


# Pseudolabeling Machine Learning Algorithm for Predictive Maintenance of Relays

FABIAN WINKEL <sup>1</sup>, OLIVER WALLSCHEID <sup>2</sup> (Senior Member, IEEE), PETER SCHOLZ <sup>1</sup>,  
AND JOACHIM BÖCKER <sup>2</sup> (Senior Member, IEEE)

<sup>1</sup>Technology Development, Phoenix Contact Electronics, 31812 Bad Pyrmont, Germany

<sup>2</sup>Department of Power Electronics and Electrical Drives, Paderborn University, 33098 Paderborn, Germany

CORRESPONDING AUTHOR: FABIAN WINKEL (e-mail: fwinkel@phoenixcontact.com)

**ABSTRACT** Predictive maintenance (PdM) has become an important industrial feature. Existing methods mainly focus on remaining useful life (RUL) regression or anomaly detection to achieve PdM in a given application. Those approaches assume monotonic degradation processes leading to a single catastrophic failure at the system's end of lifetime. In contrast, much more complex degradation processes can be found in real-world applications, which are characterized by effects like self-healing or noncatastrophic anomalies. A important example of devices with complex degradation are electromechanical relays. As established PdM solutions failed when applied to a real-world relays degradation data set, the maintenance algorithm for unlabeled data (MAUD) is presented to detect signs of wear and enable a service in time. In detail, MAUD is based on an artificial neural network (ANN), which is trained semisupervised. Experiments with measurement data from 546 relays show that MAUD is superior to various existing methods: The static B10 threshold, which represents the state of the art in relay maintenance, is surpassed by a 17.07 p.p. increase in utilization while reducing failures by 6.42 p.p. Methods based on machine learning, such as RUL estimation and anomaly detection, achieved much lower utilization (up to 31.83 p.p.) compared with MAUD while maintaining the same failure rate.

**INDEX TERMS** Electromechanical relay, artificial neural network (ANN), predictive maintenance (PdM), failure risk, pseudolabeling.

## I. INTRODUCTION

### A. MOTIVATION

Predictive maintenance (PdM) is a key technology enabling the achievement of economic and ecological goals at the same time. Due to the increasing digitalization and major advances in machine learning (ML), ML implementations of PdM gain increasing attention.

In this article, a PdM approach for electromechanical relays is presented. There is a high demand for relays as they are used for electrical switching operations in a wide range of applications, which is reflected in an annual market volume for electromechanical relays of over 6 billion US dollars [1]. The failure of a single relay can lead to the failure of an entire plant. This is why PdM is so important for relays.

But unfortunately, the current state of the art regarding PdM is not applicable to relays, as they exhibit complex degradation mechanisms: In part, the degradation is reversible, which

is why the estimation of the monotonic remaining useful lifetime (RUL) with supervised ML methods is not successful here [2]. Furthermore, relays are very heterogeneous in their characteristics, which causes anomaly detection with unsupervised ML methods to fail [2].

Therefore, a semisupervised ML-based PdM method for relays is presented in this article. Specifically, a pseudolabeling of the dataset is performed to distinguish between conspicuous and inconspicuous operating states. In this way, a need for maintenance can be indicated in time.

### B. STATE OF THE ART: PDM

The approaches for PdM are strongly dependent on the particular component to be analyzed and the associated (measurement) data. Therefore, in the following, the state of the art for PdM is differentiated into two groups: For one group, monotonic wear is evident in the data, and therefore RUL can

be estimated with this. In the other group, sudden changes can be observed, which is why anomaly detection is applied here. To estimate the RUL, a method of ML is typically used to learn a mapping of the input data to the RUL. Consequently, it is a supervised procedure that requires the RUL to be known. The definition of the RUL is dependent on the component in question, one possibility being the remaining operating time to failure. In the context of relays, the remaining switching cycles until failure would be used, since the degradation takes place significantly in the switching process.

The NASA turbofan engine degradation simulation dataset [3] is particularly commonly used for RUL estimation. It contains simulated run-to-failure data from several hundred turbofan engines. Furthermore, a lifetime is given so that the RUL can be estimated. Various ML techniques have already been applied for this purpose: An echo state network for RUL prediction is used in [4]. Moreover, in [5], recurrent neural networks were used to consider the history of parameters for the prediction of RUL and to filter outliers. Another frequently used dataset is the FEMTO bearing dataset [6], which consists of 17 run-to-failure datasets of bearings. Different ML approaches are developed for this dataset, in [7] and [8] support vector machines are used to estimate the RUL, in [9] a summation wavelet-extreme learning machine is used to perform estimation/prediction tasks.

The estimation of RUL is not possible for all datasets. Therefore, there is an area of research that deals with anomaly detection assuming that a single catastrophic failure event can be predicted by a prior operation anomaly [10]. Recently, significant progress has been made in this area through the use of autoencoders (AE). AE is an unsupervised ML approach in which typically two convolutional neural networks (CNN) are trained. One CNN is used to transform the input data into a low-dimensional latent space and is therefore called the encoder. The other CNN has the opposite task of reconstructing the input data from the encoder output data, hence the name decoder. The AE learns to encode and decode the normal operating states during training and is accordingly good at it. Unknown states, i.e., anomalies, can thus be detected by an unusually high reconstruction error. This approach allows unsupervised learning, no labels are necessary.

An example of a successful anomaly detection with an AE for PdM is given in [11], where a model was trained with correct working machine data of production press machine. Further information on PdM fundamentals and applications can be found in survey papers, e.g., [12], [13], [14].

### C. STATE OF THE ART: RELAY MAINTENANCE

Due to its practical relevance, the B10 value is presented first. It is often used in connection with the safety of machinery [15]. It provides a statistical value, indicating a lifetime, which 90 % of the components achieve. Accordingly, 10 % of the components fail before the B10 value is exceeded. The lifetime is given in a different unit depending on the component; for a relay, it is typically the amount of switching cycles [15]. The B10 value is used as a threshold applied

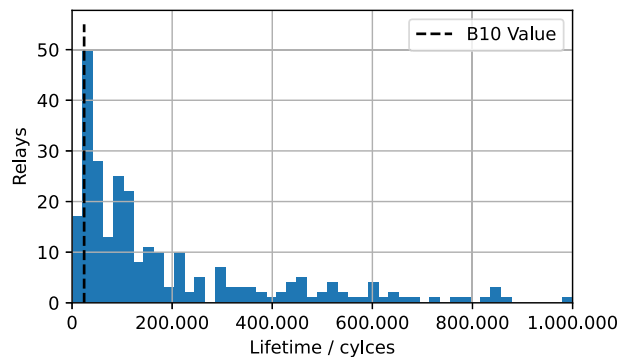


FIGURE 1. Histogram of relay lifetime from training dataset [16] with B10 value plotted.

equally to all components of one type. Thus, the individual degradation of a single device is not considered.

The inadequacies of the B10 value can be illustrated by Fig. 1. The histogram shows the lifetime of the relays from the training dataset, i.e., a total of 266 relays (see Section II-C). The B10 value is about 24 000 switching cycles, i.e., if this is selected as the threshold value for maintenance, then 10 % of the relays fail before maintenance; at the same time, 90 % have a longer lifetime and only a fraction of the possible lifetime (28 %) is used.

Due to the shortcomings of the B10 value, alternative methods have already been developed to improve relay maintenance. Two approaches can be identified: first, a PdM approach for relays was presented in [17], in which health indicators are compared with a recorded dataset. Based on similarity measures, the relays can thus be assigned to known degradation states, which in turn can be used to predict reliability. Second, in [18], the RUL of relays was already successfully estimated with a temporal convolution network and the high potential for an ML-based PdM could be shown. Both approaches consider only failures caused by continuous degradation of the relays. Therefore, the above approaches are not suitable for practical application, because studies show that several error modes occur (cf., Section II-D). These can essentially be divided into the following two groups.

First, relays fail due to continuous degradation; second, they can fail spontaneously. Related to the state of the art for PdM, problems arise: no continuous aging is observed in the relays that fail spontaneously, so estimating monotonic RUL is not possible. Furthermore, reversible processes, such as the formation and release of deposits, complicate the application of RUL estimation. Anomaly detection is also not promising for relay data. Relays exhibit high heterogeneity among themselves in the data: Data from a new relay may strongly resemble a worn one, so anomalous data do not always announce an impending failure. Furthermore, anomalous but nonfaulty operating states are included in the dataset to a considerable extent. This complicates the use of an AE to learn how to reconstruct the anomalous operating states, which actually lead to faults. Therefore, the use of anomaly

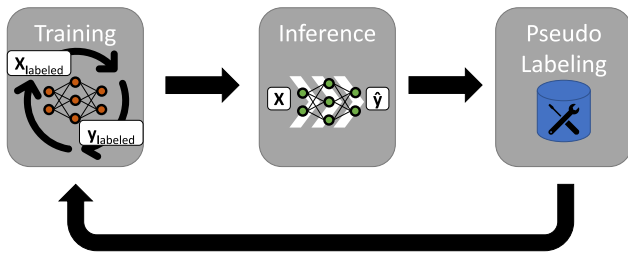


FIGURE 2. Pseudolabeling training procedure.

detection techniques for early indication of impending failure of a relay is not promising. A detailed analysis of this problem, including exemplary implementations of procedures for RUL estimation and anomaly detection, has already been published in [2].

