogen and nitrogen-based gas mixtures. *Rev. Metal.* **2011**, *47*, 29–37. [CrossRef]

20. Muthupandi, V.; Bala Srinivasan, P.; Shankar, V.; Seshadri, S.K.; Sundaresan, S. Effect of nickel and nitrogen addition on the microstructure and mechanical properties of power beam processed duplex stainless steel (UNS 31803) weld metals. *Mater. Lett.* **2005**, *59*, 2305–2309. [CrossRef]

21. Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27:1–27:27. [CrossRef]

22. Van Ooyen, A.; Nienhuis, B. Improving the convergence of the back-propagation algorithm. *Neural Netw.* **1992**, *5*, 465–471. [CrossRef]

23. Olden, J.D.; Joy, M.K.; Death, R.G. An accurate comparison of methods for quantifying variable. *Ecol. Model.* **2004**, *178*, 389–397. [CrossRef]