Artificial Intelligence in Production Engineering

Group Report
Group: Predictive Quality Battery

Nalivaika, Jan
03694590
nalivaika@outlook.de

Paździerkiewicz, Przemyslaw
03718188
p.pazdzierkiewicz@tum.de

Pielmeier, Sebastian
03693728
pielmeier.sebastian@t-online.de

Tcvetkov, Nikita
03689859
nikita.tcvetkov@tum.de

Wittner, Simon
03696129
simon.wittner@gmx.net

21.07.2022

Contents

List Of Abbreviation

CWT = Continuous Wavelet Transform
FFT = Fast Fourier Transformation
KDD = Knowledge Discovery in Databases
LBW = Lase Beam Welding
ML = Machine Learning
NN = Neural Network
OOT = Out Of Threshold
RMS = Root mean square
STD = Standard Deviation

List of Figures

1. Introduction, Business Case and Objectives

1.1 Introduction

Operating profitably in the current, highly dynamic market, requires the capability to quickly adapt to increasingly individualized customer demands, strict adherence to deadlines, and expected quality requirements. Failure to fulfill these requirements will result in a decreased competitiveness, a smaller market share and an overall lower profitability. [19, 20]
In the current state of "Industrie 4.0" and Big-Data, multiple opportunities arise to improve speed and accuracy in the production environment. Adaptive process scheduling, for example, can be used to achieve an optimal utilization of machinery and adherence to the production schedule. Both of these effects will benefit the costumer, as the average production cycle time and cost will be reduced. [1, 6, 10]
Quality control is one of the steps of the production chain where data-driven algorithms can support the identification of rejects or uncover possible areas of improvement in the production process. The significant advancements in computer science, especially in Machine Learning (ML), can be transformed and adapted to the specific needs in quality control. Thus, greater accuracy and efficiency in identifying defective products can be achieved than would be possible manually. [14]
Machine learning includes those algorithms that are capable of solving tasks without being explicitly programmed to do so. They are based on pattern recognition and their performance improves as more data is available. This property is especially advantageous as more and more data is provided by the increasingly digitized production environment. [20]

1.2 Business Case

As a consequence of ever-stronger emission regulations by national governments and international organizations, car manufacturers around the world are forced to accelerate their efforts towards more efficient vehicles. For the moment, the majority of these companies have opted to "go electric" and change their source of power from fossil fuels to batteries, which has led to a strong increase in the demand for battery cells. Before being integrated into the vehicle, the individual cells must be combined into a battery pack. In practice, this requires the manufacturing of thousands of electric contacts between the battery terminals and the thin copper foil connecting them, which is achieved by welding them together. The quality of these connections is highly critical for the performance and the safety of the whole battery pack. Already, the failure of one contact can lead to a thermal runaway and consequently the destruction of the pack and potentially the whole vehicle. Hence, being able to repeatedly produce high quality connections is critical for the safety and therefore the economic viability of battery pack manufacturing. [4, 16, 25]

Considering the properties of the battery cells (chemically reactive, temperature sensitive, mostly aluminum connectors) and the copper foil (very low thickness, copper), multiple requirements for the welding process are derived:

- Low heat input
- Low mechanical stress
- Repeatability

Laser-Beam welding (LBW) is a promising method which can satisfy these specific requirements. Its local power input is very high and leads to an almost instantaneous welding of both materials, before a significant heat input into the part takes place. Besides that, no electric current is required, which is beneficial for welding battery conductors. Since LBW is a contactless welding method, the mechanical stresses exerted on the welded parts are minimal. By refocusing the beam to the desired position, a high rate of welds per minute can be achieved. LBW has therefore many benefits for battery pack mass production. For validating the quality of the welded connections, sensors can record data during the welding process. In combination with machine learning algorithms, these sensor readings can serve as a basis to classify if the desired quality is achieved and further increase the process reliability. [8, 9, 17, 18]

1.3 Objectives

This report provides an exemplary use-case for classifying welds in the domain of laser beam welding. This process can easily be applied to any other classification problem, just by adapting a few variables. From a given data set, multiple pre-processing and feature extraction steps are performed. The procedure shown in the following adheres to the general KDD-process (Knowledge Discovery in Databases). Objective 1 is to find a correlation between Signal 1 and Signal 2 from the given data set. The second and the main Objective is to classify the weld seam in the first line in three categories:

- OK
- NOK
- OOT

Afterwards, a classification regarding a possible lubricant contamination is performed:

Alternative 1

- Lubricant
- No Lubricant

Alternative 2

- WD40
- Gleitmo

2. Methods and Experiments

The KDD-process is structured in four steps. First, Data Cleansing is performed to detect data quality issues or remove anomalies. The second step is Data Transformation to select the possible input data. Further, Data Mining is performed to enlarge the database. Finally, the evaluation is performed to categorize the success of a model. In the following sections, the KDD-process is used as a guideline to predict the quality of a weld with the help of a given dataset.[3, 15]

2.1 Data Cleansing and Preprocessing

2.1.1 Dataset Overview

As mentioned in the task description, the dataset contains measurements from two different photodiods recording back-reflected laser radiation during LBW. Additionally, denoised values for signal 1 are also given. Each welding seam is divided into five equidistant sections, whose measurements are individually recorded as a time series in the dataset.
Each signal log is categorized into four different, equally distributed binary categories: not OK, signal value exceeded, WD40 pollution and Gleitmo pollution. It is noticed that all samples for which the signal value is exceeded are uniformly marked as OK. The labels of the lubricants are found to be exclusive, meaning that for no welding seam both WD40 and Gleitmo are present. Out of the total of 1350 measurements, 77 (6%) are marked as "out of threshold", 1073 (79%) as "OK", and 277 (21%) as "not OK". 450 (33%) weld seams contained no pollution. 450 (33%) are marked with WD40 and 450 (33%) with Gleitmo.

