Summary

Group CS-23-SW-6-15

February 18, 2026

For our Bachelor's project, we chose to collaborate with the company Migatronic to investigate and hopefully solve a problem of theirs. Migatronic is a leading company in the field of automated welding with high-quality robot-welders among their products.

Migatronic's machines are equipped with different sensors, which can measure, e.g., gas flow, current, and voltage, but up until this point, the company does not provide neither data-based nor any other type of automatic quality assessment of produced welds. Thus, the purpose of this study has been to analyze the data of a set of welding samples to provide the user with an estimate of the homogeneity of the samples. This focus on the relative homogeneity between samples, rather than an absolute good/bad evaluation, stems from the assumption that most welds are made correctly, so the real question is whether there are any welds that appear to be significantly different from the rest.

The focus on homogeneity between samples brings a new approach to the table. When doing our research on the problem, we found that the issue of data-based classification of welds is not new to academia; rather, we were able to find quite a few articles related to the issue. However, the focus in *all* of these articles, was on the absolute classification of individual welds of one specific type, i.e., to be able to classify single samples as either correctly or poorly made. Intuitively, this is a very static approach to data-based weld classification.

With our study, we investigate how a relative classification of welding samples can be made. This way, our algorithm works not on individual samples, but rather on sets of samples that are then compared to each other internally.

Inspired by the related work in the field, we decided on a two-step approach to the classification with a preliminary data analysis followed by a classification performed by a machine learning classifier. For the data analysis, we calculate a set of *predictors* for each sample using statistical methods. The predictors are calculated based on the values of the individual samples relative to the mean values of the total set of samples. To make the comparison more robust, outlier removal is performed on the data set, before the mean values are calculated. After this data analysis, the sets of sample predictors are used as input to a *decision tree classifier*, which is then responsible for the actual classification.

In order to implement our solution, we have made two *Python* programs: an application to train the decision tree classifier, and a user application in which the classifier has been implemented with a graphical user interface, allowing users to upload and examine files containing welding data.

In our solution, we have used different Python libraries, such as *pandas* for data import and allocation, *numpy* and *scipy.stats* for statistical analysis, and *sklearn.tree* for decision tree training. Finally, we have used *matplotlib* for data visualization in the user application.

As input to the training application, we had to generate welding data, as the company had not collected any data from their machines. Data collection and labeling is in general mentioned as a challenge in machine learning, along with things like computational power. In our group, we took a welding course and then went to produce data on Migatronic's site. We made a rather detailed plan ahead of the visit to their site to ensure a reasonable distribution of similar versus diverging samples. We were able to produce and record data from 121 samples during the day we set aside for data collection, and a welding expert carried out the labeling of the produced samples. The decision tree classifier was then trained on this data.

The collaboration with Migatronic has been an overall good experience. We have had several meetings with them, both online and on their site. Especially in the beginning of the study, we spent quite some time defining the problem and their expectations for a solution.

Based on the time and resources that have been available to us, we have developed an algorithm that with 77.2% certainty can identify welds that are different from the majority of the samples. In the test sets, we have tested a combination of 16 good welds and four diverging welds, thus this means that approximately one out of 20 samples, i.e., 5%, might be of a different quality without being caught by the algorithm.

With this, we can conclude that the algorithm is not yet powerful enough for professional use, but we see a strong indication that the chosen method has the potential to solve this problem, even though it requires further development.

Relative Welding Formation Assessment Based on Sets of Current and Voltage Data

Emilie S. Steinmann, **Ibrahim A. A. Rached**, **Julie H. Bengtsson**, **Mikkel H. Dolmer**
{estein19, irache20, jbengt20, mdolme20}@student.aau.dk
Department of Computer Science
Aalborg University

February 18, 2026

***Abstract*—The manufacturing sector is currently going through a digitization phase known as The Fourth Industrial Revolution. This means that cyber-physical systems are being used more widely than ever before. An example of such a cyber-physical system is the Migatronic CoWelder which is a collaborative robot arm connected to a welding machine. Even though the CoWelder is an automated machine, there is still a possibility that different defects could occur in a welding formation. In some cases, it is crucial that the welding formations are strong and reliable. This is often assessed with destructive shear tests or microscopic examination tests which are quite expensive and time-consuming. To ease the process of assessing the quality of a welding formation, a classification system is needed to point out which formations might be likely to have defects.**

The purpose of this study is to use data mining and machine learning to predict which welding formations are good and which are likely to contain defects. This is accomplished in two steps: First, a statistical analysis is made of the current and voltage data obtained through a welding process. Second, a decision tree is generated and evaluated to ensure that it is suitable for Migatronic. The proposed solution consists of a simple user interface in which the decision tree is implemented. This makes it possible for Migatronic and their customers to upload a folder with data from different welds and get the predicted classification. The results indicate that on a data set of 121 welding formations, the welding formations can be classified with an accuracy of 80.2% and a recall of 77.2%.

***Index Terms*—metal inert gas welding; statistical analysis; decision tree classification.**

I. Introduction

The Fourth Industrial Revolution is the current phase in the digitization of the manufacturing sector. This wave of change has led to the emergence of cyber-physical systems[1] in the manufacturing sector, which has given completely new opportunities for manufacturing industries [10].

This is also the case for the welding industry, where semi-automated solutions have been implemented. An example of this is the Migatronic CoWelder [11], which is used in this study. The CoWelder is a combination of a collaborative robot and a welding machine. The CoWelder can be configured with different types of welding processes, such as Metal Active Gas welding, Tungsten Inert Gas welding, or Metal Inert Gas welding. In this study, only Metal Inert Gas (MIG) welding has been investigated.

Even in an automated welding environment, defects, such as cracks, porosity, and incomplete penetration, might occur. These can be caused by several different factors, including erratic wire feeding, contaminated surfaces, or improper gas shield [13]. In order to identify these defects, it is beneficial to have a monitoring system that supports the quality assurance process [18]. The successful development of such a system reduces the need for traditional quality assessment methods, such as the destructive shear, peel, etching, and microscopic examination tests, which are all expensive and time-consuming [15]. Due to their destructive nature, these methods also impose the issue that they cannot assess the actual quality of each individual weld that a user wishes to preserve. In order to do this, the user must resort to methods like x-ray radiography, ultrasonic inspection, or the Eddy current technique, which allow for testing individual welds to assess their exact quality [17]. However, these methods all require additional, highly specialized equipment [17]. Thus, these methods are also expensive and time-consuming for the users. With these observations as the starting point, academia has been seeking simpler and more accessible alternatives to the existing welding quality evaluation methods. The efforts have led to a completely new approach to welding quality evaluation.