#### D. PSEUDOLABELING

A method for PdM of relays must indicate an impending failure at an early stage so that, depending on the application, appropriate countermeasures can be taken. Therefore, it is desirable to develop a method with which a boundary between conspicuous and inconspicuous operating states can be learned in a semimonitored manner. So that it is possible to react in case of conspicuous states. Pseudolabeling is suitable for this purpose. It is a subset of ML that trains on data that is both labeled and unlabeled. This allows the use of datasets where not all labels are known. In terms of relay degradation analysis, this approach is of particular interest since it requires training with many millions of switching cycles and their labels are unknown.

According to the taxonomy for semisupervised classification by Van Egelen et al. [19] the so-called wrapper methods are suitable for PdM of relays. In these, supervised ML methods are trained to use unlabeled data. The training process for this is described below.

A three-step training process is shown in Fig. 2. First, the training step is performed, which involves supervised ML with inputs  $X_{\text{labeled}} \in \mathbb{R}^{n \times l}$  and labels  $y_{\text{labeled}} \in \mathbb{R}^l$ , where  $n \in \mathbb{N}^+$  is the number of features and  $l \in \mathbb{N}^+$  is the number of labeled switching cycles of a relay. Then, the trained model is used for inference on the entire dataset  $X \in \mathbb{R}^{n \times l+u}$  to generate predictions  $\hat{y} \in \mathbb{R}^{l+u}$ , where  $u \in \mathbb{N}^+$  are the unlabeled switching cycles of a relay. Based on these, the labels  $y \in \mathbb{R}^{l+u}$  are modified. These new labels are also called pseudolabels. Finally, the procedure is restarted and training is performed with the labels and pseudolabels. The procedure is continued until a termination criterion is reached. The advantage of this methodology is that practically all supervised ML procedures can be used, since they are independent of the pseudolabels.

A comprehensive overview of wrapper methods has been published by Triguero et al. [20]. One successful application is the approach presented by Yarowsky for disambiguating word

sense in text documents, where the meaning of words is predicted based on their context [21]. Many other methods have been proposed. For example, in [22] Lee presented a method for selecting pseudolabels based on the largest assignment to a class in each case by the trained artificial neural network (ANN). Despite this rudimentary approach, the conventional semisupervised methods were beaten in terms of performance on the MNIST dataset. Arazo et al. [23] presented a similar method, in which the pseudolabels are used along with the normal labels for training, always ensuring that a minimum of normal labels are included in the minibatches. A successful PdM application is described in [24], where a semisupervised learning method for error diagnostics of gearboxes with limited labeled samples was demonstrated.

#### E. CONTRIBUTION

The following contributions are made in this work.

- 1) Illustrate the complexity of relay degeneration and the resulting problems in applying the current state of PdM to relays.
- 2) Presentation of the MAUD approach, a pseudolabeling-based method that enables PdM of relays.
- 3) Experimental evaluation of MAUD with evidence of superiority over the current industrial state of the art based on a comprehensive real-world dataset.

#### F. OUTLINE

The rest of this article is organized as follows. In Section II general information about electromechanical relays and the experimental setup is given. Based on this, an analysis of the used dataset is performed. In Section III the pseudolabeling approach is applied to the problem, the resulting method maintenance algorithm for unlabeled data (MAUD) is described in detail. The evaluation of MAUD is done in Section IV. Finally, Section V concludes this article.

## II. ELECTROMECHANICAL RELAYS AND MEASUREMENTS

### A. ELECTROMECHANICAL RELAY

In the following, a brief description of the experimental setup and an analysis of the generated data is given. More details regarding the test setup and data are published in [2].

Different types of relays have been developed depending on the application. This study focuses on a monostable relay type with three contacts, which is widely used in the automation industry for switching low-power actuators, such as lights, fans, or valves. A simplified structure of such a relay is visualized in Fig. 4. The relay has three contacts on the load side: normally open (NO), common (COM), and normally closed (NC). For this work, current flows through COM and NO, NC is not used. Further information on the basics on relay design and operation can be found in [25], which cannot be reproduced here for the sake of a concise paper structure.

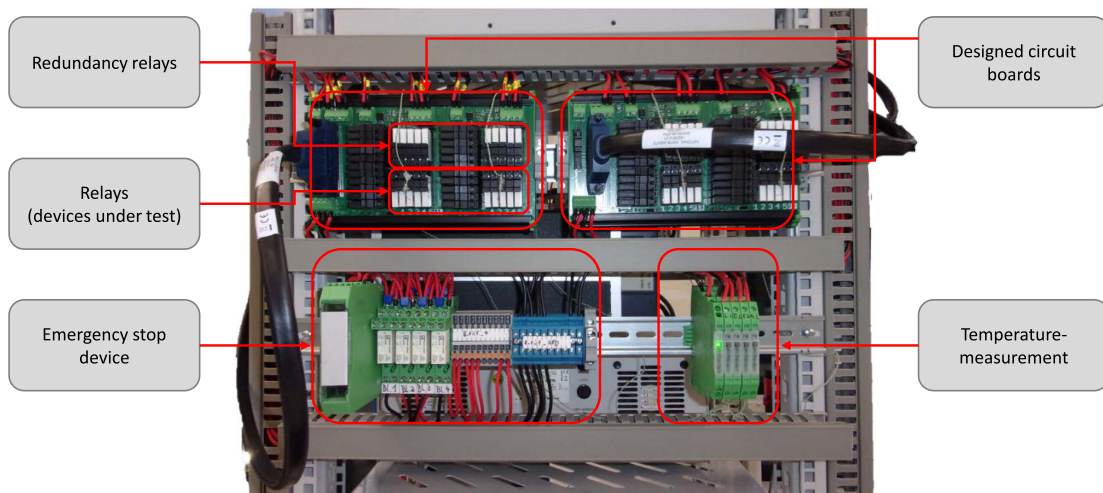


FIGURE 3. Illustration of one of the test setups used.

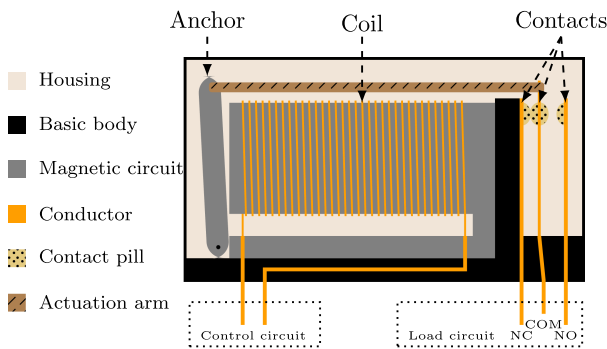


FIGURE 4. Schematic structure of a relay.

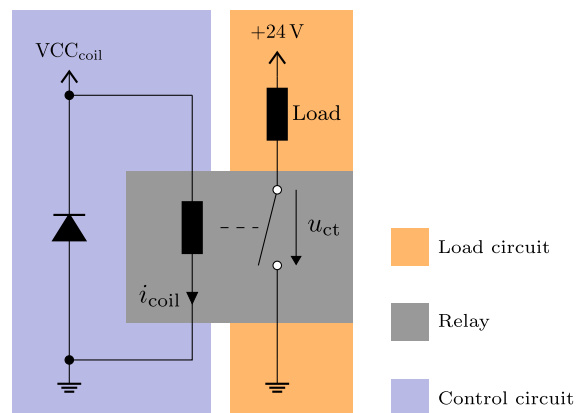


FIGURE 5. Control and load circuit of the used measurement setup.

Typically, many phenomena have an impact on the degradation of electrical relay contacts, with some of them being listed [26]: contact bounce, mechanical structural fatigue, contact stiction, frictional polymer formation, thermal-mechanical damage, and electromigration. The mechanisms leading to failure may overlap so that they are not always traceable. Furthermore, the investigation of a failed relay is difficult: e.g., welded contacts can be separated by vibrations. Therefore, these mechanisms are not investigated in this work and black box modeling is performed.

