Classification of motor vibration with machine learning methods and simulating the vibration using statistical models

Christoph Kammerer[a], Micha Küstner[a,b], Michael Gaust[b], Pascal Starke[b], Roman Radtke[a] and Alexander Jesser[a]

[a]*University of Applied Sciences Heilbronn, Max-Planck-Str. 39, 74081 Heilbronn, Germany*
[b]*CeraCon GmbH, Talstraße 2, 97990 Weikersheim, Germany*

Abstract
Reducing costs is an important part in todays buisness. Therefore manufacturers try to reduce unnecessary work processes and storage costs. Machine maintenance is a big, complex, regular process. In addition, the spare parts required for this must be kept in stock until a machine fails. In order to avoid a production breakdown in the event of an unexpected failure, more and more manufacturers rely on predictive maintenance for their machines. This enables more precise planning of necessary maintenance and repair work, as well as a precise ordering of the spare parts required for this. A large amount of past as well as current information is required to create such a predictive forecast about machines. With the classification of motors based on vibration, this paper deals with the implementation of predictive maintenance for thermal systems. There is an overview of suitable sensors and data processing methods, as well as various classification algorithms. In the end, the best sensor-algorithm combinations are shown.

Keywords
Predictive Maintenance, Industry 4.0, Internet of Things, Big Data, Industrial Internet

1. Introduction

The topic of predictive maintenance (PMA) is becoming more and more important for industrial plants and is the key topic in mechanical engineering from the Industry 4.0 aspect [1]. PMA is defined as condition-based maintenance which is carried out on the basis of a wear or service life forecast [2].

PMA uses methods that allow for individual maintenance intervals of an industrial plant to be determined and the maintenance process to be initiated automatically. As part of a R&D cooperation project between CeraCon GmbH and the Heilbronn University of Applied Sciences, a thermal system is to be set up under automation and a PMA strategy is to be implemented, which should then be adaptable to other industrial plants [1]. Due to the complexity of industrial

[1]Sponsored by ZIM ZF4644801.

plants, an intelligent solution is required in order to be able to offer individual maintenance strategies depending on the state of the plant. For this reason, the project uses machine learning (ML) methods.

The essential steps of an intelligent PMA strategy are the digital acquisition of (sensor) data, their evaluation, the analysis of the acquired data and the prediction of probable events.

First, possible component defect combinations (CDC) of the industrial plant were analyzed using standard technical risk analysis methods (FMEA, risk graph, fault tree analysis) [3]. CDC is the assignment of a wear component of the industrial system to a potentially occurring defect. Depending on the number of possible defects, a component can therefore have several CDCs. Each CDC was assigned an potential detection measure, e.g. physical vibration measurement or electrical current measurement. Suitable sensors were selected for the analyzed detection measures and analyzed with regard to the PMA strategy. CDC's with the same detection methods were combined and measurement data recorded with the respective sensors.

The core of this work is the evaluation of a combination of detection measures for data processing methods and ML algorithms. The optimal combination of these is a prerequisite for an efficient PMA strategy that can be used for the respective industrial plant.

1.1. State of the Art

A study by Bearingpoint [4] shows that PMA implementations capture 76% of the relevant data using suitable sensors, although only 59% of the process, measurement and machine data are evaluated in a targeted manner. There are three basic approaches to implementing a PMA strategy [5].

A basic approach is to use the already implemented sensors of the plant for process monitoring. This passive method is particularly suitable for systems that are already in operation. Another passive approach is to introduce dedicated sensors into the system. The additional sensors are introduced to monitor defined wear components and to detect potential defects. In the third approach, a test signal is actively fed into the system. The degree of wear of the components to be monitored can be deduced from the feedback. An example of this is Time Domain Reflectometry (TDR) [5].