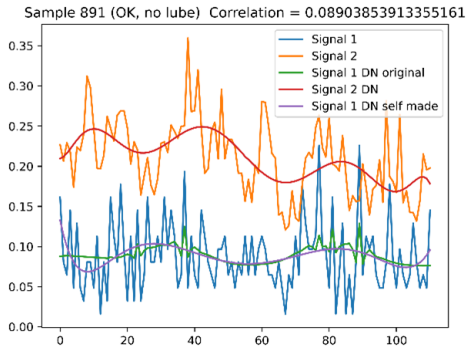
2.1.2 Data Visualization

The figures below show plotted signal values for randomly chosen samples with varying labels. As can be seen in figure 2.1, attempts are made to reproduce the noise reduction procedure used for signal 1 for signal 2. Additionally, the automatically calculated Person correlation coefficients between the values for Signal 1 and 2 are also displayed.

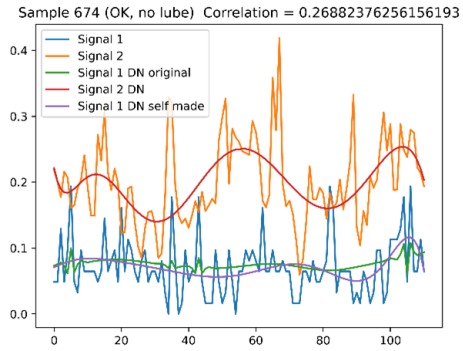
Based on data inspection, it is concluded that signal 2 is found to be typically much noisier than signal 1. Samples marked as "not OK" typically display higher values for all signals; however, no determining boundary distinguishing the two labels is identified. Furthermore, the correlation coefficients between signals carry little information about weld quality. Based on this information alone, it is difficult to predict weld seam quality.

2.1.3 Data Cleansing

After examination of the dataset, an inconsistency regarding the labeling of signals in the "out of threshold" category is identified: multiple samples across the dataset displayed only the value "1" for all signal readings of all the sensors. Although most such samples are labeled as "out of threshold" (OOT), there is a significant number of



(a) Visualization plot for sample 891



(b) Visualization plot for sample 674

Figure 2.1: Visualization of data

logs where this is not the case. As signal values are the only source of input features for the prediction model, such inconsistency poses a significant limit to the achieved accuracy. Thus, a new condition for the labeling of the signals is chosen based on the original set of OOT values and the data is relabeled accordingly. The new condition is set at 13 occurrences of the value "1" in signal 1 logs, which is the minimum among the samples originally labeled as OOT.

2.2 Data Transformation Methodology

Data Transformation aims at extracting meaningful features from the dataset that can be used for training the classification model. This step is especially important while dealing with noisy and unstructured time series data. The choice of features, can greatly affect both the learning speed and the output quality of the model.

2.2.1 Statistical Features

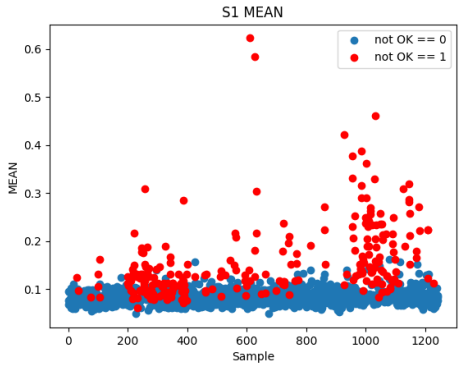
Extracting statistical features allows transforming a time-series into a single characteristic value (feature). For the calculation of each feature, an affiliated method from the pandas, NumPy and SciPy modules are used. The following subsections show several potential features for the classification task. Correlation between Signal 1 and 2 is also shown. Visualization plots for the remaining metrics are provided in the appendix.

Correlation Between Signal 1 and Signal 2

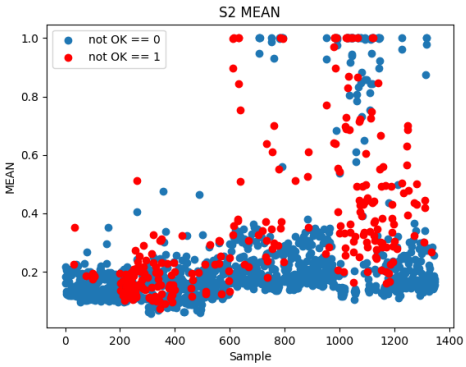
In order to answer the first Objective of the research, correlation for the two signals has to be found out. Computing the Pearson's as well as Spearman's coefficient for all the samples is necessary. Figure 6.5 shows no clear correlation between two Signals since the coefficient values are broadly spread all over the plot lying from -1 to 1.

Mean

One of the most prominent statistical approaches is the arithmetic mean. Figure 2.2 shows the distribution of the mean values of signal one and two across all measurements in the pre-processed dataset, with color-coding between OK/NOK labels.



(a) The mean values of signal 1 across samples



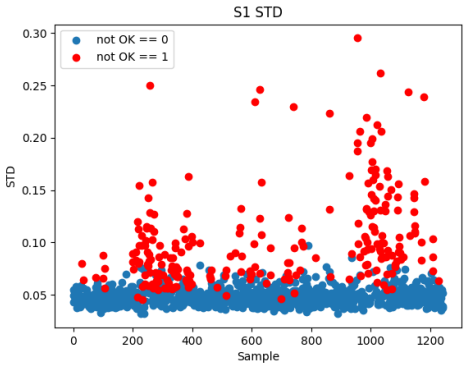
(b) The mean values of signal 2 across samples

Figure 2.2: Mean values

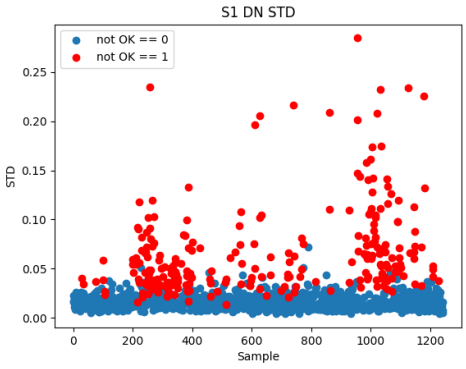
As seen in figure 2.2 (a), for signal 1, the data seems to show a significant correlation between the mean value of the signal and the quality of the weld; poor quality welds tending to correspond to higher mean values. This is especially noticeable across samples affected by pollution with the lubricant Gleitmo (samples 200 through 401 and 950 through 1125). The region, where both OK and NOK are found, is likely to cause inaccuracies in a classification model when no further features are provided. On the other hand, although similar tendencies can be identified on the plot for signal two, the overall distribution is much noisier. Since this trend is prominent for all the examined features, signal two values are rejected as a source of the input features.