The new generation of evaluation methods is based on the relationship between the welding process stability and the welding quality, which, e.g., Gao et al. [3] show for hybrid welding. In their article, they see a correlation between the welding formation and the process stability, which allows for the evaluation of the welding based on the welding process

[1]A cyber-physical system is a system with physical elements being controlled by software through sensors [7].

stability. In general, a process is said to be stable, if the mean and variance of relevant parameters do not change over time [14]. For assessment of *welding* process stability, some commonly used parameters are images of the arc [3, 21], sound recordings [19], and measurements of welding current and voltage throughout the process [4, 9, 13, 18]. Measurements of welding current and voltage are especially popular due to their simplicity and reliability [13].

Regarding the evaluation of welds with current and voltage as stability parameters, we have seen a two-step approach in multiple articles, including [4] and [18]. First, a preliminary data analysis is made, followed by a classification of the data set using a machine learning-based (ML) classifier.

For the preliminary data analysis, two main approaches have been explored: cyclogram representations [8, 9, 13] and statistical measures [9, 18]. In the articles using cyclograms for analysis, we have not seen the use of ML classifiers, but with inspiration from the field of biology, we know that it is possible to use the properties of the cyclogram as input in a neural network [1]. On the other hand, the statistical properties of a data set can be evaluated using a decision tree, as seen in [18]. In this article, Sumesh et al. develop a decision tree with absolute values as splitting criteria, thus their classifier is developed for one specific welding scenario and is designed to assess single welds of the chosen type. This approach seems to be characterizing for all of the work that we have come across.

With the use of welding robots, we are experiencing that most welds are made correctly. Therefore, in this study, we propose a different approach to welding assessment. In contrast to [4, 15, 18, 19], we propose an algorithm that does not assess the absolute quality of a single weld, but rather identifies any diverging samples from a set of welds. The applied methods follow and extend the methods from [18], in particular.

The main contributions of this paper are as follows:

- An algorithm that allows for easy detection of diverging welds by first calculating statistical predictors from current and voltage data. The predictors are then used to generate a decision tree classifier.
- The solution provides a new, dynamic approach to welding classification, as it calculates predictors relatively to the entire data set.
- The use of outlier removal performed using the modified z-score remarkably improves the efficiency of the algorithm.
- The algorithm is trained to be tolerant of the natural variation between welding samples through experiments.
- Experiments on a set of 121 real welding samples suggest that the proposed algorithm can achieve 80.2% accuracy, 50.1% precision, and 77.2% recall.

The rest of the paper is organized as follows. We first introduce preliminaries in section II. The proposed methods used in the experiments are described in section III. Section IV covers the experimental study. Lastly, a conclusion is drawn in section V.

II. Preliminaries

In section I, we mentioned different analysis techniques. The relevant ones for this study are statistical analysis and decision tree classification, which we give a theoretical introduction to in this section.

A. Statistical Analysis

In this paper, the preliminary data analysis includes calculations of the statistical measures *mean*, *median*, *standard deviation*, *skewness*, and *kurtosis*, as is also seen in [18]. The *mean* is found by dividing the sum of the values in a set by the number of values [2]. The *median* is the middle value of the ordered data set, or, in the case of an even number of values, the mean of the two middle values [2]. The *standard deviation* is a measure of the spread of a data set, i.e., the difference between each value in the data set and the mean. The standard deviation is computed as follows [2]:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \tag{1}$$

where it appears that the standard deviation is a measure of the square root of the mean squared difference between each element, x_i, and the calculated mean, $\bar{x}$.

Skewness and *kurtosis* are two attributes of the distribution of the values. Skewness describes symmetry, and can be either positive or negative depending on whether an excessive number of values are found on the right or the left side of the mean, respectively [2]. Skewness can be computed using the Fisher-Pearson coefficient, which is calculated as follows [14]:

$$S_k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} \tag{2}$$

where $\bar{x}$ is the mean and s is the standard deviation.

Kurtosis, on the other hand, describes how heavy or light the tails of the distribution are. Kurtosis is calculated as follows [14]:

$$kurtosis = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3 \tag{3}$$

where a low value indicates a heavy-tailed distribution, i.e., a rather even distribution of values around the mean, and a high value indicates a light-tailed distribution, i.e., having a substantial share of the values close to the mean.

B. Decision Tree Classification

A decision tree is a simple and traditional method for supervised classification learning. The objective of a decision tree is to correctly classify elements of the domain [16].

A decision tree starts at the root node consisting of all the samples in the training data set. From here, the tree is recursively partitioned on some splitting criterion into purer subsets, also called internal nodes, until all the samples in the partition belong to the same class or a stopping criterion is met. The nodes in a decision tree which can not be further split are called leaf nodes. Each leaf node is associated with exactly one class. When classifying an element, the element traverses through the tree, starting from the root node, until it ends up in a leaf node. The class of the leaf node is then the prediction for that element [16].

After the tree has been designed, it can be evaluated by classifying known elements from a test data set. The predictions can result in one of four outcomes, depending on the correctness of the classification:

- True positive (TP): A correct positive prediction.
- True negative (TN): A correct negative prediction.
- False positive (FP): A wrong positive prediction.
- False negative (FN): A wrong negative prediction.

Based on these values, there are different metrics that can be used to evaluate the performance of the classifier, such as accuracy, precision, and recall. These metrics are defined as shown in Equation 4, 5, and 6:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

These metrics are often weighted differently, depending on the classification problem. Having a high accuracy is generally important, since this is a measure of how many of the elements were correctly classified. A high precision is important if the amount of FP should be minimized, as this describes how many of the elements classified as "positive" that were actually positive. A high recall is important if the amount of FN should be minimized, as this describes the number of correct "positive" predictions out of all "positive" predictions that should have been made [16].

III. Methods

The classification of the welding formations is designed as a two-step algorithm based on the theory described in section II. The algorithm is provided in Algorithm 1. This section describes how the methods are applied, and it is ordered accordingly with the two steps of the algorithm, which consist of a preliminary data analysis followed by the construction of a decision tree classifier.