2. Data Collection

2.1. Sensor Resolution

When buying industrial sensors, you often have to commit to a sensor resolution. This requires that you have a basic understanding of what accelerations occur on the component. For this purpose, the effects were previously considered in an experiment when an accelerometer with an insufficient resolution is used. In this case, the sensor generates vibrations that exceed the sensor resolution. A CDC of the fan motor is that the fan wheel has an imbalance. This fault situation was simulated by attaching an unbalance to the fan blade.

The result of this simulation is shown in figure 1 (a). There are shown the measured acceleration values in x- and y- axis of an accelleration sensor with a maximum resolution of $\pm 2G$. The red values show the vibrations of the motor without an imbalance and the blue values show
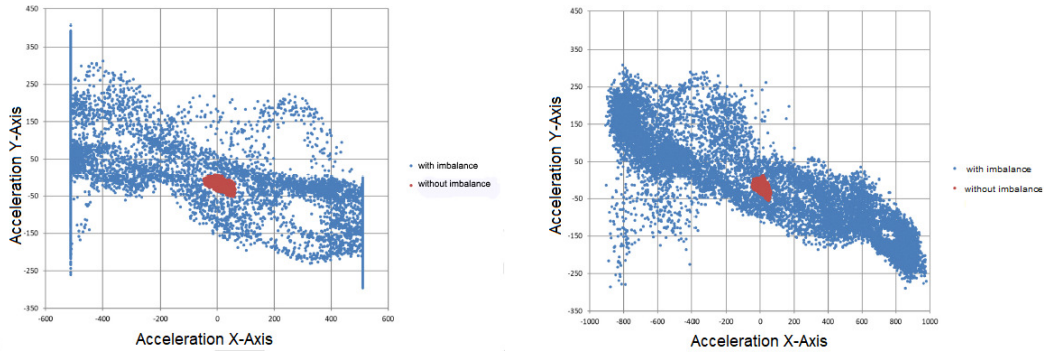
Figure 1: Comparison of an unbalanced fan with a resolution of $\pm2G$ (a) and a resolution of $\pm4G$ (b).

the vibrations which occurs with an imbalance. It can be clearly seen that the vibrations on the motor increased due to the imbalance. It can also be seen that vibrations that go beyond the set sensor resolution of $\pm2G$ were not recorded correctly. They are in line with the maximum acceleration of $\pm2G$. The measured values that did not exceed the maximum resolution were not affected by this. The experiment shows that the CDC "imbalance" can cause very strong vibrations. The vibrations are so strong that they exceed a sensor resolution of $\pm2G$. If a sensor is used that can only record values up to a resolution/acceleration of $\pm2G$, these are recorded incorrectly. The values that exceed the maximum resolution are then incorrectly saved in the data record [6]. To prevent such problems, it is important to see how large the vibrations can be. The sensor resolution should have at least this value with a safety buffer. In figure 1 (b), instead of the resolution of $\pm2G$, the double resolution of $\pm4G$ was chosen for the same motor level. In the picture you can see that no "lines" have formed and therefore the vibrations were not greater than the sensor resolution. The resolution of $\pm4G$ is therefore much more suitable than the resolution of $\pm2G$. The experiment has shown that a correctly selected sensor resolution is a prerequisite for obtaining meaningful results. If the vibrations are greater than the resolution of the sensor, the incorrectly stored measured values cannot be classified correctly [6].

2.2. Sensors and Test Set-Up

The requirement for a condition-based PMA is a structured data collection of sensor values. The following sensors were used to obtain status data:

- Three-axis acceleration sensors (Accelerometer):
    - LIS 3DH, MMA 8451, ADXL 343, ADXL 345

- Three-axis acceleration sensors with three-axis yaw rate sensor (gyroskope)
    - MPU 60.50