Standard deviation

Standard deviation (STD) is a metric of the spread of values within a sample. It helps distinguish stable measurements from highly varying ones. Figure 2.3 shows the distribution of standard deviation values across the dataset.



(a) The Standard deviation of signal 1 across samples



(b) The Standard deviation of denoised signal 1 across samples

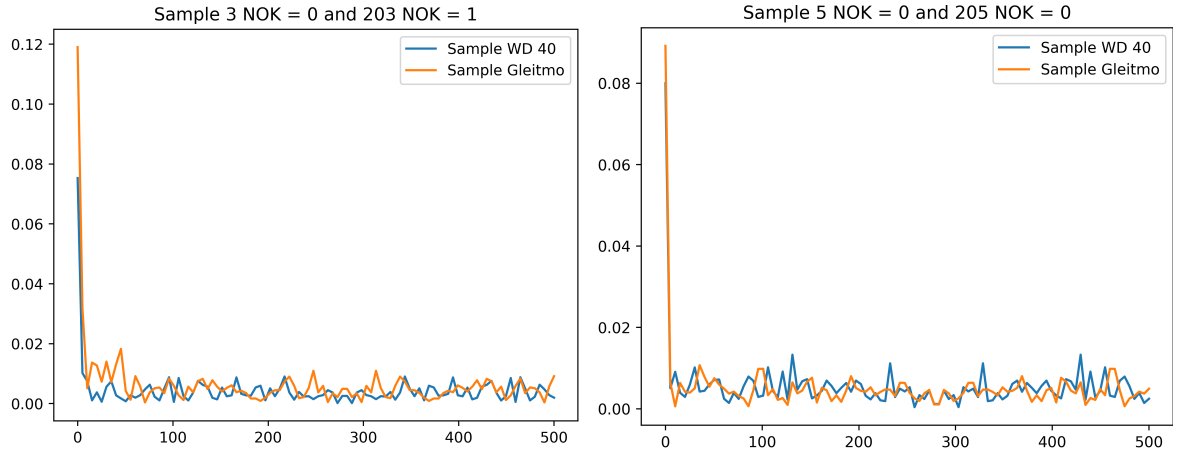
Figure 2.3: Distribution of standard deviation

Based on the results in figure 2.3 the signals standard deviation serves as a better feature for classification than the mean. The boundary between the two clusters is

more protruding. The effect is again augmented by the presence of Gleitmo lubricant. Still, the confusion region, especially for the denoised signal, appears very prominent.

2.2.2 Fast Fourier Transformation

Fast Fourier Transformation (FFT) is an algorithm to transform a signal from a time-space domain into the frequency domain. It's used to analyze stationary systems and its vibrations. To transform the signal, the python scipy.fftpack package with default parameters is used to perform the FFT. Figure 2.4 shows four different samples of signal 1. The signals have multiple peaks all over the spectrum. In figure 2.4a, Samples 3 and 203 are shown. Sample 3 is contaminated with WD40 and has NOK = 0, while Sample 203 is contaminated with Gleitmo and its NOK = 1. From the two FFT's, no obvious differences can be registered. Both samples have peaks at frequency = 0 which can be interpreted as an oscillation with a very large period. In comparison to figure 2.4a, both samples are NOK = 0. Similarly, no huge differences are detected between these classes.[12]



(a) OK and NOK samples

(b) OK samples

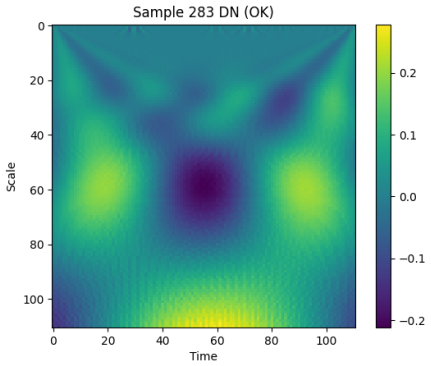
Figure 2.4: FFT of 4 different Samples

2.2.3 Continuous Wavelet Transformation

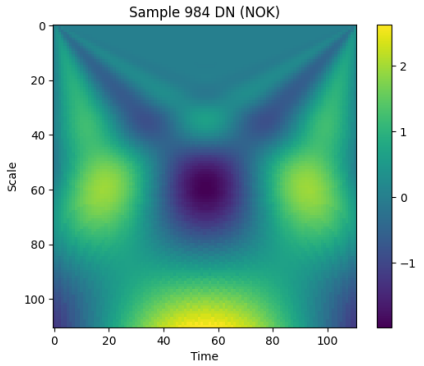
Similarly to FFT, the wavelet transform maps the signal from time to frequency domain; the key difference being that is also provides time resolution of the transformed signal, which makes it more suitable for analyzing non-stationary systems. The transform is carried out with help of the pywavelets module using the Morlet wavelet and varying ranges of scaling factors (up to 1000). The transformed data is subsequently visualized with the intention of identifying any possible underlying patterns. Figure 2.5 shows exemplary plots for two randomly chosen samples with differing labels. As shown in figure 2.5, no obvious correlation between the labeling and the wavelet transform coefficients of the signal can be identified. [23]

2.2.4 Feature Selection Summary

As discussed in this section, there are limits to how effectively signals can be classified based on a single feature. Still, given the visualization process, the mean, STD, and



(a) CWT visualization for sample 283 (Signal 1 denoised)



(b) CWT visualization for sample 984 (Signal 1 denoised)

Figure 2.5: CWT visualization

RMS values appear promising for the task. Moreover, signal one denoised typically provides more consistent metrics than the other sources. Finally, the combination of mean and STD can also be used as an additional source of information for the model. The chosen features can then be compared against the raw signal in making predictions.
Although the mentioned features are studied in the context of the OK/NOK classification, they are equally applied as inputs for the second classification task. The main reason for that is that no feature seems particularly effective in classifying either the presence or the type of the lubricant. Still, since there exists a correlation between the polluted seams and the ones marked as NOK, the inputs may carry significant information for the second classification.