Algorithm 1 Algorithm for classifying welding formations.

Input: Data sets for a set of welding samples, D
Output: a decision tree classifier, clf

// preliminary data analysis
$basic\_statistics \leftarrow$ compute according to Equation 1-3 in addition to mean and median for every element in D

$mean\_statistics \leftarrow$ compute mean of $basic\_statistics$ with outlier removal following Equation 7
$predictors \leftarrow$ compute according to Equation 9

// classifier generation
clf $\leftarrow$ generate decision tree classifier from $predictors$

return clf

A. Preliminary Data Analysis

The first step of the algorithm is the preliminary data analysis, which is responsible for calculating the predictors that will be used to determine the splitting criterion for each node in the decision tree.

As discovered in section I, current and voltage signals are commonly used to assess welding quality. By analyzing the current and voltage signals of welding samples, it is clear to see that the uniformity of the transients[2] is an indication of the quality of the welding sample. As shown in Figure 1, a welding sample with an undisturbed arc has uniform transients and the current and voltage signals are stable throughout the welding process.

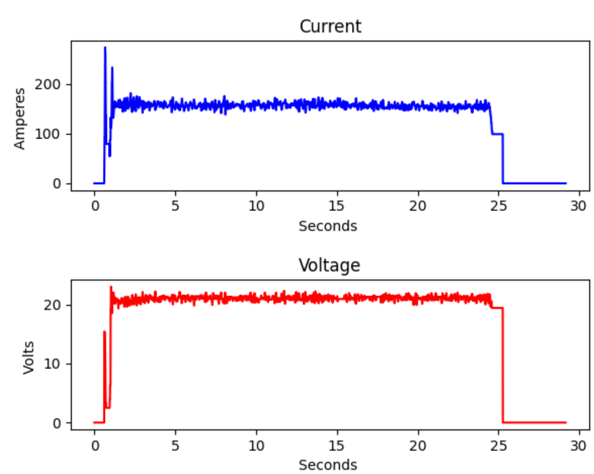

Fig. 1. A welding sample with an undisturbed arc (sample 77).

[2]A momentary variation in current, voltage, or frequency.

In contrast, a welding sample with defects is unstable because of significant drops in the current and voltage signals, as seen in Figure 2.

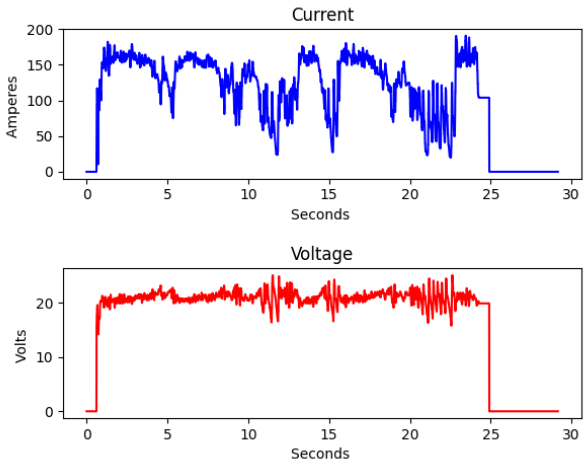

Fig. 2. A welding sample with a disturbed arc caused by erratic wire feeding (sample 113).

It is exactly these variations in the measured data of samples with and without defects that we aim to bring forward in the calculation of the predictors.

The calculation of the different predictors is accomplished in three phases. First, statistical measures are calculated for the individual data sets. Second, the mean values of the results of the first phase are calculated. Finally, in the third phase, the individual results are compared to the mean values across all samples. This results in a set of relative measures for each sample by which they can be classified. This process is illustrated in Table I.

TABLE I
PREDICTOR CALCULATION PHASES.

	s	k	$\lvert 1 - s_i/\bar{s} \rvert$ **(9)**	$\lvert 1 - k_i/\bar{k} \rvert$ **(9)**
S1	s_1	k_1	$\lvert 1 - s_1/\bar{s} \rvert$	$\lvert 1 - k_1/\bar{k} \rvert$
S2	s_2	k_2	$\lvert 1 - s_2/\bar{s} \rvert$	$\lvert 1 - k_2/\bar{k} \rvert$
mean	$\bar{s}$	$\bar{k}$		

Table I provides a simplified example of how the two predictors standard deviation, s, and kurtosis, k, are calculated for the two samples $S1$ and $S2$. In the first phase, basic statistical measures are calculated, as shown in the part highlighted in yellow. Second, the mean values across samples are calculated for each of the yellow columns. These are highlighted in green. Finally, in the third phase, the predictors are calculated based on the basic statistical measures from the first phase and the mean values from the second phase. The results of the third phase are highlighted in orange. Exact descriptions of each phase are given in the following subsections.

1) Statistical Measures: In the first phase of the preliminary data analysis, which is highlighted in yellow in Table I, statistical measures for voltage (U), current(I), power ($U \cdot I$), and resistance (U/I) are used to capture the trends of the data. The statistical measures used in this paper are *mean*, *median*, *standard deviation*, *skewness*, and *kurtosis*.

2) Mean of Statistical Measures: In order to identify diverging welds, it is necessary to calculate the mean of the aforementioned statistical measures for the whole set. This phase is highlighted in green in Table I.

To ensure that anomalous data points do not significantly influence the mean values of the whole set, modified z-scores are calculated for the data set in order to detect and remove outliers. The modified z-score is a measure of how much a particular value differs from the median. It is calculated as [5]:

$$\text{modified z-score} = \frac{0.6745 \cdot (x_i - \widetilde{x})}{\text{MAD}} \tag{7}$$

where $\widetilde{x}$ is the median, and MAD is the median absolute deviation, which is calculated as [5]:

$$\text{MAD} = \text{median}_i\{|x_i - \widetilde{x}|\} \tag{8}$$

It is recommended that values with modified z-scores less than -3.5 or greater than 3.5 are labeled as potential outliers [5], so when calculating the mean in this phase, we only include data points with a modified z-score that lies between -3.5 and 3.5.