- Three-axis magnetic field sensor (magnetometer)
    - MLX 90393
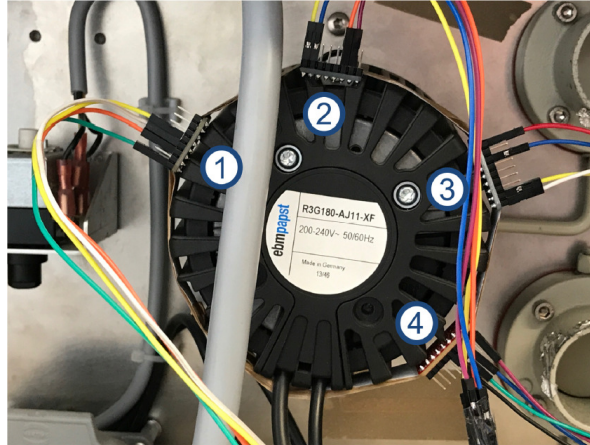
Figure 2: Measurement setup with the three-axis acceleration sensors LIS3DH, MMA 8451, ADXL 345 and ADXL 343.

Table 1
A Section of the Data measured with the MPU 60.50 Gyroskope.

60.50 AccelX	60.50 AccelY	60.50 AccelZ	60.50 GyroX	60.50 GyroY	60.50 GyroZ	Target
416	6088	60146	2709	62478	65187	Motor 000
426	6026	60148	2733	62469	65185	Motor 000
404	6110	60146	2727	62478	65192	Motor 000
470	6046	60140	2720	62486	65188	Motor 000

- Multi sensors with three-axis acceleration, three-axis yaw rate and three-axis magnetic field measurement

    – MPU 92.65, BNO 055, GY 250, GY 521

These recorded the acceleration, the rotation rate and the surrounding magnetic field of the fan motor R3G180-AJ11-XF from ebm-papst Mulfingen GmbH & Co. KG used in the thermal system. Figure 2 shows the measurement set-up with the selected three-axis acceleration sensors [6].

The fan motor was operated at fixed speeds, which were divided into 7 classes. This classification was based on the specific values 0%, 50%, 60%, 70%, 80%, 90% and 100% of the maximum engine speed. During the operation of the fan motor the vibration of the crankcase was sensed and recorded by the sensors. More than 980,000 structured sensor data sets per measurement series and sensor type were recorded. A total of more than 2.6 million data sets have thus been recorded for all sensor types. A section of a full data set is shown in Table 1.

An example of a recorded data set is shown in figure 3. It shows the measured acceleration from the housing vibration in the spatial x- and z- orientation. The individual classes are highlighted in color to make a distinction possible. Due to the highest spatial coverage, it can be seen that the measurement results for class 80% can be assigned to the resonance range of the fan motor, since the acceleration values in the x- and z- alignment are at their maximum values
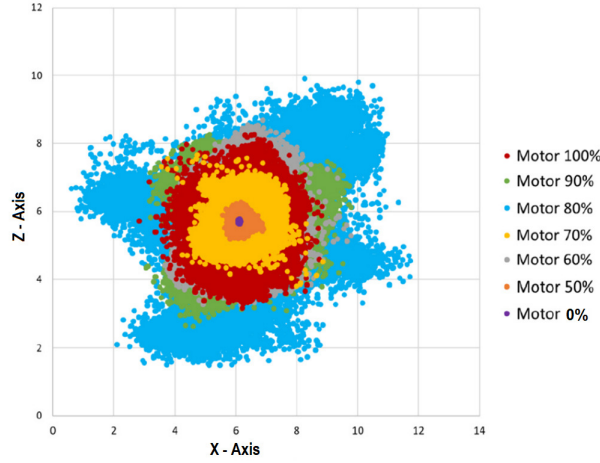
Figure 3: Measurement results from the acceleration sensor MMA 8451.

here.