2.3 Classification Models

In accordance with the KDD process, different models for finding patterns in the preprocessed data are considered. The three different models presented in the following chapters are chosen based on publications regarding predictive quality classification in production engineering.

Logistic Regression Model

Logistic Regression is a method for the binary classification of an arbitrary input X_i. It uses the logistic function σ (sigmoid function) to transform the continuous output of a Linear Model to values between 0 and 1. This output is the predicted probability Pr of the input X_i belonging to the class Y: $Pr(Y = 1|X_i)$. The final discrete classification is based on the whether the probability Pr is great or equal/smaller than 0.5. [5, 11]
For fitting the Logistic Regression Model, the regression coefficients of the linear model must be estimated. An ideal fit is achieved by maximizing the likelihood function of the model via iterative gradient-based methods. In the presented case the Limited-memory Broyden-Fletcher-Goldfarb-Shannon (LBFGS) optimizer is used. [11]

Gradient Boosting Method

Gradient boosting is a recent machine learning technique used for regression and classification tasks based on ensemble learning. Multiple weak prediction models are iteratively generated based on the gradient of the model errors. The final strong prediction model is obtained by calculating the weighted average of all weak model. "XGBoost", a gradient boosting method based on decision trees is chosen based on its performance in machine learning competitions and its computational efficiency. [24, 7, 13]
The main hyperparameters of the "XGBoost" gradient boosting model are:

- The number of gradient boosted trees
- The learning rate

Considering the limited computational resources at hand, they are optimized using a randomized non-exhaustive grid search. For the final model the number of gradient boosted trees is set to 600 and the learning rate to 0.1.

Neural Network

Another kind of model which can be used for the classification of data are Neural Networks (NNs). They consist of an input layer, multiple hidden layers and an output layer. The number of neurons in the hidden layers is arbitrary, while in the input and output layer it corresponds with the dimension of the input/output. Depending on the task of the NN (Regression or Classification), the so-called activation function in the output layers is chosen. The activation function determines the mathematical operations performed at each node. For the hidden layers, the choice of the activation function is independent of the application of the NN.
For a NN, numerous hyperparameters must be determined. The main ones are the optimizer, the learning rate, the activation function and the number of hidden layers with their respective number of neurons. [21, 22]
Based on multiple tests and own experiences from previous research, the widely used ADAM optimizer is chosen together with a learning rate of 0.001. For the binary classification in the output layer the sigmoid activation function is selected. As activation functions of the hidden layers, the hyperbolic tangent (tanh) and the Rectified Linear Unit (ReLu) function are considered. Their performance with different NN architectures (1-7 layers, 5 – 100 neurons) is evaluated in multiple tests using a cross validation. The results show that a higher mean prediction accuracy and a lower standard deviation of the results can be achieved using the ReLu function. With more complex NN architectures a shift in the shares of the false positives and negatives is observed which may indicate at least a partial overfit of the model on the dataset. Therefore, a model architecture with only one hidden layer and 75 neurons, using the ReLu activation function in all hidden layers is chosen. [21, 22, 2]

3. Results

In the following chapter the performance of the three different classification models is evaluated. For this comparison the accuracy metric (share of correct predictions) and its spread are considered, as well as the share of false positives and false negatives. Intercomparability between the results of the different models is ensured by using the same ten-fold stratified cross validation for all input/output combinations and models. All the results are obtained using the denoised data from sensor 1. The predictions with this data set proved to be continuously superior or equal to the ones obtained using the other data sets (maybe cite data chapter with plots). Prior to the training, all samples which are 'out of signal threshold' and the ones considered to be incorrectly classified as "within signal threshold" are removed from the data set in accordance with the procedure described in chapter 2.1.3.
Apart from the sample quality, another classification regarding a possible lubricant contamination is realized. Here, 'Alternative 1' (c.f.1.3) is chosen since detecting the sheer presence of lubricant at the weld seam is deemed more important for the product quality than knowing the kind of lubricant.

3.1 Classification of Sample Quality

Table 3.1 shows the mean prediction accuracy and the standard deviation of the prediction accuracy of the three different classification methods for the weld-seam quality label ("OK", "Not OK"). Apart from the raw data, different statistical metrics are considered as input features. Since the gradient boosting model is not compatible with one-dimensional features, it is not possible, to use Mean, STD and RMS as Input in this model. The hyperparameters are set according to chapter 2.3. Regardless of the input, high mean prediction accuracies above 82% are achieved. Additionally, the calculated standard deviations are low ($< 2.5\%$). In contrast to the other methods, the prediction accuracy of the NN improves when using the derived statistical metrics as input.

X	Neural Network	Logistic Regression	Gradient Boosting
Raw	89.39 / 2.18	90.11 / 0.88	92.13 / 3.35
Mean	88.66 / 1.23	81.91 / 0.97	-
STD	92.20 / 1.97	80.63 / 0.86	-
Mean & STD	92.77 / 2.41	83.36 / 1.18	94.30 / 2.21
RMS	90.35 / 1.23	83.12 / 1.01	-

Figure 3.1: Mean prediction accuracy in [%] and standard deviation for the weld-seam quality label (OK, Not OK)

The shares of the false positives (parts incorrectly classified as "OK") and false negatives (parts incorrectly classified as "Not OK") are shown in table 3.2. All methods exhibit a tendency towards incorrectly classifying more samples as "Not OK". This phenomenon is most prevalent for the predictions using the statistical metrics as an input and especially for the Logistic Regression and Gradient boosting methods. In

the case at hand this does however not equal a 100% bias, since a few samples is correctly classified as positive.