3) Deviation from Mean: The last phase of the preliminary data analysis, which is highlighted in orange in Table I, is to calculate how much the statistical measures of a single welding sample deviate from the mean values of the statistical measures for the whole set. Regarding the mean current for each sample, this is also compared to the desired current for the welding, which is defined by the user. The deviation is calculated as follows:

$$d = |1 - \frac{x_i}{\bar{x}}| \tag{9}$$

where x_i is the basic statistical measure for the i'th sample (calculated in phase 1), and $\bar{x}$ is the mean (calculated in phase 2). The deviation is calculated for all predictor types for all samples. The deviations are then used as predictors in the decision tree classifier. In total, there are 21 predictor types.

B. Decision Tree Classifier

The second step of the algorithm is to generate the decision tree classifier based on the predictors calculated in the preliminary data analysis. All predictors are taken into consideration when the tree is designed, but only those resulting in high-quality splits, are included in the final decision tree classifier.

The decision tree classifier can be configured with different parameters, which will affect how the tree is generated. The parameters we found to be the most relevant for our classifier were:

- Criterion: The function used to measure the quality of a split.
- Height: The maximum height of the tree.
- Class weight: Weights associated with the classes.

In order to have an efficient decision tree classifier, the tree must be split in a way that maximizes information gain. The criterion parameter is therefore important, since it determines the strategy used for splitting at each node. The strategies considered in this study are gini index and entropy.

Additionally, the height of the tree is an important parameter of the configuration. If the height is not limited, then the decision tree will continue to grow until all the nodes are pure, which would result in an overfitted tree. Conversely, if the tree is too shallow, then the nodes would have high impurity, and the tree would therefore be underfitted.

In the case of an uneven number of samples in each class, the class weight can be used to balance out the classes. Furthermore, if it is desired to have one class in favor of another, the class weight of this class can be increased.

The exact values of the parameters used in this paper are described in subsection IV-D.

IV. Experimental

A. Setup

The MIG welding experiment was conducted using the Migatronic CoWelder [11] consisting of a collaborative robot arm from Universal Robots [20] and a welding machine from Migatronic [12]. The welding machine and the robot arm were connected through a robot control interface (RCI). The RCI was connected to both the welding machine and a programmable logic controller that captured the data from both the welding machine and the robot arm. The captured data included measurements of the welding current, voltage, gas flow, wire inch speed, time, and arc detection. The weld data was captured at a sampling frequency of 120 Hz.

Prior to the experiment, a professional welder installed the CoWelder correctly and set up the program for this specific experiment. In the program, the welding current was set to 150 A and the voltage was set to 20.9 V. The welding head moved with a velocity of 6 mm/s. To ensure conformity between samples, a customized plate holder had been installed on the welding table. The metal plates for each sample could then be placed accurately each time. The exact welding setup is shown in Figure 3.

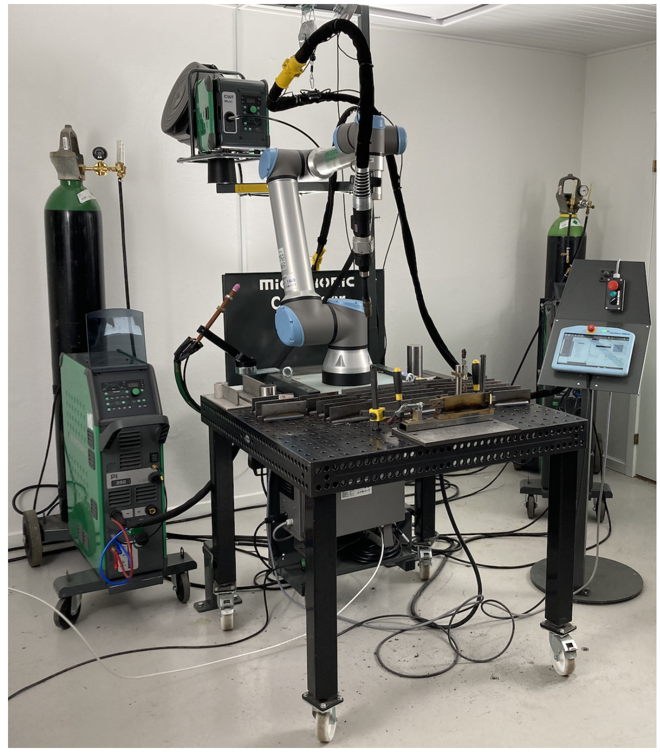

Fig. 3. The welding setup.

In the figure, the green welding machine can be seen to the left, and the blue collaborative robot arm on the table. The customized plate holder is placed on the table in the front. For creating a sample we simply had to place the plates in the holder, and press a button on the screen, seen to the right.

B. Samples

For this experiment, data from 121 welding samples was collected. For all samples, 3 mm thick steel plates were used, and the weld joints were designed as a tee joint [6] where only one side of the material was welded. The plates were 150x50 mm and were welded together along the long side resulting in a welding formation with a length of 140 mm. Two welding samples are shown in Figure 4 and Figure 5.

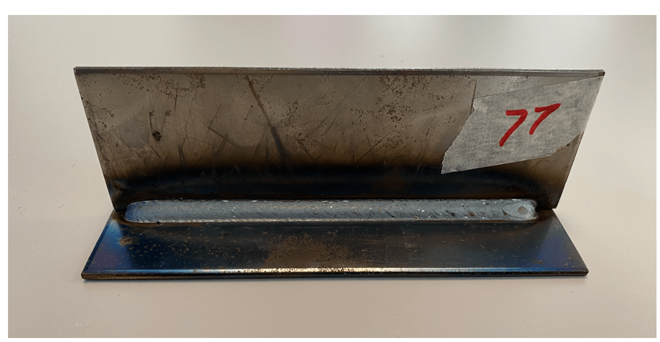

Fig. 4. Welding sample with no defects (sample 77).

Fig. 5. Welding sample with a defect caused by erratic wire feeding (sample 113).

To generate deviating welds, some of the welding defects described in section I were recreated in this experiment. Out of the 121 samples, 14 samples were sprayed with paint to simulate a contaminated surface on the steel plates. For 16 samples, the gas shield was disturbed by creating a variable airflow around the arc. Incomplete penetration was simulated by grinding holes of different sizes in the steel plates. 10 samples were made with one hole of sizes between 3-21 mm, and 11 samples had two holes of sizes between 5-26 mm. Finally, erratic wire feeding was simulated on 9 samples by holding back the wire spool to create an uneven welding formation. The remaining 61 samples were made without attempts at enforcing any defects.