3. Data Conditioning

In order to be able to better differentiate the individual classes, it is in some cases advantageous if the data records are processed before classification. The methods used for data conditioning are presented here:

One possibility to process the data sets consists of the differencing and absolute value formation of subsequent values according to equation 1.

$$X_i = \quad |X_i - X_{i+1}| \tag{1}$$

X_i and X_{i+1} are the successive sensor values. Another processing method is the integration of the data according to equation 2. Here the area under two successive values X_i and X_{i+1} is calculated.

$$X_i = \begin{cases} X_i + 0.5 \cdot (X_{i+1} - X_i) & \text{if } X_i < X_{i+1} \\ X_i - 0.5 \cdot (X_i - X_{i+1}) & \text{if } X_i > X_{i+1} \\ X_i & \text{if } X_i = X_{i+1} \end{cases} \tag{2}$$

In both the processing methods, an additional smoothing can be carried out by calculating the moving average according to equation 3.

$$X_i = \frac{1}{G} \cdot \sum_{i-g}^{i+g} X_i \tag{3}$$

The parameter G specifies the degree of smoothing. The parameter g is the difference between the indices between the instantaneous value X_i and the maximum value $X_{g \pm i}$ specified by

47
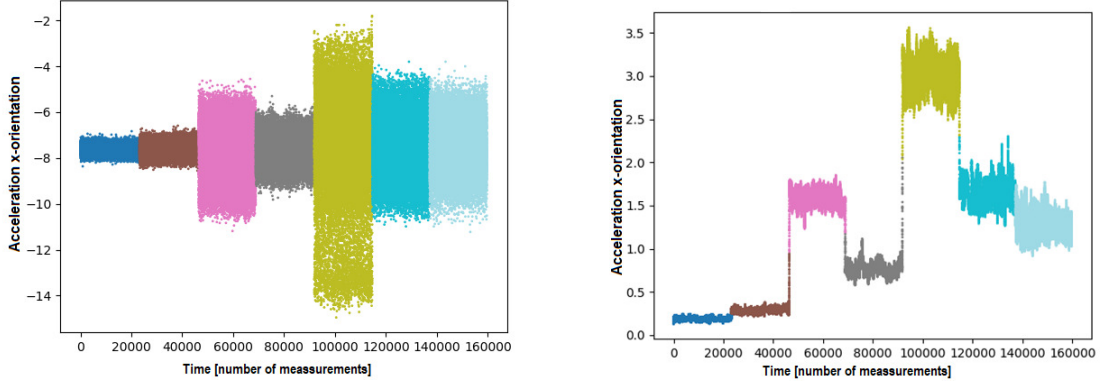
Figure 4: Comparison of raw data (a) and data prepared by differencing (b).

the degree of smoothing. Thus g depends on the degree of smoothing G and can be determined according to equation 4.

$$g = \frac{G - 1}{2} \tag{4}$$

With the degree of smoothing G, first optimizations regarding the classification of the measured sensor data can be carried out [7]. Figure 4 shows the effect of processing by means of differencing compared to the unprocessed raw data. The coloring in the pictures illustrates the different class assignments.

Figure 4 (a) shows the acquired raw data of an acceleration sensor in the x-orientation. It can be seen that a delimitation regarding the classes is not clear. For example, the acceleration in the direction of the x-axis at $-8\,m/s^2$ is not unique and can in principle be assigned to any class. In Figure 4 (b) the classes are more delimited after the differencing and smoothing and thus a class assignment is clearer. For example, the value 0.7 can be clearly assigned to the class shown in gray.

Figure 5 (a) shows the raw data from the measurements in x- and y-orientation. A classification is clearly not possible due to the overlapping point clouds. Figure 5 (b), on the other hand, shows the data prepared after the differencing. It can be seen that the point clouds are now clearly distinguishable, making visual and algorithmic class assignment easier.