X	Neural Network	Logistic Regression	Gradient Boosting
Raw	5.22 / 5.38	1.60 / 8.24	1.87 / 5.35
Mean	0.88 / 10.45	0.00 / 18.15	-
STD	0.64 / 7.16	0.00 / 19.43	-
Mean & STD	1.12 / 6.11	0.00 / 16.69	1.87 / 3.74
RMS	0.80 / 8.84	0.00 /16.94	-

Figure 3.2: Share of false positives and negatives in [%] for the weld-seam quality predictions (OK, Not OK)

3.2 Classification of Lubricant Presence

The results for the prediction accuracy of the presence of lubricant at the weld seam are presented in table 3.3. When using either the Logistic Regression method or the Neural Network, all evaluated metrics (mean prediction accuracy, standard deviation of the prediction accuracy, share of false positives and negatives) remain unchanged, regardless of the input data used. The reason for this behavior is that all inputs are incorrectly assigned to the same class ("Lubricant present"). The obtained mean prediction accuracy of 64.43% equals the share of samples with a lubricant contamination. The Gradient boosting method does not assign all inputs to the same class. A low prediction accuracy of around 60% is obtained. Table 3.4 shows that for both inputs ("Raw" and "Mean & STD") the distribution of false positives and false negatives is approximately identical.

X	Neural Network	Logistic Regression	Gradient Boosting
Raw	64.79 / 0.29	64.79 / 0.29	60.13 / 2.64
Mean	64.79 / 0.29	64.79 / 0.29	-
STD	64.79 / 0.29	64.79 / 0.29	-
Mean & STD	64.79 / 0.29	64.79 / 0.29	59.82 / 4.21
RMS	64.79 / 0.29	64.79 / 0.29	-

Figure 3.3: Prediction accuracy in [%] and standard deviation of the presence of lubricant at the weld seam

X	Neural Network	Logistic Regression	Gradient Boosting
Raw	35.21 / 0.00	35.21 / 0.00	24.06 / 15.78
Mean	35.21 / 0.00	35.21 / 0.00	-
STD	35.21 / 0.00	35.21 / 0.00	-
Mean & STD	35.21 / 0.00	35.21 / 0.00	29.68 / 19.52
RMS	35.21 / 0.00	35.21 / 0.00	-

Figure 3.4: Share of false positives and negatives in [%] for the presence of lubricant

4. Discussion

In the following chapter the different methods and procedures are evaluated based on the obtained results. Furthermore, their applicability and limitations are discussed.

4.1 Data Preparation and Selection

Prior to the modelling, the available data is thoroughly analyzed and pre-processed. Most notably, additional signal samples are relabeled as 'out of threshold' in accordance with the procedure shown in chapter 2.1.3. All samples labeled as 'out of threshold' are removed from the data set. For the Logistic Regression model this data pre-processing decreases the standard deviation of the result recorded during the ten-fold cross validation. The accuracy and the shares of false positives and negatives remain nearly unchanged. This is partially in contrast to the behavior of the NN. Here, the shares of false positives and negatives are more equal when using the pre-processed data set, while the recorded standard deviation and mean accuracy vary only slightly. This data cleaning step has less influence on the final result than e.g. the choice of the sensor (1 vs 2). From a data quality standpoint however, it must be assumed that these signal series which are already labeled as 'out of threshold' or are incorrectly labeled as 'within threshold' will lead to an overall worse data quality and hence less representative prediction. For the data of sensor 1 the number of data series already labeled as 'out of threshold' is 77. Another 31 are determined to be incorrectly labeled as 'within threshold'. As a consequence, about 8% of the original data set are not available for training and testing. However, achieving representative, repeatable predictions thanks to a good data quality is deemed more important than input data size. It is therefore proposed to consequently remove these inputs.

As mentioned in chapter 3, only the denoised data of sensor 1 is used for the evaluation of the different classification methods. This choice is based on the observation that the highest prediction quality is obtained with this data set. Based on the correlation coefficients shown in chapter 6, there is no distinct correlation between the readings of both sensors. However, the signals display a significant redundancy in other statistical features as shown in in figure 2.2 Consequently, using both sensor signals as an input in an attempt to increase the data sample size has shown to lower the overall prediction accuracy, compared to the results using only sensor 1. Since both sensors use a similar photo diode to detect the emissions from the same process there is an inherent risk for redundancy. Disregarding the nosier, redundant data has proven to be beneficial both for the prediction accuracy as well as the required computational time.

4.2 Sample Quality Prediction

The pre-processed data set used for the classification of the sample quality contains four times as many instances with the quality label 'OK', as it contains instances with the quality label 'Not OK'. This imbalance of the data is inherent to the process. Since LBW is a well-known, albeit relatively new, manufacturing technique, it is expected that the majority of the samples manufactured does meet the set quality requirements.

The same over-representation of samples with a positive label does not occur in the incorrect classifications of the models. All models show a tendency of varying strength to over-classify samples as negatives. The relations between false positives and false negatives range between 0.96 and 0.084. No clear explanation is found for this behavior. However, from a practical standpoint the recorded outcome may be desirable to some extent, depending on the application: Defect weld seams which are incorrectly classified as 'OK' (= false positives) are more undesirable and will reduce the trustworthiness of all predictions. In the long term this may even lead to the failure of the practical implementation of a machine learning algorithm. With all the models presented this is avoided. On the other hand, more samples are incorrectly rejected, which means that good quality products are disposed of. Considering the possible business case presented in 1.2, the cost of one battery cell is negligible, compared to the cost and the possible consequences of the failure of a whole battery pack. Due to the accumulation of incorrectly disposed units over several years of production, this will however effect the overall profitability of the process.

The prediction accuracy for the part quality achieved with all three models is in the same range or above the accuracy documented for similar classification task. The same applies for the standard deviation of the results which is found to be equal or lower than the results recorded in previous research. [18]
Using different randomized train-test split of the data (80/20 and 70/30 ratio) yields results in the same ranges as obtained with the ten-fold cross validation. This indicates that the proposed model-feature combinations are robust and not overly sensitive to varying sample sizes.

Apart from the accuracy metrics shown in the result chapter 3 the required computational resources for the predictions with different model-feature combinations must be considered. Ideally, the predictions are carried out in-situ or near in-situ to avoid a prolonged cycle-time. The required times for predicting the test output using the raw signals as inputs range between $0.001\,s$ (Logistic Regression) and $0.002\,s$ (Neural Network) (CPU: i7-7500U, 4 Processors, 2.70 GHz; RAM: 8 GB). While the absolute differences between the calculation times are small, these may be significantly more pronounced when deploying the model on the programmable logic controller of a production machine.