In total, data from 121 samples was collected. The samples were classified as either "approved" or "not approved" by a professional welder. 65 of the 121 samples were classified as "approved" and 42 samples were classified as "not approved". The remaining 14 samples had indications of instabilities in the process, and were considered "not approved" in the solution. The full schema of the samples can be found in Table IV.

In order to test our methods, we partitioned our sample data into a training and a test set. In their article, Sumesh et. al. use a split of 80% of the data for training and 20% for testing [18]. Inspired by this, we chose a split with around 75% of the data in the training set (91 samples) and 25% of the data in the test set (30 samples), to allow for making different subsets of the test data for evaluation, as described in the next subsection. We chose the samples for the training and test sets, such that they had an equal distribution of approved and not approved samples, and such that the different types of errors that we had inflicted on the samples were likewise as equally distributed as possible. In the training set, we have 42 "not approved" samples and 49 "approved" samples. In the test set, we have 14 "not approved" samples and 16 "approved samples".

C. Performance Metrics

To evaluate the decision tree, we use the metrics accuracy (Equation 4), precision (Equation 5), and recall (Equation 6). In this paper, the "positive" class are the welding samples that are "not approved" and the "negative" class are the samples that are "approved". Therefore, a FN would be a "not approved" welding sample that has been classified as "approved". It is of high importance that as few FN's as possible occur, since these are an indication of how many welding defects that are not found by the classifier. This means that the most relevant metric for this classifier is recall.

In the experiments, samples in the test set are traversed through the tree to see how many of them are classified correctly. In order to get a more robust evaluation, the decision tree is tested on different subsets from the test set. More precisely, the tree is tested on 1000 combinations of 20 welding samples, where 16 of them are "approved" welding samples and four of them are randomly selected "not approved" welding samples. The mean values of accuracy, precision, and recall from the 1000 combinations are then calculated.

D. Results

To test our algorithm, we used the training data as input to a decision tree generator in Python. We explored different configurations of the tree through experiments with the parameters described in subsection III-B. The best configuration that we found was the following:

- Criterion: entropy
- Height: 5
- Class weight: 1.365:1 (not approved:approved)

Here, especially the class weight was important, since our two classes were imbalanced with more "negatives" (approved samples) than "positives" (not approved samples) in our training set. To counter this over-representation of "positive" samples, class weighting was defined for the decision tree classifier, such that the training data would be more balanced, and even slightly in favor of the "positive" class, such that it would increase recall.

The evaluation result of this configuration is seen in Table II. The evaluation is performed in accordance with subsection IV-C, and the table shows the mean results. From this we can see, that more than three out of four "not approved" (NA) samples are classified correctly, resulting in a recall of 77.2%. Approx. three out of the 16 "approved" (A) samples are misclassified as "not approved", so that brings the precision down to 50.1%, though the overall accuracy is still 80.2%.

TABLE II
EVALUATION RESULT OF THE CLASSIFIER.

	NA (actual)	A (actual)
NA (predicted)	TP = 3.088	FP = 3.056
A (predicted)	FN = 0.912	TN = 12.944

Accuracy	0.802
Precision	0.501
Recall	0.772

In order to achieve these results, we decided to implement the outlier removal in phase two of the statistical analysis, as described in subsection III-A. As can be seen from comparing Table II and Table III, this had a huge impact on the efficiency of the classifier.

TABLE III
EVALUATION RESULT OF CLASSIFIER MADE W/O OUTLIER REMOVAL.

	NA (actual)	A (actual)
NA (predicted)	TP = 2.613	FP = 9.993
A (predicted)	FN = 1.387	TN = 6.007

Accuracy	0.431
Precision	0.386
Recall	0.653

Table III shows the evaluation of a decision tree classifier with the same configuration and input as the one evaluated in Table II. The only exception is that for this classifier, outlier removal had not been performed in phase two of the statistical analysis. As can be seen, this massively affects the performance metrics. Starting with recall, only 2.6 "not approved" samples were classified correctly on average, bringing recall down to 65.3%. The high class weight on the "not approved" class does for this tree lead to a high misclassification rate for the "approved" class, such that approx. 10 out of 16 "approved" samples are classified as "not approved" each time. Therefore, both accuracy and precision score below 50%. This underlines the importance of outlier removal in the calculation of the predictors.

The decision trees for the classifier with and without outlier removal can be seen in Figure 6 and Figure 7, respectively. Focusing again on the best classifier, which makes use of outlier removal in the statistical analysis, we can examine the predictors which have proved relevant to the problem. As described in subsection III-A, the predictors are calculated as statistical measures on the four data sets of the welds: current, voltage, resistance, and power.

The relevancy of each predictor is reflected in the hierarchy of the decision tree in Figure 6, where the predictors used for splitting closest to the root node are contributing more than those used farther away from the root. Intuitively, the features that are not used at all do not contribute to the classification. Additionally, some predictors appear multiple times in the decision tree, which increases their contribution.

From Figure 6, we can see that for the statistical measures, standard deviation and skewness are the most dominant features, followed by mean and kurtosis. The deviation from the desired current and the median measures are not included in the tree at all. Out of the four data sets for each weld, voltage and resistance plays the most prominent role, followed by power. Current is only directly involved in one split, making it the least prominent data set.

V. Conclusion and Future Work

The solution presented in this paper is a two-step algorithm, that first performs a statistical analysis followed by classification via a decision tree. In the statistical analysis, predictors are calculated relatively to reference values describing the entire data set. Modified z-score is used to perform outlier removal when calculating the reference values. This improves the robustness and overall quality of the solution. With this solution, experiments show that 80.2% of the welds are classified correctly (accuracy) and 77.2% of the welds that stand out from the majority are successfully identified (recall). To achieve this relatively high value for recall, we have compromised on the precision, so that only around 50.1% of the welds classified as "not approved" actually belong in that class. As the solution stands now, it is not precise enough for professional use. However, the current solution gives a clear indication of the potential of this kind of analysis and classification in the field of automated welding.

In future research, it is of interest to apply the proposed algorithm to different welding setups and to explore more decision tree configurations.