**B. TEST SETUP**

Fig. 3 shows a picture of one test setup developed and used to collect the dataset. The core of the system consists of two printed circuit boards (PCBs) with ten relay test locations each, the PCBs are mounted side by side on a top-hat rail. For each relay tested, an additional redundant relay is installed to be able to disconnect the current path in the event of a fault. Furthermore, an emergency stop system was installed and a corresponding temperature monitoring system. Both

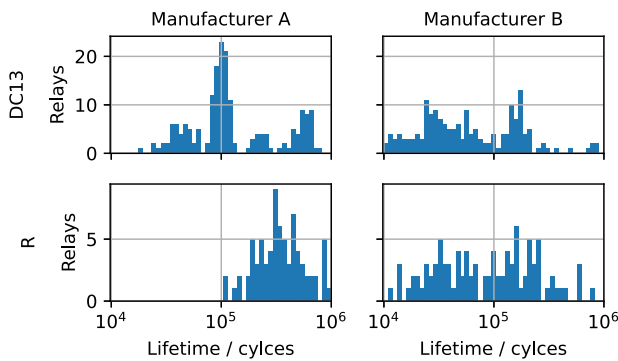
are necessary to be able to use the test setup in continuous operation.

The circuit layout developed for data generation is described in Fig. 5 by providing a simplified overview: The control circuit consists of the relay coil and a freewheeling diode required to protect against induced overvoltages through the coil. The coil current  $i_{coil}$  is measured by a shunt, whose influence on the switching behavior of the relay is negligible. The load circuit consists of load and relay contacts, and the load is interchangeable so that the dataset can be diversified. The contact voltage  $u_{ct}$  is measured directly. Both circuits are connected via the relay. The measured variables of a switching cycle are shown as an example in Section III-A.

For testing the manufacturers, the loads, and the load currents are varied. Two manufacturers with the same technical specifications are used, which are therefore interchangeable in practice. For the loads, only components with resistive or inductive characteristics are considered, since relays are not suitable for switching the high inrush currents of capacitive

**TABLE 1. Operating Conditions**

| Number | Manufacturer | Load | Current / A |
|--------|--------------|------|-------------|
| 1      | A            | DC13 | 1.70        |
| 2      | A            | DC13 | 1.25        |
| 3      | A            | DC13 | 1.00        |
| 4      | A            | R    | 6.00        |
| 5      | A            | R    | 4.00        |
| 6      | B            | DC13 | 1.70        |
| 7      | B            | DC13 | 1.25        |
| 8      | B            | DC13 | 1.00        |
| 9      | B            | R    | 6.00        |
| 10     | B            | R    | 4.00        |



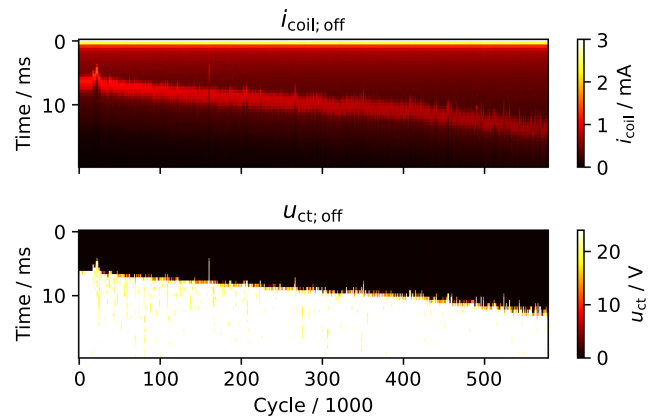
**FIGURE 6. Histograms of relay lifetime for different manufacturers and loads.**

loads. Therefore, the DC13 load, a coil with defined inductance, is used, which has established itself as a test load with various relay manufacturers [27]. Furthermore, resistive loads (R) are used. In summary, a total of ten different operating conditions are investigated, as given in Table 1.

According to the two operating states of a relay, there are the following two failure patterns. On the one hand, a switched on relay can carry no current and on the other hand, a switched OFF relay can carry current. A relay might continue to work successfully after a faulty switching cycle. One example of this is hooked contacts. Operation can roughen the surfaces of the contacts to such an extent that they become hooked and do not open when switched OFF. Light vibrations, such as switching ON again, can cause the contacts to become detached from each other, so that the relay works again after a faulty switching operation. For the purposes of this work, a relay is considered as being defect if a failure has occurred at least twice in a row.

### C. DATASET ANALYSIS

As a basis for the following investigation, comprehensive experimental measurements were carried out on 546 relays. These have been published as an open dataset, which is available to the public [16]. The lifetime of the relays is strongly related to the switched load and the manufacturer, as shown in Fig. 6. The columns are assigned to the manufacturers and the rows to the load, the abscissa is scaled logarithmically, since early failures are much more frequent than late failures and there are high variations in the lifetime depending on the load.



**FIGURE 7. Visualization of the time series during degradation for a relay switching a DC13 load.**

For manufacturer A, there are significant differences between the loads, for the DC13 load four segments can be divided, which have an average life of about 100 000 switching cycles. In contrast, for the resistive load (R), segmentation is not possible and the relays have an average life of about 500 000 switching cycles. The differences in lifetimes are not as apparent for manufacturer B, but they are present here as well; overall, the lifetimes are more evenly distributed over a wider range. The same can be seen when comparing the loads.

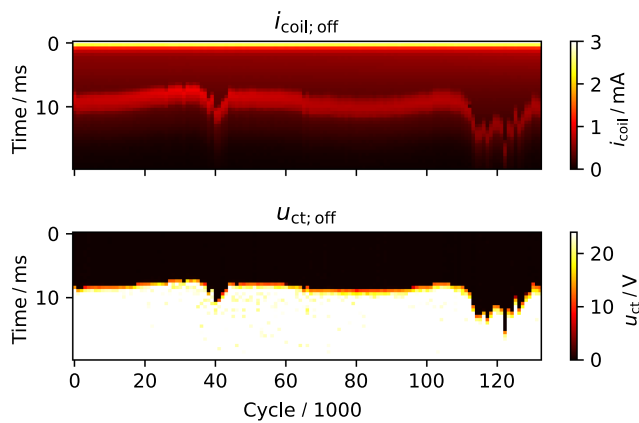
Ultimately, different lifetimes can be expected depending on the manufacturer and load. This can be attributed to design differences and the physical properties of the loads. With the DC13 load, for example, an arc is generated when the load is switched OFF, which heats the contacts extremely at certain points so that material is burned off.

### D. VISUALIZATION OF DEGRADATION

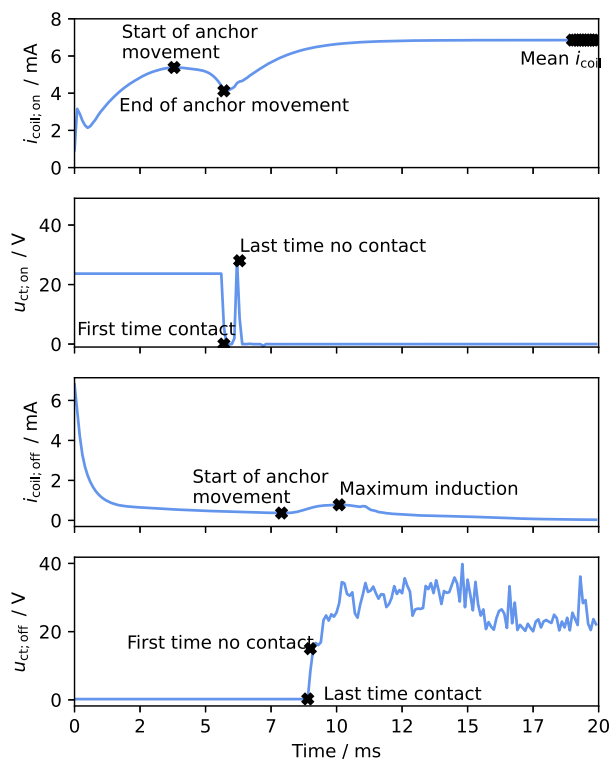
The following shows an example of the data when two relays are switched OFF over their entire lifetime; similar conclusions can be drawn for the data when they are switched ON.

In Fig. 7, a continuously degenerated relay operated with a DC13 load is shown. There are clear trends in the data, the local maximum in the  $i_{coil}$  at switch-OFF occurs later and later, and the time at which  $u_{ct}$  at switch-OFF reaches 24 V again is delayed with lifetime. This can be explained by the burning of the contacts; with each switching cycle, some material of the contact pills is burned. Consequently, the spring force with which the contacts are separated decreases, so that the anchor detaches from the coil later and more slowly.