In consideration of the above conclusions drawn from the results, case-dependent recommendations for model-feature combinations are given: In the case of highly limited computational resources, the combination of the Logistic Regression model and raw data inputs will provide results with a high prediction accuracy ($\approx 90\%$) and a very low standard deviation (0.88), while providing the shortest possible prediction time. This could be a viable solution for older, already running LBW machines with comparatively outdated electronics.
The Gradient boosting method together with the mean and standard deviation of the recorded signal as input features will give the highest accuracy. This combination is most beneficial for applications with a high individual production unit cost, e.g. limited production runs of battery packs with an experimental cell chemistry, since the smallest share of samples is incorrectly classified as 'Not OK'. This will reduce the number of functional, high cost units disposed of.
Lastly, the Neural Network together with the standard deviation of the signal as an

input feature should be considered for high volume productions of units with low individual costs. This combination provides a high prediction accuracy of $\approx 92\%$, while the share of false positives is minimal (0.64%). The predictions can be characterized as more conservative, since they rather classify a sample as 'Not OK' than 'OK' and thereby minimize the number of faulty units leaving the production machine. A possible application are LBW machines for manufacturing of standardized battery packs used in in electric scooters, electric bicycles, laptop batteries etc..

Regardless of the use case and the model-feature combination, a second line of quality control is suggested, since even the highest prediction accuracy of 94% is deemed not sufficient for the production of products which can pose a significant risk in case of a malfunction.

4.3 Lubricant Presence

The predictions for the presence of lubricant at the weld seam presented in chapter 3.2 show, that regardless of the combination of model and input features, no meaningful predictions are achieved. Both for the Logistic Regression method and the NN a 100% bias towards classifying all inputs as positives ('lubricant contamination') is detected. The Gradient boosting method does not exhibit this complete bias but only achieves a mean accuracy of around 60%. Performing a principal component analysis on the data to reduce the data dimensionality and extract further information also yields no improvements.

Considering the multitude of different approaches and features tested, it is concluded that the sensor data available does not include a sufficient amount of information on a possible lubricant contamination of the weld. With the present data recording setup it is hence not possible to reliably correlate the sensor data with the label 'Lubricant'.

5. Summary and Outlook

5.1 Summary

This report illustrates an exemplary workflow for the prediction of product quality based on the data gather from a manufacturing process. Specifically the quality of weld seams produced by a LBW machine are analyzed.
After obtaining an initial overview of the provided dataset, the data is cleansed. Hereby, all data samples considered to be above the maximum signal value detectable by the sensors used for data collection are removed.
As a part of the following data transformation step, several statistical features are extracted from the dataset and their meaningfulness is evaluated with the help of different visualizations. The best visible separation of data points is obtained with the mean, the standard deviation and the root mean square values. They are therefore selected as input features, along with the raw signals. Neither FFT, nor CWT uncover any obvious signal characteristics. Similarly, no clear correlation is found using the Pearson's and Spearman's Correlation coefficient.
A Logistic Regression model, the Gradient boosting method and a Neural Network are considered for the classification of the data. For the classification of the sample quality all models achieve a mean accuracy above 89%. The highest accuracy (94%) is obtained using the Gradient boost method and both MEAN and STD as input features, while the combination of the Logistic Regression and the raw data signals as yields the smallest STD of the results (0.88%). All the models, regardless of input, show a bias of varying strength towards predicting more false negatives than false positives.
Regarding the classification of a possible lubricant presence of the sample, it is found that no meaningful predictions are possible, regardless of the combination of model and input features. All attempts result in either a 100% bias of the prediction or a low mean accuracy (60%).

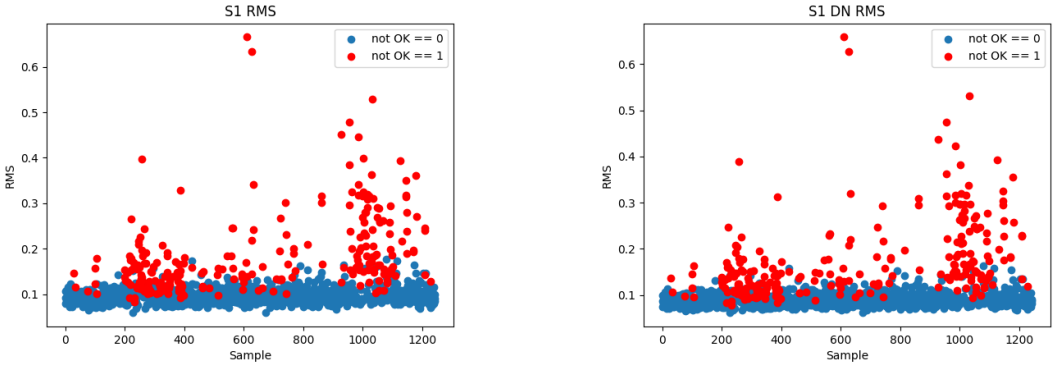
5.2 Outlook

The work with databases and data-driven algorithms requires significant efforts to capture and understand all details and peculiarities of a specific application. Covering all possible options which can be applied in a certain case is not possible within the scope of this work. Considering the limited analysis presented in this report, there is still a noteworthy potential for improving the obtained results. This concerns in particular the selection of the ideal mix of model input features, the selection of the classification model itself and the optimization of hyperparameters like the learning rate. Regarding the data capture process more capable and/or additional sensors should be evaluated and the size of the dataset increased. Furthermore, analyzing the performance of the classifier with changed process parameters like laser beam power or welding different materials could give deeper insights regarding the flexibility and capabilities of the proposed models. In summary, there is a huge variety of options that can be explored and parameters that can be studied in depth to validate the performance of the models. Hence, further research is necessary to create a fully generalizable model.