4. Evaluation

The evaluation of the ML algorithms with regard to the respective sensors and the data processing was divided into a training and a test phase. In the training phase, the data records were divided evenly by feeding every tenth data value of the respective training method to the ML algorithm. As a result, the respective ML algorithm was trained with 10% of the data. The complete data set was then evaluated in the test phase. Several ML algorithms were considered for the recorded data sets. These included decision trees [8, 9], the gradient boost method
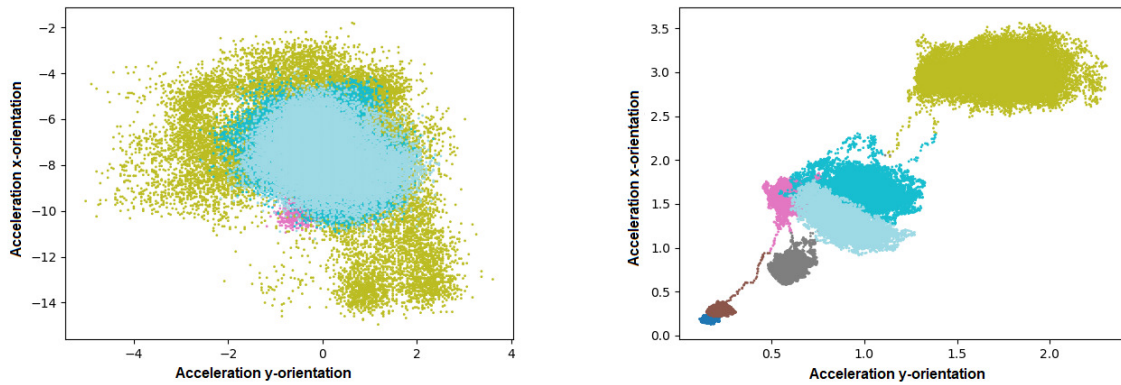
Figure 5: Comparison of raw data (a) and data prepared by differencing (b) in two axes.
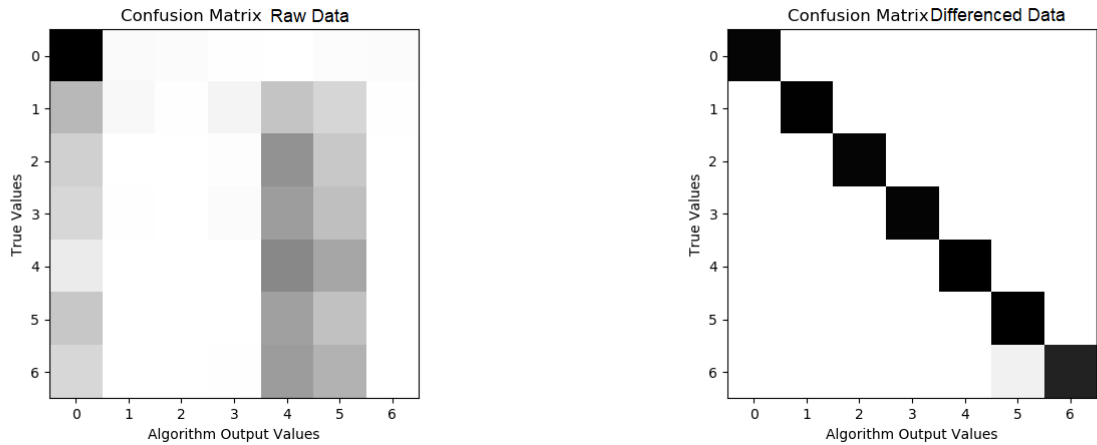
Figure 6: Comparison of the prediction results of the focus cluster algorithm with raw data (a) and data prepared by differencing (b).

[8, 10], a focus cluster algorithm [11] and artificial neural networks (ANN) [12]. The investigations revealed that ANNs are less suitable for data sets with low attribute numbers due to the long duration in the training phase. Therefore, only the decision trees, the gradient boost method and the focus cluster algorithm were used for the further experiments [7].

Figure 6 shows two confusion matrices [13] for the focus cluster algorithm, which show the distribution between the actual class and the class determined by the algorithm. The numbers on the axes correspond to the seven defined classes in which the data records have been categorized. The darker an area, the more often the ML algorithm has assigned data records to a class. A correct assignment is obtained if the assigned class corresponds to the actual class. Ideally, you would get a black diagonal from top left to bottom right.