There are no clear trends in Fig. 8 with increasing lifetime. There are spontaneous changes, but they fade away, e.g., at 40 thousand cycles of switching OFF, a change is observed that no longer exists at 50 thousand cycles of switching OFF. At the end of life, a further, stronger change can be observed. However, a clear trend over the lifetime cannot be seen. This is a typical example of a relay with a spontaneous failure, as this is only announced in the short term.



**FIGURE 8.** Visualization of the time series during degradation for a relay switching a DC13 load.



**FIGURE 9.** Extracted features marked on exemplary measurements.

### III. MAINTENANCE ALGORITHM FOR UNLABELED DATA

In the following paragraph, a new PdM method for relays is presented. The individual failure risk of a relay is estimated based on measured data. When operating a relay, maintenance could be triggered based on this failure risk so that faults can be prevented. Since the used dataset is not labeled, a pseudolabeling is performed.

The data processing pipeline is shown in Fig. 10. First, features are extracted from the time series of each switching operation. These are used together with the time series in an initialization phase to classify the manufacturer and load of

the relays. The need for this comes from the previous statistical study in Section III-A: manufacturer and load have an impact on degeneration. The initialization phase takes place during the first 2 h of operation (720 cycles) of each relay. In addition to the classification, the initial values of the characteristics of each relay are determined here. After initialization, inference can begin and the risk of failure for each switching operation can be estimated. For this purpose, the features are first augmented and then used for prediction with an multilayer perceptron (MLP), a simple type of ANN consisting of layers of neurons with activation functions. But before an MLP can be used for prediction, it must be trained. In the case of MAUD, pseudolabeling is performed using a wrapper method. The following sections contain detailed information on both the data processing pipeline and training.

### A. FEATURE EXTRACTION

First, a number of 12 features  $v \in \mathbb{R}^{12}$  are extracted from the voltage and current time series with the goal of reducing complexity as much as possible. The extracted features are marked in Fig. 9 and described below.

- 1) *Start of the anchor movement*: The local maximum in the coil current is related to the start of the anchor movement and can thus, among other things, allow conclusions to be drawn about the entire magnetic circuit, such as the rest position of the anchor.
- 2) *End of anchor movement*: The end of the anchor movement can be seen as a local minimum. If there is contamination or debris between the anchor and the coil core or between the contact pairs, the trajectory of the anchor would be altered. An example of this is the migration of contact material, which reduces the contact gap to such an extent that the anchor can no longer strike the coil core and, therefore, oscillates. This change is reflected in the coil current.
- 3) *First time contact*: The time of the first contact is detected by the first voltage drop. It allows conclusions to be drawn about changes in the contact distance.
- 4) *Last time no contact*: This feature determines the duration of the bounce. For this reason, the time specified here corresponds to the time difference to “First time contact.” A high level of bouncing is a sign of wear.
- 5) *Start of anchor movement*: This feature provides information on the release of the anchor from the coil and thus allows conclusions to be drawn about the mechanical tension of the contacts.
- 6) *Maximum induction*: The maximum induction depends on the anchor movement and is therefore considered.
- 7) *Last time contact*: The time at which the contacts are released when switched OFF is related to the mechanical contact voltage.
- 8) *First time no contact*: This feature indicates how long it takes for the contact voltage to rise above 24 V again for the first time. The measured time here is the time difference to “Last time contact.”

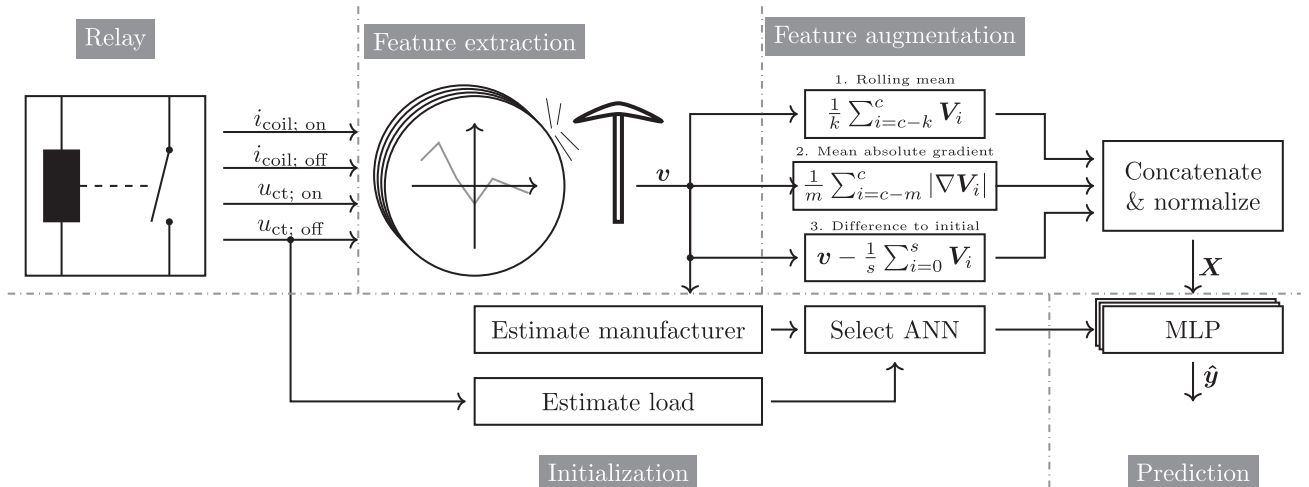


FIGURE 10. Overview of the MAUD data processing pipeline.

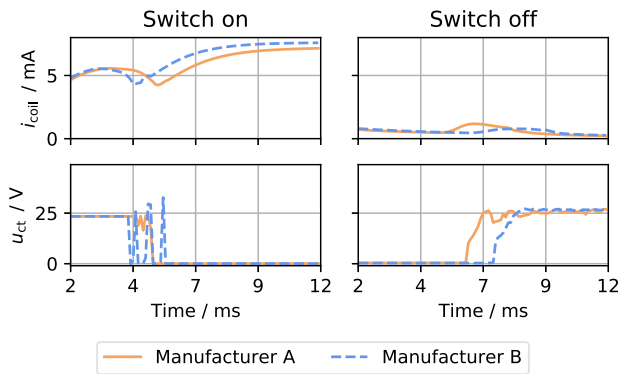


FIGURE 11. Comprehension of time series for two different manufacturers.

For the four features based on the coil current  $i_{coil}$ , both time and current are determined. However, for those four based on the contact voltage  $u_{ct}$ , only the time is recorded. This is because the contact voltage is an almost binary signal, so the voltage does not contain any significant information. Thus, a total of eight time points and four current values are extracted from the measured variables of a switching operation.

In addition, the mean value of  $i_{coil}$  is marked in Fig. 9. It is only used for data correction purposes and therefore not listed as a feature. The mean value is important, because it depends, among other things, on the supply voltage and the coil temperature. Therefore, fluctuations of the two influencing variables are corrected by dividing the values of the current characteristics by the mean value.

## B. INITIALIZATION

### 1) ESTIMATE MANUFACTURER

The relays are differentiated by manufacturer, in Fig. 11 the time series for two new relays from both manufacturers are

shown, for  $i_{coil}$  different time series can be seen when switching ON and OFF. These are due to the geometries of the magnetic circuits and the inductances of the coils. The spring forces of the contacts also effect the movement of the anchor. When switching on,  $u_{ct}$  shows that the relay from manufacturer B bounces more strongly, an observation that also applies to the majority of the other relays. With  $u_{ct}$  switched OFF, it can be seen that the contacts separate at different time points.

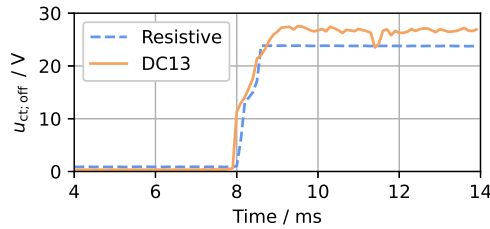
All in all, it can be concluded that it seems possible to distinguish relays by manufacturer, so the features described in Section III-A are used for this classification task. As an ML method, an MLP is trained with a hidden layer of 25 neurons and a rectified linear unit (ReLU) activation function and one neuron as output with sigmoid activation function. MAUD thus contains two MLPs, one for the pseudolabeling and one for the manufacturer estimate. The latter achieves 99.99 % accuracy in the training dataset and 99.90 % accuracy in the validation dataset. Since the manufacturer estimation is part of the initialization of MAUD, a majority voting can be performed during the first 720 cycles. In this way, 100 % of the relays in the test dataset can be assigned to the correct manufacturer.