6. Appendix

6.1 Visualization Plots for Statistical Features

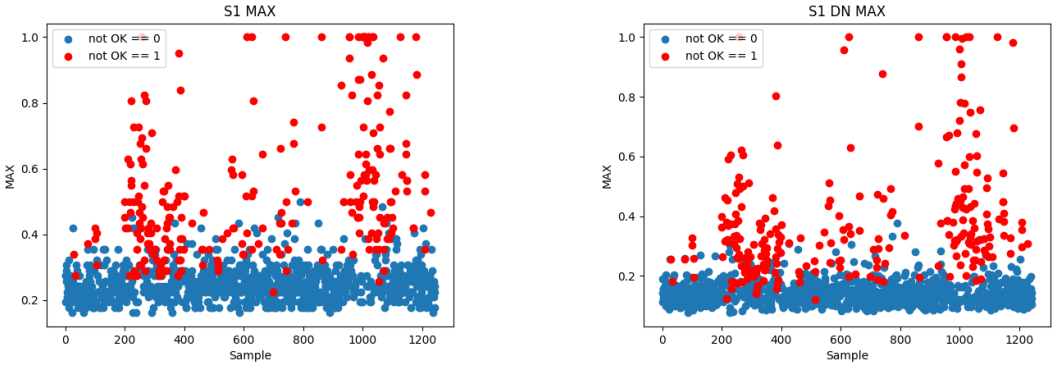
The figures below show the distributions of further examined statistical features across the dataset for selected signals. Some other investigated features include the percentile values (Q25, Q75) or the median. In many cases little statistical significance is found between them and the labels.



(a) The RMS values of signal 1 across samples

(b) The RMS values of denoised signal 1 across samples

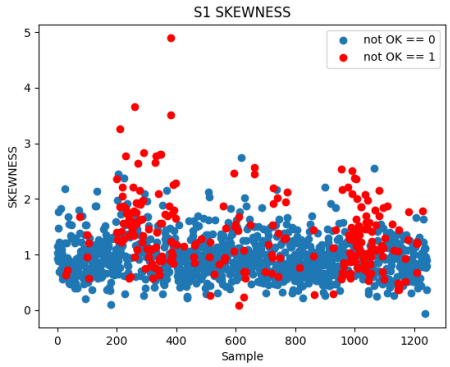
Figure 6.1: RMS values



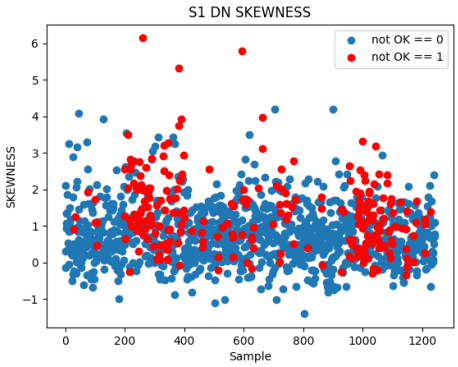
(a) The maximum values of signal 1 across samples

(b) The maximum values of denoised signal 1 across samples

Figure 6.2: Maximum values

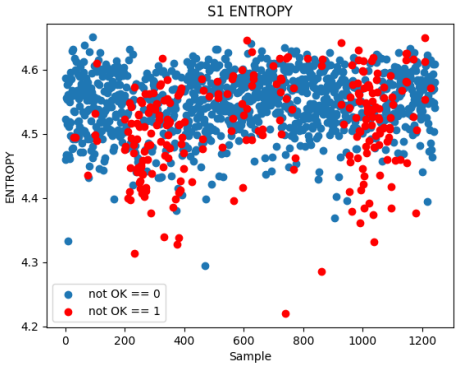


(a) The Skewness of signal 1 across samples

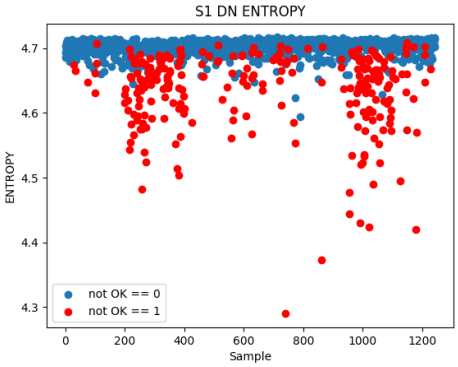


(b) The Skewness of denoised signal 1 across samples

Figure 6.3: Skewness

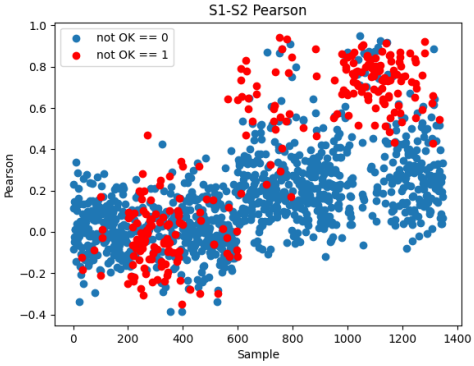


(a) The Entropy of signal 1 across samples

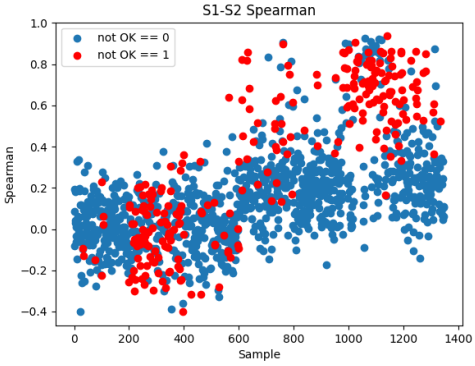


(b) The Entropy of denoised signal 1 across samples

Figure 6.4: Entropy



(a) The Pearson correlation coefficients for signal one and two across samples



(b) The Spearman correlation coefficients for signal one and two across samples

Figure 6.5: Pearson and Spearman's correlation coefficients

Bibliography

[1] *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2016.

[2] Oludare Isaac Abiodun et al. "State-of-the-art in artificial neural network applications: A survey". In: *Heliyon* 4.11 (2018), e00938. ISSN: 2405-8440. DOI: 10.1016/j.heliyon.2018.e00938. URL: https://www.sciencedirect.com/science/article/pii/s2405844018332067.

[3] FH-Bielefeld-MIF-SW-Engineering-2017. *Knowledge Discovery in Databases · Klausurthemen*. last visited on: 20.07.2022. URL: https://fh-bielefeld-mif-sw-engineerin.gitbooks.io/klausurthemen/content/ai/knowledge-discovery-in-databases.html.