Figure 6 (b) shows the result for the data sets prepared after differencing and smoothing. It

can be seen that the majority of the data records were assigned to the actual classes, the hit rate here was over 98%. On the other hand, it can be seen in Figure 6 (a) that a significantly lower hit rate has been achieved for the unprepared data sets.

5. Using Statistical Methods for Prediction

The preceding test were performed by measuring the vibration of a stand-alone motor on a workbench, which was not build into a working machine. Therefore this data cannot be used to make a prediction for maintenance, but the feasibility of categorizing a motor by its vibration and magnetic field was studied. To get a better picture of the real working conditions of such a motor a larger data set was collected by mounting three multifunction sensors (GY 521, BNO 055, MPU 92.65) on a thermal system which is used in day to day operations. These data sets were then used to build a statistical model based on auto regression and moving average (ARMA [14], [15]) of the vibration. The statistical models were created for every sensor orientation separately to get an optimal result for each time series. As the metric to compare the different models was chosen the maximum relative deviation (MRD) according to equation 5.

$$MRD = \max_i \left| \frac{X_i - Y_i}{X_i} \right| \tag{5}$$

In equation 5 X_i are the measured sensor values used to build the model and Y_i are the values predicted by the model at this time-step.

6. Results

To compare the results of the ML algorithms for the respective sensors, a matrix with the relevant properties was created for each combination of ML algorithm and sensor:

- Classification accuracy (performance) of the algorithms

- Computing time for the training and testing phase of the algorithms

- Smoothing factor G

Table 2 shows the comparison of the processing methods examined using the combination of the multifunction sensor GY 521 and the gradient boost method.

It can be seen that the highest performance is achieved when using the raw data. In the training and test phases the integrated and the differenced data are slightly faster. A comparison was made for each sensor and ML algorithm combination. The best performing data processing method was then selected for each combination. The comparison tables of the sensors which have achieved the best results of the sensor types examined are listed in Tables 3 to 5. These were the ADXL 345 (accelerometer), the MPU 60.50 (gyroscope) and the GY 521 (accelerometer, gyroscope and magnetic field). The processing method with the highest performance for the respective algorithm is shown for each of the sensors.

Table 2

Comparison of the data preparation methods with the combination of the multifunction sensor GY521 and the gradient boost method.

Sensor GY 521 with Gradient Boost method	Performance (%)		Time training (s)		Time testing (s)	
	$G = 0$	$G = 99$	$G = 0$	$G = 99$	$G = 0$	$G = 99$
Raw data	98.51	83.97	4.85	6.88	1.11	0.9
Integration	82.84	85.15	5.45	3.82	1.07	0.66
Differencing	79.76	85.1	5.04	3.83	1.5	0.61

Table 3

Comparison of the ML algorithms on the combination ADXL 345 (accelerometer) and the best performing data processing method.

ADXL 345	Decision tree	Gradient boost method	Focus cluster algorithm
Conditioning method	Differencing, $G = 99$	Differencing, $G = 99$	Differencing, $G = 99$
Performance	93.79%	96.6%	98.98%
Computing time training	1.78s	5.88s	0.32s
Computing time testing	31.99s	1.44s	34.88s

Table 4

Comparison of the ML algorithms on the combination MPU 60.50 (gyroskope) and the best performing data processing method.

MPU 60.50	Decision tree	Gradient boost method	Focus cluster algorithm
Conditioning method	Raw data, $G = 0$	Raw data, $G = 0$	Differencing, $G = 99$
Performance	91.46%	95.24%	90.89%
Computing time training	0.09s	5.54s	0.91s
Computing time testing	0.28s	1.46s	31.99s

The result of the examinations according to Table 3 was that all acceleration sensors achieved the greatest performance with smoothing ($G = 99$) and data prepared by differencing. The focus cluster algorithm achieved the highest performance.

