**NN**

Another kind of model which can be used for the classification of data are Neural Networks (NNs), also referred to as Multilayer Perceptrons. They are composed of an input layer, multiple hidden layers and an output layer. The number of neurons, also called nodes, 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 neural network (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.

The main advantages of Neural Networks are their universality, which is a result of the nonlinearities in the neurons and their ability to efficiently approximate functions in high-dimensional spaces. These properties are especially useful when considering complex datasets.

For a NN, a large number of hyperparameters must be determined. The main ones are:

* Architecture of the NN (number of hidden layers and respective number of neurons)
* Activation function
* Optimizer and Learning rate

Based on engineering practices presented in the courses “AI in Production Engineering” and “Physics-informed machine learning”, as well as the paper “Real-time prediction of…”, the widely used ADAM optimizer is chosen together with a learning rate of 0.001.

Since a binary classification must be conducted, the “Sigmoid” activation function is used in the output layer.

For the activation function of the hidden layers the hyperbolic tangent (tanh) and the Rectified Linear Unit (ReLu) function are considered. For comparison multiple tests are conducted using the 10-fold stratified cross validation as described in XX. Additionally, different NN architectures (1-7 layers, 5 – 100 neurons) are evaluated.

The test results show that especially with bigger NNs, a higher mean prediction accuracy can be achieved using ReLu (~ 93%), while the accuracy of the networks using the tanh function reaches its maximum accuracy of around 91% already with two hidden layers and then stagnates.

The variance of the prediction accuracy ranges between ~ 1.2% and ~4.1%. In average, it is around 1% higher when using ReLu.

For the NN architectures using only one hidden layer, the shares of false positives and false negatives are similar (~ 5.8% vs ~ 4.8%). However, already with three hidden layers, the share of false negatives decreases to around 2.8%, while the share of false positives increases to around 6.2%. Especially with higher numbers of hidden layers this effect is more pronounced when using the ReLu activation function. This indicates at least a partial overfit of the model on the dataset.

Considering the planned application of the model, an emphasis is put on the robustness and reliability of the expected prediction. Therefore, a model architecture with only one hidden layer and 75 neurons, using the tanh activation function is chosen. The expected prediction accuracy is 89%, which is around four percent lower than the results obtained with more complex models. However, the expected very low variance of the results (1.33%) and the balance between the shares of false negatives and false positives (6.1% vs 4.9%) ensure the reliability of the prediction and minimize the risk of possible overfitting. Especially in production engineering, a bias in the predictions introduced by overfitting can have detrimental effects on the practical usability of a model and may nullify its practical use in the long term.

**Results**

Table X.X shows the prediction accuracy for different labels y using the NN with the hyperparameters described in X.X. All predictions were done using the raw, but denoised data of sensor 1 or derived statistical metrics as the input of the NN as this gave the highest accuracies, regardless of the data label.

It is not possible to correctly predict for the label ‘WD40’ and ‘Lubricant’. In both cases the NN classifies all outputs the same. This problem persists, regardless of the input.  
The predictions for the label ‘Gleitmo’ are slightly better, albeit also with a strong tendency for false negatives.

In summary, correctly predicting the presence or the kind of lubricant at the weld seam is not possible with the available data sets and used statistical metrics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **y** | | | |
| **X** | Not OK | WD40 | Gleitmo | Lubricant |
| Raw | 88.99 | 64.63 | 71.30 | 63.66 |
| Mean | 89.31 | 65.43 | 72.51 | 64.79 |
| STD | 92.76 | 65.43 | 75.40 | 64.79 |
| Mean + STD | 92.93 | 65.43 | 74.67 | 64.79 |
| RMS | 90.91 | 65.43 | 73.31 | 64.79 |