[4] Martin J. Brand et al. "Welding techniques for battery cells and resulting electrical contact resistances". In: *Journal of Energy Storage* 1 (2015), pp. 7–14. ISSN: 2352-152X. DOI: 10.1016/j.est.2015.04.001. URL: www.sciencedirect.com/science/article/pii/s2352152x15000055.

[5] Jason Brownlee. "Logistic Regression for Machine Learning". In: *Machine Learning Mastery* (31/03/2016). URL: https://machinelearningmastery.com/logistic-regression-for-machine-learning/.

[6] Rainer Drath and Alexander Horch. "Industrie 4.0: Hit or Hype? [Industry Forum]". In: *IEEE Industrial Electronics Magazine* 8.2 (2014), pp. 56–58. ISSN: 1932-4529. DOI: 10.1109/mie.2014.2312079.

[7] GitHub. *dmlc/xgboost: Scalable, Portable and Distributed Gradient Boosting (GBDT, GBRT or GBM) Library, for Python, R, Java, Scala, C++ and more. Runs on single machine, Hadoop, Spark, Dask, Flink and DataFlow*. last visited on: 20.07.2022. URL: https://github.com/dmlc/xgboost.

[8] Sophie Grabmann, Lazar Tomcic, and Michael F. Zaeh. "Laser beam welding of copper foil stacks using a green high power disk laser". In: *Procedia CIRP* 94 (2020), pp. 582–586. ISSN: 22128271. DOI: 10.1016/j.procir.2020.09.080.

[9] E. Haddad et al. "Laser micro welding with fiber lasers for battery and fuel cell based electromobility". In: *Journal of Advanced Joining Processes* 5 (2022), p. 100085. ISSN: 26663309. DOI: 10.1016/j.jajp.2021.100085.

[10] Mario Hermann, Tobias Pentek, and Boris Otto. "Design Principles for Industrie 4.0 Scenarios". In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2016. DOI: 10.1109/hicss.2016.488.

[11] Gareth James et al. *An Introduction to Statistical Learning: With Applications in R*. 2nd ed. 2021. Springer eBook Collection. New York, NY: Springer US and Imprint Springer, 2021. ISBN: 978-1-0716-1418-1. DOI: 10.1007/978-1-0716-1418-1.

[12] Thomas Kuttner. *Praxiswissen Schwingungsmesstechnik*. Wiesbaden: Springer Fachmedien Wiesbaden, 2015. ISBN: 978-3-658-04637-8. DOI: 10.1007/978-3-658-04638-5.

[13] NVIDIA Data Science Glossary. *What is XGBoost?* last visited on: 20.07.2022. URL: https://www.nvidia.com/en-us/glossary/data-science/xgboost/.

[14] *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things*. IEEE, 2015.

[15] Prof. Dr. Richard Lackes. "Definition: Knowledge Discovery in Databases (KDD)". In: *Springer Fachmedien Wiesbaden GmbH* (last visited on: 20.07.2022). URL: wirtschaftslexikon.gabler.de/definition/knowledge-discovery-databases-kdd-40959.

[16] Philipp A. Schmidt and Michael F. Zaeh. "Laser beam welding of electrical contacts of lithium-ion batteries for electric- and hybrid-electric vehicles". In: *Production Engineering* 9.5-6 (2015), pp. 593–599. ISSN: 1863-7353. DOI: 10.1007/s11740-015-0637-4. URL: https://link.springer.com/article/10.1007/s11740-015-0637-4.

[17] Christian Stadter et al. "Correlation analysis between the beam propagation and the vapor capillary geometry by machine learning". In: *Procedia CIRP* 94 (2020), pp. 742–747. ISSN: 22128271. DOI: 10.1016/j.procir.2020.09.138.

[18] Christian Stadter et al. "Real-time prediction of quality characteristics in laser beam welding using optical coherence tomography and machine learning". In: *Journal of Laser Applications* 32.2 (2020), p. 022046. ISSN: 1042-346X. DOI: 10.2351/7.0000077.

[19] Jiafu Wan, Hu Cai, and Keliang Zhou. "Industrie 4.0: Enabling technologies". In: *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things*. IEEE, 2015. DOI: 10.1109/icaiot.2015.7111555.

[20] Peng Wang, Yiran Yang, and Narges Shayesteh Moghaddam. "Process modeling in laser powder bed fusion towards defect detection and quality control via machine learning: The state-of-the-art and research challenges". In: *Journal of Manufacturing Processes* 73 (2022), pp. 961–984. ISSN: 1526-6125. DOI: 10.1016/j.jmapro.2021.11.037. URL: https://www.sciencedirect.com/science/article/pii/s1526612521008446.

[21] Sun-Chong Wang. "Artificial Neural Network". In: *Interdisciplinary Computing in Java Programming*. Ed. by Sun-Chong Wang. Springer eBook Collection. Boston, MA: Springer, 2003, pp. 81–100. ISBN: 978-1-4613-5046-0. DOI: 10.1007/978-1-4615-0377-4{\textunderscore}5.

[22] Sun-Chong Wang, ed. *Interdisciplinary Computing in Java Programming*. Vol. 743. Springer eBook Collection. Boston, MA: Springer, 2003. ISBN: 978-1-4613-5046-0. DOI: 10.1007/978-1-4615-0377-4.

[23] weisang.com. *Kontinuierliche Wavelet Transformation (CWT) - weisang.com*. last visited on: 20.07.2022. URL: https://www.weisang.com/dokumentation/timefreqspectrumalgorithmscwt_de/.

[24] *XGBoost*. last visited on: 20.07.2022. URL: https://www.kaggle.com/code/dansbecker/xgboost/notebook.

[25] M.F.R. Zwicker et al. "Automotive battery pack manufacturing – a review of battery to tab joining". In: *Journal of Advanced Joining Processes* 1 (2020), p. 100017. ISSN: 26663309. DOI: 10.1016/j.jajp.2020.100017. URL: https://www.sciencedirect.com/science/article/pii/s2666330920300157.