Acknowledgments

We would like to thank Migatronic [12] for supporting this study and for helping out with data gathering by providing the necessary materials and letting us borrow a CoWelder [11]. We would also like to thank Migatronic [12] for classifying all the samples used in the experiments with the proposed algorithm. Finally, we would like to thank our supervisor, Tianyi Li, for all of her feedback and experience with paper writing.

References

[1] Thiago Bruno Caparelli and Eduardo Lázaro Martins Naves. "Reconstruction of gait biomechanical parameters using cyclograms and artificial neural networks". In: *Research on biomedical engineering* 33.3 (2017). ISSN: 2446-4732.

[2] William P. Fox. *Probability and Statistics for Engineering and the Sciences with Modeling Using R.* Textbooks in Mathematics Ser. Milton: CRC Press LLC, 2023. ISBN: 1-003-31790-1.

[3] Xiangdong Gao et al. "Process stability analysis and weld formation evaluation during disk laser–mag hybrid welding". In: *Optics and lasers in engineering* 124 (2020). ISSN: 0143-8166.

[4] Yong Huang et al. "A Weld Quality Classification Approach Based on Local Mean Decomposition and Deep Belief Network". eng. In: *Journal of materials engineering and performance* 30.3 (2021). ISSN: 1059-9495.

[5] Boris Iglewicz and David C. Hoaglin. *How To Detect And Handle Outliers*. 1993.

[6] Universal Technical Institue. URL: https://www.uti.edu/blog/welding/joint-types. (Accessed on 04/04/2023).

[7] IotOne. *Cyber Physical Systems (CPS)*. 2023. URL: https://www.iotone.com/term/cyber-physical-systems-cps/t145. (Accessed on 23/05/2023).

[8] Seamkong Kuoch, Eakkachai Warinsiriruk, and Sutep Joy-A-Ka. "Digitalized Evaluation of Welder Skill by using Cyclogram Characteristics". In: *MATEC Web of Conferences* 269 (2019). ISSN: 2261-236X.

[9] Jianxiong Li et al. "Effect of torch position and angle on welding quality and welding process stability in Pulse on Pulse MIG welding–brazing of aluminum alloy to stainless steel". In: *International journal of advanced manufacturing technology* 84.1-4 (2016). ISSN: 0268-3768.

[10] McKinsey. *What are Industry 4.0, the Fourth Industrial Revolution, and 4IR?* 2022. URL: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-are-industry-4-0-the-fourth-industrial-revolution-and-4ir. (Accessed on 29/03/2023).

[11] Migatronic. URL: https://cowelder.com/. (Accessed on 30/03/2023).

[12] Migatronic. URL: https://www.migatronic.com/. (Accessed on 23/05/2023).

[13] Syed Quadir Moinuddin and Abhay Sharma. "Arc stability and its impact on weld properties and microstructure in anti-phase synchronised synergic-pulsed twin-wire gas metal arc welding". In: *Materials in engineering* 67 (2015). ISSN: 0261-3069.

[14] National Institute of Standards and Technology. *NIST/SEMATECH e-Handbook of Statistical Methods*. 2012. URL: http://www.itl.nist.gov/div898/handbook/. (Accessed on 28/03/2023).

[15] Y. J. Park and H. Cho. "Quality evaluation by classification of electrode force patterns in the resistance spot welding process using neural networks". In: *Proceedings of the Institution of Mechanical Engineers. Part B, Journal of engineering manufacture* 218.11 (2004). ISSN: 0954-4054.

[16] David Poole and Alan Mackworth. *Artificial Intelligence: Foundation Of Computational Agents*. 2017.

[17] John Stavridis, Alexios Papacharalampopoulos, and Panagiotis Stavropoulos. "Quality assessment in laser welding: a critical review". In: *International journal of advanced manufacturing technology* 94.5-8 (2018). ISSN: 0268-3768.

[18] A. Sumesh et al. "Decision tree based weld defect classification using current and voltage signatures in GMAW process". In: *Materials Today: Proceedings*. Vol. 5. 2. Elsevier Ltd, 2018.

[19] A. Sumesh et al. "Use of Machine Learning Algorithms for Weld Quality Monitoring using Acoustic Signature". eng. In: *Procedia Computer Science* 50 (2015). ISSN: 1877-0509.

[20] Univeral Robots. URL: https://www.universal-robots.com/da/. (Accessed on 24/05/2023).

[21] Li Zhiyong et al. "On the Use of Arc Radiation to Detect the Quality of Gas Metal Arc Welds". In: *Materials and manufacturing processes* 26.7 (2011). ISSN: 1042-6914.

Appendix A: Samples

February 18, 2026

Table IV is an overview of the planned type of samples and the actual type of samples. Originally, we planned to collect the type of samples in the “Planned”-column but we ended up creating the sample types seen in the “Actual”-column. In the table, XA stands for “excessive amount”, which refers to the amount of paint on paint samples. In total, there are 121 samples, 61 with no defects, 14 with paint, 10 with one hole, 11 with two holes, 16 with improper gas shield, and nine with thread spool defects.

The “Label”-column shows the classification for each sample. Here, “A” stands for “approved”, “NA” stands for “not Approved”, and “IN” stands for “indication of fault”. In total, 65 samples were classified as “approved”, 42 as “not approved”, and 14 samples were classified as “indication of fault”.

From the table, it is clear that we mixed up the types of samples so that all samples with, e.g., holes did not get consecutive numbers. We did this to minimize bias from the welder who performed the labeling of samples at the end of the process, where all samples were lined up in numerical order.

TABLE IV: The table shows an overview of the planned type of samples, the actual type of samples and their classifications. Some numbers are missing as these were not captured correctly.

