\section(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).   
Additionally, all samples considered to be incorrectly classified as “within signal threshold” are removed in accordance with the procedure described in chapter X.X.

\subsection(Classification of sample quality (OK, Not OK) )

Table X.X 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. The hyperparameters are set according to chapter X.X.

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 | - |

The shares of the false positives (parts incorrectly classified as “OK”) and false negatives (parts incorrectly classified as “Not OK”) are shown in table X.X. All methods exhibit a bias towards incorrectly classifying 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.

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | Neural Network | Logistic Regression Model | Gradient boosting method |
| 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 | - |

\subsection(Classification of lubricant (WD40, Gleitmo, no lubricant) )

The results for the prediction accuracy of the presence of WD40 lubricant at the weld seam are presented in table X.X. Regardless of the model and the input data, all evaluated metrics (mean prediction accuracy, standard deviation of the prediction, share of false positives and negatives) remain unchanged. The reason for this outcome is that all inputs are assigned to the same class (“No WD40 present”).  
The inverse case is observed when trying to predict the presence of lubricant, not considering its exact designation. Here, all inputs are assigned to the class “Lubricant present”, again regardless of the input (mean prediction accuracy: 64.79%, standard deviation of the results: 0.29).

The predictions for the presence of Gleitmo are slightly better when using a NN (mean prediction accuracy: > 71%, standard deviation of the results: < 3.14), albeit also with a distinct tendency for false negatives (share of false positives > 2.17%, share of false negatives < 26.45%)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Tables for Lubricant

|  |  |  |  |
| --- | --- | --- | --- |
| **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 | - |

|  |  |  |  |
| --- | --- | --- | --- |
| **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 | - |