### 2) ESTIMATE LOAD

Since different physical phenomena occur depending on the switched load, different degradation are to be expected. Therefore, the relays are differentiated by load. For differentiation,  $u_{ct}$  was used when switching off, since for the DC13 inductive loads, arcing always occurs, which is noticeable by noise, as shown in Fig. 12. For the DC13 load the standard deviation is 3.25 V and for the resistive load 0.03 V, in this way both loads can be distinguished.

## C. FEATURE AUGMENTATION

Based on  $v$ , a further processing takes place. Since the features refer only to a single switching operation, the changes between the switching operations do not emerge from it.


**FIGURE 12.** Contact voltages during switch-off for two different loads.

**TABLE 2.** MLP Topology

| Layer | Type    | Neurons | Activation | Dropout-rate |
|-------|---------|---------|------------|--------------|
| 1     | Dense   | 50      | ReLU       | -            |
| 2     | Dense   | 50      | ReLU       | -            |
| 3     | Dense   | 50      | ReLU       | -            |
| 4     | Dropout | -       | -          | 0.25         |
| 5     | Dense   | 1       | sigmoid    | -            |

Besides, outliers are partly to be found out, which must be corrected. Therefore, the history  $V_r \in \mathbb{R}^{12 \times l_r}$  of each relay  $r \in \mathbb{N}^+$  with lifetime  $l_r \in \mathbb{N}^+$  is used to extend the dataset for each cycle  $c_r \in \mathbb{N}^+$ .

Three augmentations are made: the moving average of the feature with a window width  $k \in \mathbb{N}^+$ , a measure of the magnitude of change in the feature (as a mean gradient of  $m \in \mathbb{N}^+$  cycles), and the deviation from the initial state (which is averaged over the first  $s \in \mathbb{N}^+$  cycles). In this way, the dataset contains a total of  $3 \times 12$  features per switching cycle as input data. These are finally normalized, with the mean and standard deviation calculated for each feature over the entire training set.

#### D. ANN TOPOLOGY

For the application of the method, the inference of the MLP is of particular importance. The costs of the necessary periphery for inference must be proportionally to the relay costs. Therefore, the choice has fallen on a MLP whose hardware requirements are so low that it can be executed on a micro-controller.

The MLP topology is given by Table 2. A random search was performed to optimize the hyperparameters of the MLP. Within 20 random draws, the number of neurons per layer (5, 10, 25, 50, 75, and 100), the number of layers (1, 2, 3, 5, and 10), the activation function (ReLU and sigmoid) and the dropout rate (0.05, 0.10, 0.15, 0.20, and 0.25) were optimized. A uniform distribution was used in each case. Accuracy for the first iteration of MAUD was used as a metric for performance.

#### E. TRAINING

The failure risk is used as the target value, which is a binary classification. It is used to identify signs of spontaneous failure. During the manual analysis of the dataset, anomalies can be found in some places that should be recognized as part of the classification. The assignment of the labels is demanding, since a manual check of the extensive dataset with millions

**TABLE 3.** Training Parameters

| Name                                | Value                |
|-------------------------------------|----------------------|
| Library                             | TensorFlow / Keras   |
| Optimizer                           | Adam                 |
| Epochs                              | 1000                 |
| Batch size                          | 10 000               |
| Data shuffling                      | True                 |
| Early stopping                      | True                 |
| Early stopping patience             | 20                   |
| Early stopping restore best weights | True                 |
| Error function                      | Binary cross entropy |

of switching cycles is not possible. Therefore, a semisupervised approach is pursued, with which pseudolabels for the switching cycles are assigned by the MLP itself. The training procedure corresponds to the three steps from Fig. 2.

Two assumptions are made for the initialization: On the one hand, it is assumed that the last cycles are conspicuous and can be labeled as such. Therefore,  $\lambda \in \mathbb{N}^+$  is used to (1) and (2). On the other hand, it is assumed that most of the cycles in the first half of the relay's life are inconspicuous, which is reflected in  $\sigma$ . The initial target value  $y$  for the cycle  $c_r$  is thus

$$y_{c_r} = \begin{cases} 0 \text{ (inconspicuous)} & \text{if } c_r < \sigma \\ 1 \text{ (conspicuous)} & \text{if } c_r > l_r - \lambda \\ \text{not used} & \text{else.} \end{cases} \quad (1)$$

The MLP's predictions  $\hat{y}$  for the training or validation set are used to update the labels

$$y_{c_r} = \begin{cases} 0 & \text{if } c_r < \epsilon \text{ or } \hat{y}_{c_r} < \alpha \\ 1 & \text{if } \hat{y}_{c_r} > 1 - \alpha \text{ and } c_r > \sigma \\ 1 & \text{if } c_r > l_r - \lambda \\ \text{not used} & \text{else} \end{cases} \quad (2)$$

whereby the first  $\epsilon \in \mathbb{N}^+$  cycles are labeled as inconspicuous and the last  $\lambda$  cycles are labeled as conspicuous. All cycles in between are relabeled based on  $\hat{y}$ . By this labeling it is possible, e.g., to label the conspicuous cycles in Fig. 8 at 40 thousand accordingly and to permit at the same time that following again an inconspicuous phase is run through, which cycles are marked depends on the threshold  $\alpha \in ]0, 0.5[$ . The restriction that only  $c_r > \sigma$  can be labeled as conspicuous is because otherwise many cycles in an early life phase would also be labeled as conspicuous and thus little life of a relay would be used. At the same time, cycles that were initially incorrectly labeled as inconspicuous are removed from the dataset in this way so that both labels can be better distinguished.

The training parameters used are listed in Table 3. For each iteration, a new MLP is trained because it has been shown that otherwise overfitting occurs quickly. Furthermore, equal numbers of conspicuous and inconspicuous labels are included in the training and validation datasets, so that none of the labels is overrepresented. The choice of the cycles used is random.

#### IV. EXPERIMENTAL RESULTS

In the following section, the results of the MAUD procedure are presented. First, ablation studies for the choice of the



**TABLE 4. Number of relays in training, validation and test set by manufacturer and load**

| Manufacturer & Load | Training   | Validation | Test       |
|---------------------|------------|------------|------------|
| A-DC13              | 93         | 49         | 49         |
| B-DC13              | 87         | 46         | 46         |
| A-R                 | 38         | 20         | 21         |
| B-R                 | 48         | 24         | 25         |
| <b>Total</b>        | <b>266</b> | <b>139</b> | <b>141</b> |

**TABLE 5. Ablation Study of Parameter  $\alpha$**

| $\alpha$ | Utilization / % | Failures / % |
|----------|-----------------|--------------|
| 0.02     | 31.67           | 2.80         |
| 0.05     | 45.15           | 2.80         |
| 0.10     | 35.74           | 2.80         |
| 0.20     | 26.70           | 4.26         |

threshold value  $\alpha$  and the feature augmentation are presented. Thereupon, the training of MAUD is documented. Subsequently, the individual results are first considered qualitatively and finally compared with the state of the art.

The following two metrics are used to compare the results.

- 1) *Utilization*: The proportion of the lifetime that the relay has been operated before maintenance.
- 2) *Failures*: The proportion of relays that would not have been maintained before their failure.

Both metrics are important: the greater the utilization, the less maintenance and resource wasting. At the same time, however, failures must be kept low, as these can lead to the failure of the entire plant.

### A. TRAINING SETUP

In Table 4, the numbers of relays that make up the training, validation and test sets are given, the latter being assigned a high proportion of relays (25 % each) to obtain higher evidence of the results. The relays sorted by manufacturer and load were randomly assigned to the sets.

The threshold  $\alpha$  is of particular importance for label assignment, so the effect of different threshold values is investigated. In Table 5 the results of MAUD with different  $\alpha$  are given.

It can be seen that for  $\alpha = 0.05$  there is the highest exploitation of the utilization with 45.15 % and at the same time the lowest undetected errors with 2.80 %. This can be explained by the fact that if the  $\alpha$  is too low, too few pseudolabels are assigned and therefore a poorer distinction between conspicuous and inconspicuous is learned. A too large  $\alpha$ , on the other hand, leads to a fast allocation of pseudolabels, which make the labeling iterations unstable and ultimately lead to worse results.

Furthermore, the features used have an influence on the performance of MAUD. For this reason, different combinations of features were used in advance for the training of MAUD to investigate their influence on the performance (see Section III-C).