	Planned		**Actual**			
Sample no	**Type**	**Defect Size**	**Type**	**Defect Size**	**Start time**	**Label**
1	No defect		No defect		08:20	A
2	Paint	3mm	Paint	Thin layer	09:25	A
3	Holes	2mm	Holes	3 mm	11:49	IN
4	No defect		No defect		08:24	A
5	Holes	4mm	Holes	6mm	11:54	NA
6	Holes	6mm	Holes		12:02	NA
7	No defect		No defect		08:29	A
8	Holes	8mm	Holes		12:03	IN
9	No defect		No defect		08:33	A
10	No defect		No defect		08:35	A
11	No defect		No defect		08:37	A
12	No defect		No defect		08:38	A
13	No defect		No defect		08:40	A
14	Holes	10mm	Holes		12:04	NA
15	No defect		No defect		08:41	A
16	No defect		No defect		08:43	A
17	No defect		No defect		08:45	A
18	Holes	12mm	Holes	11 mm	12:05	NA
19	No defect		No defect		08:47	A
20	No defect		No defect		08:48	A
21	No defect		No defect		08:50	A
22	No defect		No defect		08:52	A
23	Paint	6mm	Paint		10:12	IN
24	Paint	9mm	Paint	XA	10:22	IN

25	No defect		No defect		08:56	A
26	No defect		No defect		08:58	A
27	No defect		No defect		08:59	A
28	No defect		No defect		09:01	A
29	No defect		No defect		09:03	A
30	Paint	12mm	Paint	XA	10:01	IN
31			Gas		11:14	NA
32			Gas		11:15	NA
33	Paint	15mm	Paint	XA	09:56	NA
34			Gas		11:17	NA
35			Gas		11:18	NA
36			Gas		11:19	NA
37			2x Holes	12 mm, 15 mm	13:08	NA
38	Holes	14mm	Holes	12 mm	11:52	NA
39			No defect		12:20	A
40			No defect		12:22	A
41			No defect		12:24	A
42	Holes	16mm	Holes	16 mm	12:06	NA
43	Paint	18mm	Paint	XA	09:52	NA
44	Holes	18mm	Holes		12:08	NA
45			No defect		12:25	A
46			No defect		12:26	A
47	Holes	20mm	Holes	21 mm	11:51	NA
48	Paint	21mm	Paint	XA	10:16	NA
49	Holes		2x Holes	14 mm, 14 mm	13:05	NA
50			No defect		12:27	A
51			No defect		12:28	A
52	Holes		2x Holes	11 mm, 10 mm	13:05	NA
56	Paint	24mm	Paint	XA	09:42	NA
58	Holes		2x Holes	14 mm, 11 mm	13:04	NA
60	Paint	27mm	Paint		09:35	A
61	Holes		2x Holes	5 mm, 5 mm	13:02	IN
63	Paint	30mm	Paint	Thin layer	09:29	A
64	Holes		2x Holes	24 mm, 15 mm	13:02	NA
65	Paint		Paint	XA	10:28	IN
66	Paint		Paint	XA	10:25	IN
67	Holes		2x Holes	18 mm, 16 mm	13:03	NA
68	Holes		2x Holes	13 mm, 17 mm	13:00	NA
70			No defect		09:07	A
71	Paint		No defect		12:09	A
72			No defect		09:09	A
73			No defect		09:12	A
74			No defect		09:14	A
75			No defect		09:17	A
76			No defect		09:20	A
77			No defect		09:23	A
78			Paint	XA	10:32	A
79			Paint	XA	10:35	IN
80			Gas		10:54	NA

81			Gas		10:52	NA
82			Gas		10:59	NA
83			Gas		11:02	NA
84	Paint		No defect		12:10	A
85	Holes		2x Holes	13 mm, 26 mm	13:07	NA
86			Gas		11:04	NA
87			Gas		11:06	NA
88			Gas		11:08	NA
89			Gas		11:09	NA
90			Gas		11:10	NA
91	Paint		No defect		12:11	A
92	Paint		No defect		12:12	A
93	Paint		No defect		12:13	A
94			Gas		11:12	NA
95	Paint		No defect		12:14	A
96			Gas		11:13	NA
97	Paint		No defect		12:15	A
98	Holes		2x Holes	18 mm, 17 mm	12:56	NA
99	Paint		No defect		12:16	A
100	Holes		2x Holes	18 mm, 25 mm	12:59	NA
101			No defect		13:14	A
102			No defect		13:15	A
103			No defect		13:16	A
104			No defect	did not start correctly at first	13:22	A
105			Thread		13:23	IN
106			Thread		13:24	IN
107			Thread		13:29	NA
109			Thread		13:40	IN
110			Thread		13:41	NA
111			Thread		13:42	IN
112			Thread		13:43	IN
113			Thread		13:44	NA
114			Thread		13:45	NA
115			No defect		13:46	A
116			No defect		13:47	A
117			No defect		13:48	A
118			No defect		13:49	A
119			No defect		13:50	A
120			No defect		13:51	A
121			No defect		13:51	A
122			No defect		13:53	A
123			No defect		13:54	A
124			No defect		13:55	A
125			No defect		13:56	A
126			No defect		13:57	A
127			No defect		13:57	A
128			No defect		13:58	A
129			No defect		13:59	A