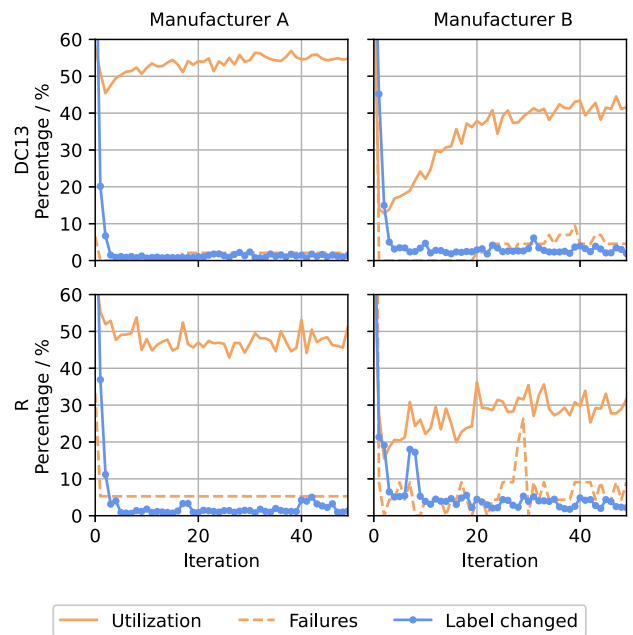
The best performance is achieved by using all features, which is why they are used in the following. The parameters used for feature augmentation are listed in Table 7.

**TABLE 6. Ablation Study of Feature Augmentation**

| Augmentation  | Utilization / % | Failures / % |
|---|-----------------|--------------|
| Rolling mean  | 32.15           | 3.55         |
| Rolling mean and mean absolute gradient                           | 29.26           | 3.55         |
| Rolling mean and difference to initial                            | 34.92           | 3.55         |
| Rolling mean and mean absolute gradient and difference to initial | 45.15           | 2.80         |

**TABLE 7. Used MAUD Parameter Values**

| Parameter  | Description                    | Value   |
|------------|--------------------------------|---------|
| $n$        | Number of features             | 12      |
| $k$        | Window length moving average   | 120     |
| $m$        | Window length average gradient | 120     |
| $s$        | Window length initial values   | 720     |
| $\lambda$  | Assumed conspicuous cycles     | 360     |
| $\epsilon$ | Assumed inconspicuous cycles   | 20      |
| $\sigma$   | Assumed normal operation phase | $l_r/2$ |
| $\alpha$   | Threshold for pseudolabels     | 0.05    |



**FIGURE 13. History of utilization, failures, and changed labels over training iterations.**

### B. TRAINING

Fig. 13 shows the results of the training iterations, divided by manufacturer and load. For each manufacturer and load combination, MAUD is run independently, at the end of each training phase three sizes are recorded for the validation set. Label changed quantifies the percentage of cycles changed by the predictions of the newly trained MLP for the next iteration. Utilization quantifies the average percentage of relay lifetime that is used, assuming that a relay is changed when it is classified as conspicuous. To accommodate real-world conditions, it was specified that labeling as conspicuous must occur at least 1 h (360 cycles) before failure. Otherwise, a relay is evaluated

as a failure, which is also specified relatively in Fig. 13. To suppress outliers, the MLP predictions were smoothed with a moving average with a window width of 120.

For manufacturer A, it can be seen that after five iterations the changed labels converge to a small percentage, these can be justified with the training of the MLP, since randomness plays a large role here, e.g., the training data are randomly selected. Consequently, the MAUD algorithm could be used to separate the conspicuous and inconspicuous cycles. Thereby, usage and failures for both loads change only slightly after convergence of the changed labels, this is also a proof of the safe differentiation of the classes.

The convergence of relabeled cycles can be seen on for manufacturer B, but the relative proportion is significantly larger, so the label distinction does not appear to be as stable. This is supported by the utilization and the failures, for the resistive load a high dispersion of the failures is observed and for the DC13 load the utilization increases during the first iterations. This behavior can be justified in two ways, first the relays of manufacturer A have almost twice the lifetime for DC13 loads and almost four times the lifetime for resistive loads, the dataset is therefore larger, which may have a positive impact on training. On the other hand, the histograms from Fig. 6 show that the relays from manufacturer A turn out to be more predictable (smaller dispersion in the histogram).

When comparing the validation and test sets, only a few percent deviations can be seen for all manufacturers and load combinations in terms of utilization, which indicates that there is no overfitting, but that MAUD has actually learned a reliable classification. A total of three more failures occur in the test set; these relays, with live times of 5000–18 000 cycles, are extreme early failures that do not occur in either the training or test set.

### C. INDIVIDUAL RESULTS

In the following, the predictions are visualized in combinations with the data. For this purpose, two relays were selected, which can be used to show how MAUD works. These are the relays shown previously in Figs. 7 and 8, also the graphics have the same structure, only this time the predictions are visualized in color.

In the case of the continuously aged relay from Fig. 14, it can be seen that it is successfully labeled as conspicuous at the end of its lifetime, which is in line with expectations. The same applies to the range between 400 and 450 thousand cycles, which has some unlabeled cycles. What is striking about this relay is that it is classified as dangerous right at the start, i.e., before 50 000 cycles. A closer look at the data reveals strong changes in the measured variables in this range, so the cycles are indeed conspicuous, but the relay has not failed. This behavior can be observed many times, the relays got conspicuous cycles, but only a fraction actually fails. This has a very negative effect on the evaluation of the utilization, but to achieve a low failure rate, the relays must be changed correspondingly early, since some relays fail with exactly these conspicuous cycles.

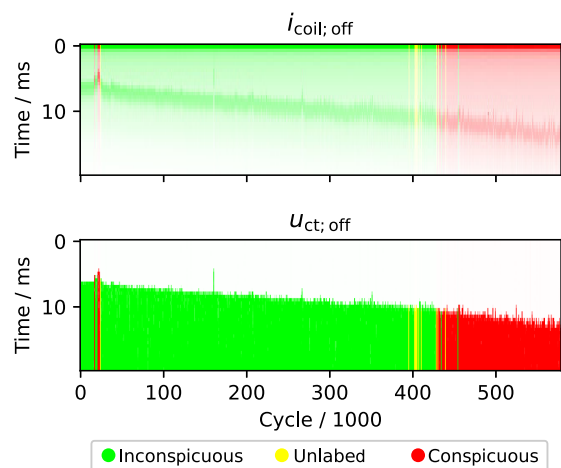


FIGURE 14. Illustration of predictions for a relay switching a DC13 load.

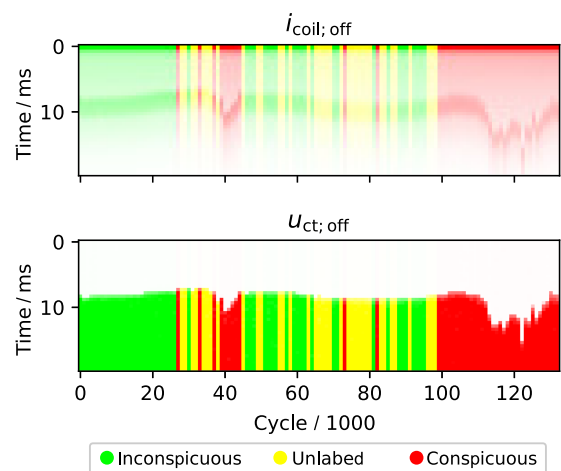


FIGURE 15. Illustration of predictions for a relay switching a DC13 load.

The phenomenon of early conspicuous cycles observed in Fig. 14 is even more evident in Fig. 15, where many more cycles were not labeled and the intervals where large changes are seen are marked as conspicuous. The pseudolabeling with MAUD is successful according to this qualitative sample, but the utilization of the relays is limited in the quantitative evaluation, since it can be seen that conspicuous cycles can occur long before the relays fail.

### D. COMPARISON WITH STATE OF THE ART

In the following section, the performance of MAUD is compared with the state of the art. For this purpose, three alternative methods were implemented and evaluated with the same augmented dataset described in Section III-A. First, the B10 value, because this is used in practice to PdM the relays. Second, the estimation of RUL, since this method is popular and has been used for PdM of relays in [18]. Third, anomaly detection, due to the popularity and the high complexity of the

**TABLE 8. Results**

| Method        | Utilization / % | Failures / % | Failures |
|---------------|-----------------|--------------|----------|
| Validation    |                 |              |          |
| MAUD (A-R)    | 53.81           | 5.26         | 1        |
| MAUD (B-R)    | 35.43           | 0.00         | 0        |
| MAUD (A-DC13) | 54.46           | 0.00         | 0        |
| MAUD (B-DC13) | 37.19           | 0.00         | 0        |
| Test          |                 |              |          |
| MAUD (A-R)    | 51.34           | 0.00         | 0        |
| MAUD (B-R)    | 33.04           | 0.00         | 0        |
| MAUD (A-DC13) | 56.88           | 6.52         | 3        |
| MAUD (B-DC13) | 37.02           | 2.22         | 1        |

**TABLE 9. RUL Estimation Topology**

| Layer | Type  | Neurons | Activation |
|-------|-------|---------|------------|
| 1     | Dense | 50      | ReLU       |
| 2     | Dense | 50      | ReLU       |
| 3     | Dense | 50      | ReLU       |
| 4     | Dense | 25      | ReLU       |
| 5     | Dense | 12      | ReLU       |
| 6     | Dense | 6       | ReLU       |
| 7     | Dense | 1       | ReLU       |

relay data. In the following, the methods are briefly presented and then compared with MAUD.

### 1) RUL ESTIMATION

In RUL estimation, a ML method is trained to perform regression in a supervised manner. The augmented features are used as input data and the remaining lifetime is used as outputs. Since the service life of relays depends on the load, among other factors, the RUL is defined relatively, i.e., scaled to the interval from 100 % to 0 % individually for each relay. Thus, each new relay has a RUL of 100 % and immediately before failure a RUL of 0 %.

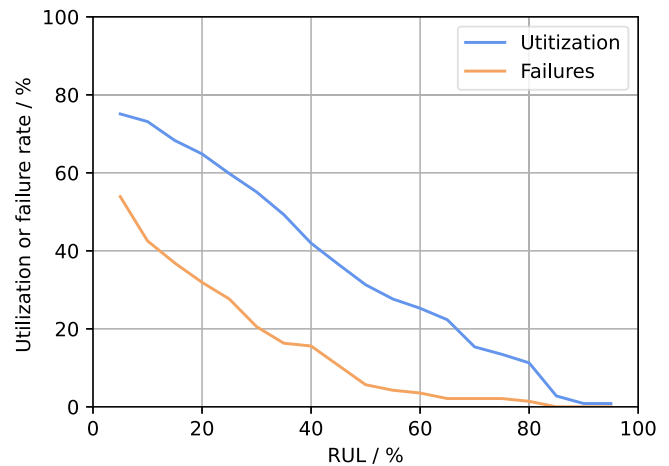
The ML method used here was an MLP whose topology is listed in Table 9. Training is analogous to the modalities in Table 3, except for the error function: mean squared error is used for RUL estimation. To suppress outliers, the MLP predictions were smoothed with a moving average with a window width of 120.

If RUL estimation is used, a threshold value has to be defined, at which the maintenance of the relays is triggered. For comparison with MAUD, Fig. 16 shows the influence of the threshold value on the metrics utilization and failures.

It can be observed that both metrics decrease with increasing threshold value. This is to be expected: the higher the threshold value, the earlier the relays are maintained and consequently there are fewer failures, although utilization is also low.

### 2) ANOMALY DETECTION

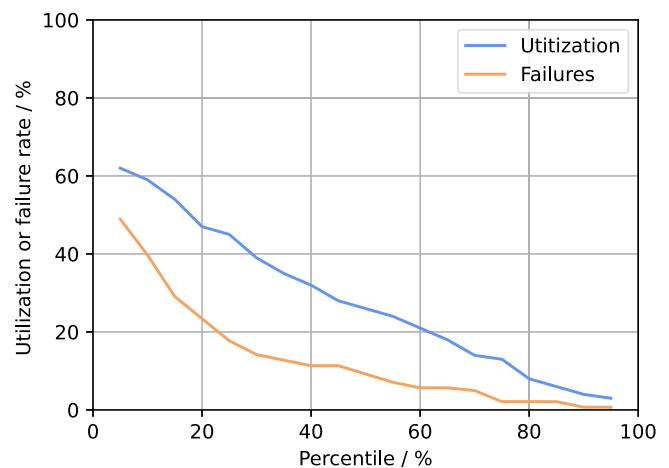
Anomaly detection is a method that involves an unsupervised learning approach. An AE is trained with the first 5000 switching cycles of each relay from the training dataset, using the augmented features from both input and output data. The



**FIGURE 16. Utilization and failures for the RUL estimation when different RUL threshold values are selected.**

**TABLE 10. AE Topology**

| Layer | Type  | Neurons | Activation              |
|-------|-------|---------|-------------------------|
| 1     | Dense | 36      | Exponential linear unit |
| 2     | Dense | 18      | Exponential linear unit |
| 3     | Dense | 9       | Exponential linear unit |
| 4     | Dense | 18      | Exponential linear unit |
| 5     | Dense | 36      | None                    |



**FIGURE 17. Utilization and failures for anomaly detection when selecting the error threshold value according to different error percentiles.**

topology of the AE is given in Table 10, and training is analogous to RUL estimation.

Anomalies are detected using the reconstruction error, i.e., from the difference between input and output data. If the error is above a threshold value, then relay maintenance is performed. For comparison with MAUD, the threshold values were determined as percentiles of the reconstruction error of the validation dataset: the 5 % percentile corresponds to the boundary between the largest 5 % of errors and the remaining 95 %. In Fig. 17, utilization and failures are given for the

**TABLE 11. Summarizing Results**

| Method                  | Utilization / % | Failures / % |
|-------------------------|-----------------|--------------|
| B10                     | 28.08           | 9.22         |
| B10 (differentiated)    | 36.75           | 9.93         |
| RUL estimation (45%)    | 35.56           | 10.64        |
| Anomaly detection (50%) | 26.13           | 9.22         |
| RUL estimation (60%)    | 25.28           | 3.55         |
| Anomaly detection (75%) | 13.32           | 2.13         |
| MAUD                    | 45.15           | 2.80         |

different percentiles. To suppress outliers, the AE errors were smoothed with a moving average with a window width of 120.

The graph shows that the higher the percentile, the lower the utilization and failures. This is to be expected, since the larger the percentile, the more values lie above the threshold value, resulting in fewer failures and utilization.

### 3) B10 VALUE

Two B10 values were calculated based on the training set: with and without differentiation of manufacturer and load. Classically, the differentiation is not carried out in this way for the B10 value. However, since it has been shown in this publication that it is possible, it is also considered for the B10 value to increase its performance and, therefore, ensuring a fair comparison.

### 4) COMPARISON

For each pseudolabeling iteration of MAUD, a trained MLPs are available for each manufacturer and load combination. Therefore, iteration selection must take place. For this, the performance on the validation set is used, considering first the minimum failures and second the maximum load, to select the iteration and thus the MLP.

For the purpose of comparison, a threshold value must be set for each of the RUL estimation and anomaly detection methods. The choice is based on the results of MAUD and the B10 value. Because the reduction of the failures is preferred, the threshold values are chosen so that the failures of the methods are similar. For MAUD, this results in an RUL threshold value of 60 % and the 75 % percentile for anomaly detection. The thresholds for the B10 values are 45 % (RUL) and 50 % (anomaly).

The results are given in Table 11. First, the B10 values are considered: the statistical parameter implies that about 10 % of failures occur for the test dataset. Consequently, both B10 values are as expected. However, they differ significantly in utilization. This is due to the fact that higher B10 values are calculated for the differentiated manufacturers and load combinations, which means that more lifetime is used on average.

Second, the methods RUL estimation and anomaly detection are considered. Compared with the B10 values, the comparison is made with the threshold values 45 % and 50 %, respectively, so that no significant differences can be found with regard to the failures. At the same time, however, no improvement can be observed with respect to the

utilization; the B10 values even have a slight advantage. The effort required for the RUL estimation and anomaly detection methods can therefore not be justified. One reason for the lack of performance improvement may be the nature of the relay degradation processes: RUL estimation assumes monotonic degradation, which not all relays satisfy (cf., Fig. 8). In anomaly detection, the fact that the relays have heterogeneous characteristics among themselves comes into play. Thus, anomaly detection detects not only abnormal operating states but also abnormal relays, thereby reducing utilization. A more detailed analysis of the application of both methods can be found in [2].

Finally, MAUD is considered: the comparison with both B10 values is clear, MAUD can increase utilization by 17.07 p.p. (8.40 p.p. for differentiated) and the failures are reduced by 6.42 p.p. (7.13 p.p. for differentiated). With regard to both metrics, a significant increase was achieved through MAUD. A similar picture emerges when comparing the RUL estimation (60 %) and the anomaly detection (75 %) such that their remaining failure rates is on eye level with MAUD, i.e., making them more conservative. However, the utilization of MAUD is 19.87 p.p. (RUL estimation) and 31.83 p.p. (anomaly detection) higher. Thus, the performance of the existing PdM methods RUL estimation and anomaly detection is much worse than that of MAUD for relays.