Appendix B: Decision Tree with Outlier Detection

February 18, 2026

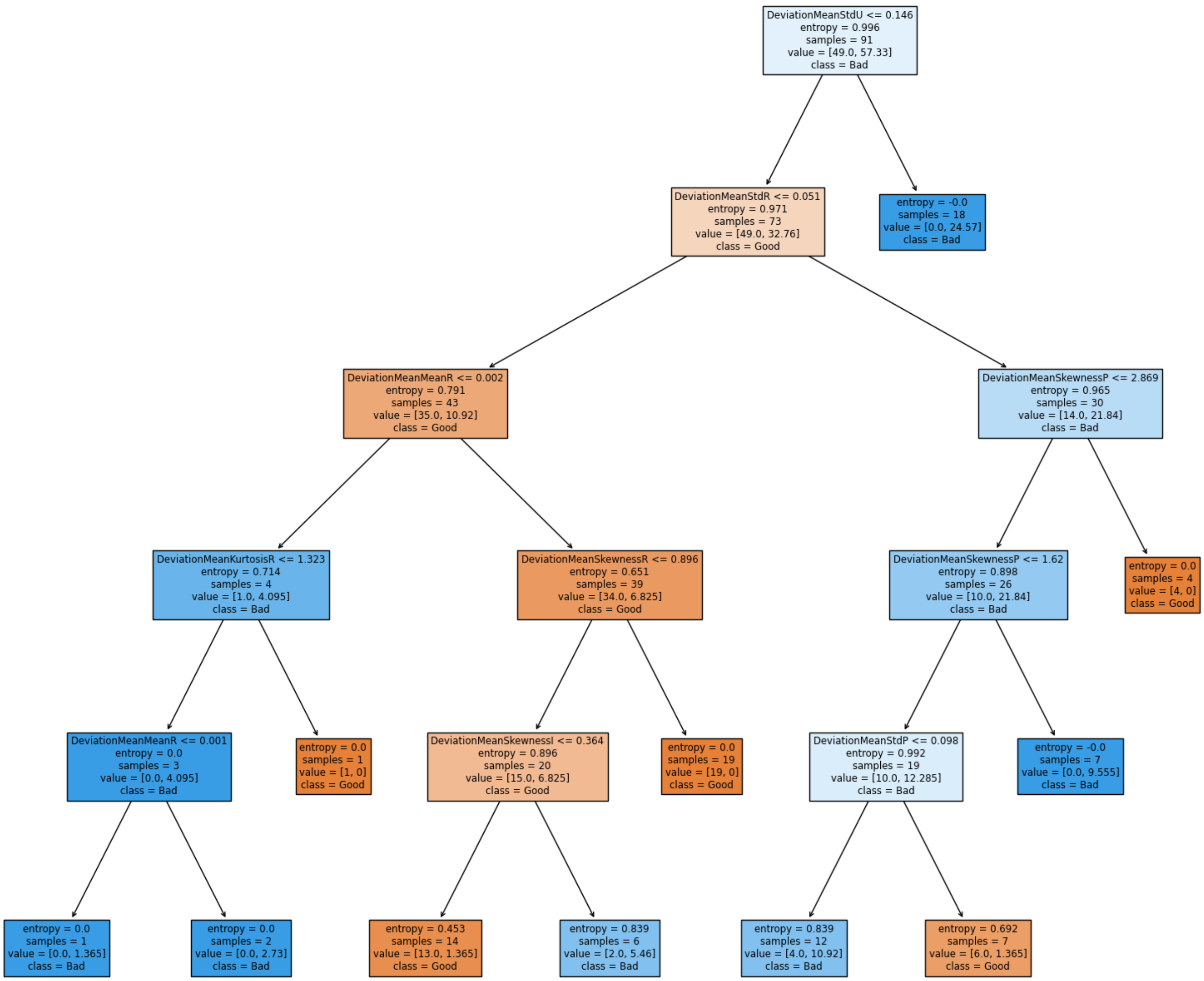

Fig. 6. The figure shows the decision tree when outlier detection is included in the statistical analysis.

Appendix C: Decision Tree without Outlier Detection

February 18, 2026

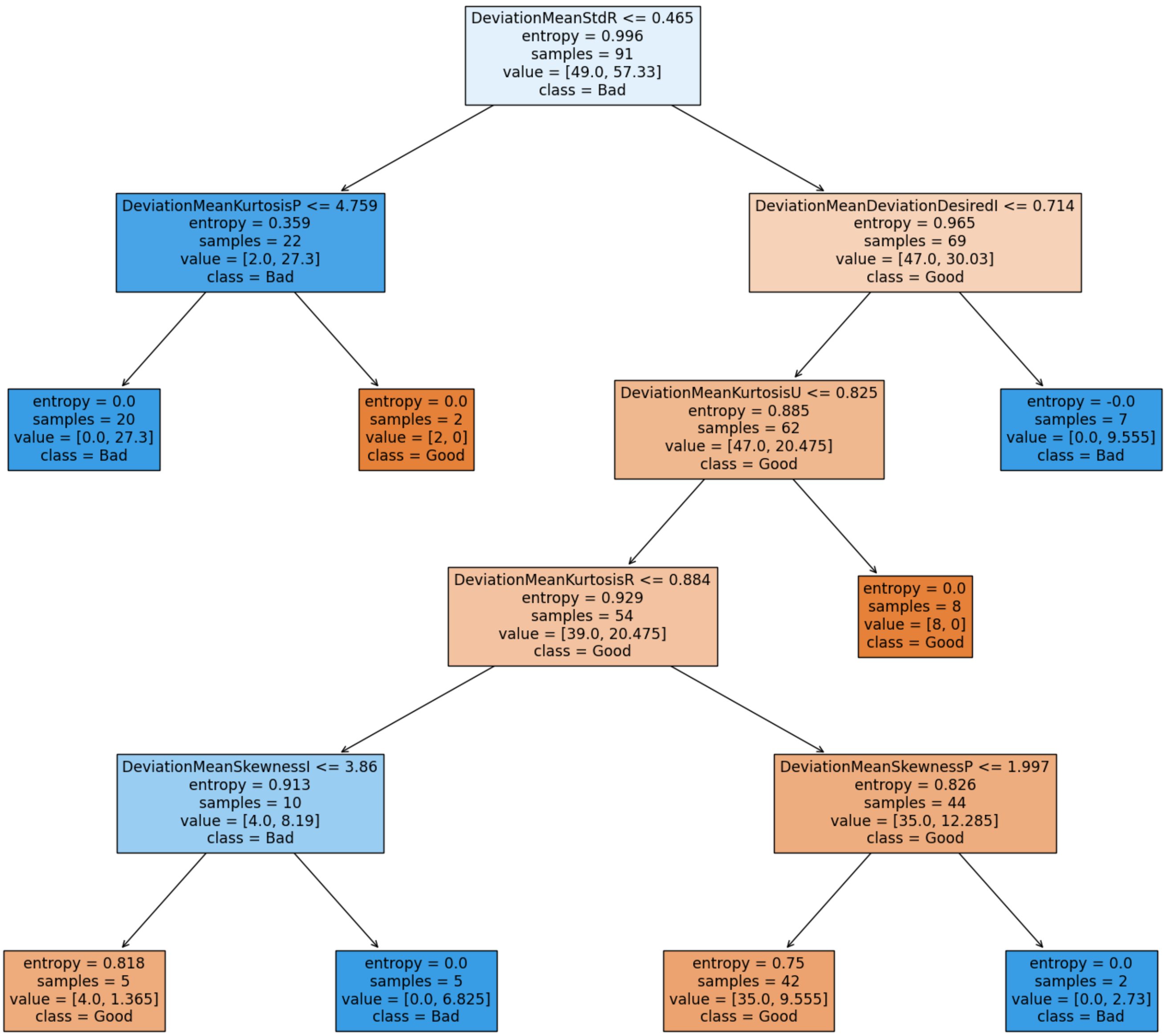

Fig. 7. The figure shows the decision tree when outlier detection is not included in the statistical analysis.