## V. CONCLUSION

Existing approaches like RUL prediction or anomaly detection are not suitable for PdM of electromechanical relays. This is mainly due to the complex degradation processes. Therefore, a new target variable, failure risk, was introduced to indicate impending failure. With pseudolabeling, a new MAUD was introduced and evaluated. MAUD is superior to the current state of the art in relay maintenance: Utilization can be improved by 17.07 p.p., and furthermore, undetected failures can be reduced by 6.42 p.p. In addition, MAUD is also superior to existing PdM methods, achieving up to 38,83 p.p. higher utilization for comparable failures.

Since real data was used in the context of this work, it can be concluded that the practical application of MAUD enables longer operation with fewer failures.

## REFERENCES

- [1] "Electromechanical relay market - forecasts from 2021 to 2026," 2021. Accessed: Aug. 10, 2023. [Online]. Available: <https://www.researchandmarkets.com/reports/5332646/electromechanical-relay-market-forecasts-from>
- [2] F. Winkel, J. Deuse-Kleinstüber, and J. Böcker, "Run-to-failure relay dataset for predictive maintenance research with machine learning," *IEEE Trans. Rel.*, early access, Mar. 28, 2023, doi: [10.1109/TR.2023.3255786](https://doi.org/10.1109/TR.2023.3255786).
- [3] A. Saxena and K. Goebel, "Turbofan engine degradation simulation data set," *NASA Ames Prognostics Data Repository*, NASA Ames Research Center Moffett Field, CA, USA, 2008. Accessed: Aug. 10, 2023. [Online]. Available: <https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository>
- [4] Y. Peng, H. Wang, J. Wang, D. Liu, and X. Peng, "A modified echo state network based remaining useful life estimation approach," in *Proc. IEEE Conf. Prognostics Health Manage.*, 2012, pp. 1–7.

- [5] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *Proc. Int. Conf. Prognostics Health Manage.*, 2008, pp. 1–6.
- [6] N. Patrick, G. Rafael, M. Kamal, and R. Emmanuel, "An experimental platform for bearings accelerated life test," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, 2012, pp. 1–8.
- [7] T. H. Loutas, D. Roulias, and G. Georgoulas, "Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression," *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 821–832, Dec. 2013.
- [8] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.
- [9] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling health monitoring approach based on vibration data for accurate prognostics," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 647–656, Jan. 2015.
- [10] C. C. Aggarwal, "An introduction to outlier analysis," in *Outlier Analysis*. Berlin, Germany: Springer, 2017, pp. 1–34.
- [11] O. Serradilla, E. Zugasti, J. R. d. Okariz, J. Rodriguez, and U. Zurutuza, "Adaptable and explainable predictive maintenance: Semi-supervised deep learning for anomaly detection and diagnosis in press machine data," *Appl. Sci.*, vol. 11, no. 16, 2021, Art. no. 7376.
- [12] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–836, 2018.
- [13] Y. Ran, X. Zhou, P. Lin, Y. Wen, and R. Deng, "A survey of predictive maintenance: Systems, purposes and approaches," 2019, *arXiv:1912.07383*.
- [14] O. Serradilla, E. Zugasti, J. Rodriguez, and U. Zurutuza, "Deep learning models for predictive maintenance: A survey, comparison, challenges and prospects," *Appl. Intell.*, vol. 52, pp. 10934–10964, 2022.
- [15] *Deutsches Institut für Normung e.V., Safety of machinery - safety-related parts of control systems - Part 1: General principles for design*, Standard ISO 13849-1:2015, 2016.
- [16] F. Winkel and J. Deuse-Kleinstaub, "Phoenix contact relay dataset," Kaggle, Phoenix Contact Electronics GmbH, Bad Pyrmont, Germany, 2022, [Online]. Available: [kaggle.com/datasets/fabianwinkel/phoenix-contact-relay-dataset](https://kaggle.com/datasets/fabianwinkel/phoenix-contact-relay-dataset)
- [17] M. E. A. Remy, J. Desforges, and F. J. M. Rolard, "Method and device for diagnosing wear of an electrical switching unit, and electrical unit comprising such a device," U.S. Patent US15 831 541, May 12, 2017.
- [18] L. Kirschbaum, V. Robu, J. Swingler, and D. Flynn, "Prognostics for electromagnetic relays using deep learning," *IEEE Access*, vol. 10, pp. 4861–4895, 2022, doi: [10.1109/ACCESS.2022.3140645](https://doi.org/10.1109/ACCESS.2022.3140645).
- [19] J. E. V. Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, 2020.
- [20] I. Triguero, S. García, and F. Herrera, "Self-labeled techniques for semi-supervised learning: Taxonomy, software and empirical study," *Knowl. Inf. Syst.*, vol. 42, no. 2, pp. 245–284, 2015.
- [21] D. Yarowsky, "Unsupervised word sense disambiguation rivaling supervised methods," in *Proc. 33rd Annu. Meeting Assoc. Comput. Linguistics*, 1995, pp. 189–196.
- [22] D.-H. Lee et al., "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Representation Learn.*, vol. 3, no. 2, 2013, Art. no. 896.
- [23] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2020, pp. 1–8.
- [24] X. Zhang, Z. Su, X. Hu, Y. Han, and S. Wang, "Semisupervised momentum prototype network for gearbox fault diagnosis under limited labeled samples," *IEEE Trans. Ind. Inform.*, vol. 18, no. 9, pp. 6203–6213, Sep. 2022.
- [25] V. Gurevich, *Electric Relays: Principles and Applications, Ser. Electrical and Computer Engineering*. Boca Raton, FL, USA: CRC Press, 2018.
- [26] B. F. Toler, R. A. Couto, and J. W. McBride, "A review of micro-contact physics for microelectromechanical systems (MEMS) metal contact switches," *J. Micromechanics Microengineering*, vol. 23, no. 10, 2013, Art. no. 103001.
- [27] *Deutsches Institut für Normung e.V., Low-voltage switchgear and controlgear - part 5-1: Control circuit devices and switching elements - electromechanical control circuit devices*, Standard IEC 60947-5-1:2016 cor1:2016, 2018.



**FABIAN WINKEL** received the bachelor's and master's degrees in industrial engineering, in 2017 and 2019, respectively, from the Paderborn University, Paderborn, Germany, where he is currently working toward the doctoral degree in electrical engineering.

He is currently employed as a Research Associate with Phoenix Contact Electronics, Bad Pyrmont, Germany. His research focuses on applying machine learning in industrial environment, in particular to predictive maintenance applications.



**OLIVER WALLSCHIED** (Senior Member, IEEE) received the bachelor's and master's degrees (hons.) in industrial engineering and the doctorate degree (hons.) in electrical engineering from Paderborn University, Paderborn, Germany, in 2010, 2012, and 2017, respectively.

Since then, he has worked as a Senior Research Fellow with the Department of Power Electronics and Electrical Drives, Paderborn University. His research includes system identification and intelligent control methods for technical systems in the areas of power electronics, drives and decentralized grids, which also includes hybrid methods utilizing machine learning (black box approaches) and classical engineering techniques (white box approaches).



**PETER SCHOLZ** received the Dipl.-Ing. degree in electrical engineering from the University of Paderborn, Paderborn, Germany, in 2006, and the Dr.-Ing. degree in electrical engineering from the Institut für Theorie Elektromagnetischer Felder, Technische Universität Darmstadt, Darmstadt, Germany, in 2010.

From 2006 to 2008, he was with the Sensor Technology Group, University of Paderborn, where he was involved in RFID and wireless power transmission projects. Since 2011, he has been with Phoenix Contact Electronics GmbH, Bad Pyrmont, Germany, where he has been involved in the development of industrial electronics products, such as isolation amplifiers, current sensors, and electromechanical switching devices. His current research interests include numerical field computation, RF systems, and predictive maintenance.



**JOACHIM BÖCKER** (Senior Member, IEEE) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Berlin University of Technology, Berlin, Germany, in 1982 and 1988, respectively.

He is a Senior Professor and the former with the Department of Power Electronics and Electrical Drives, Paderborn University, Paderborn, Germany. From 1988 to 2001, he was the head of the control engineering team of the electrical drive systems laboratory, AEG and Daimler research. In 2001, he started his own business in the area of control engineering, electrical drives and power electronics. In 2003, he was appointed as a Full Professor and retired in 2023. His research interests include electrical drives, particularly for EVs and HEVs, energy management strategies for vehicles and smart grids, and converters for power supplies, EV chargers, and renewables.