The result according to Table 4 is that the highest performance was achieved with unprocessed and unsmoothed data with the gyroscopes. The gradient boost process achieved the highest performance.

In the case of the multifunction sensors with acceleration, magnetic field sensors and gyroscope, it can be seen from Table 5 that the highest performance was achieved with smoothing ($G = 99$) and differenced data using the cluster cluster algorithm.

Table 6 shows the lowest MRDs for each separate sensor orientation with their number of auto regressive (p) and moving average (q) terms.

These results show a maximum deviation up 28.3% with the gyroscope values in z orientation. The lowest deviation was reached with acceleration in z orientation and the magnetic field in x orientation with only 10.1% and 10.6% respectively. The difference in accuracy of the

Table 5
Comparison of the ML algorithms on the combination GY 521 and the best performing data processing method.

GY 521	Decision tree	Gradient boost method	Focus cluster algorithm
Conditioning method	Raw data, $G = 0$	Raw data, $G = 0$	Differencing, $G = 99$
Performance	95.7%	98.51%	99.89%
Computing time training	0.09s	4.85s	1.27s
Computing time testing	0.28s	1.11s	37.63s

Table 6
Lowest MRD for each Sensor Orientation with their corresponding p and q Numbers.

Sensor Orientation	p	q	MRD (%)
Acc X	5	5	20.3
Acc Y	3	3	24.4
Acc Z	2	5	10.1
Gyro X	7	5	22.4
Gyro Y	4	3	24.7
Gyro Z	7	6	28.3
Mag X	4	5	10.6
Mag Y	3	5	22.6
Mag Z	7	7	28.2

models is rather high and therefore an ARMA model can only be used to model two of the nine time series measured. For the remaining seven there should be used other means of modeling.

7. Conclusion

It has been found that the processing of the raw data in the form of smoothing and differencing in combination with the focus cluster algorithm gave the best results for acceleration sensors. The gyroscopes examined showed that the unprocessed raw data without smoothing in combination with the gradient boost method achieved the highest classifiability. The multisensors examined gave the best results when using the focus cluster algorithm in combination with smoothed and differenced data. In addition it was found, that an ARMA model could be used to predict the acceleration in z orientation and the magnetic field in x orientation.

8. Outlook

Based on these results, the combination of detection measure, data processing method and ML algorithm can in the next step be used for a PMA strategy. For a complete PMA, further detection measures have to be examined. For that purpose, this procedure is continued with further sensor types in order to find an optimal combination for all necessary detection measures. In the future, a prediction model is to be developed on the basis of these results, with

which predictions can be made about the degree of wear of system components of a thermal system under automation. Formal aging and error models of the respective system components must also be created in order to map the aging process of components. These models can then be used to make probabilistic statements about the failure probabilities of the individual assemblies. Such models could be based on Dynamic Bayesian Networks (DBN) [16], auto regresssion and moving average (ARMA) [17] or, as the focus cluster algorithm has yielded such an high performance, a multi dimensional focus trajectory. In addition to that, the statistical models used to predict the motor vibration in day to day operations could be extended to auto regression, integrated, moving average to get a better result for all sensor orientations.

References

[1] S. Feldmann, R. Lässig, O. Herweg, H. Rauen, P.-M. Synek, Predictive Maintenance: Servicing tomorrow ? and where we are really at today, Technical Report, Roland Berger GmbH, München, 2017. URL: https://www.vdma.org/documents/105806/17180011/VDMA+Predictive+Maintenance+englisch.pdf.

[2] Din en 13306:2018-02: Instandhaltung - begriffe der instandhaltung; dreisprachige fassung en 13306:2017, 2018. URL: https://www.beuth.de/en/standard/din-en-13306/270274780. doi:10.31030/2641990.

[3] B. Ebert, Prozessoptimierung bei Industrie 4.0 durch Risikoanalysen: Gefährdungen erkennen und minimieren, Springer Vieweg, Berlin, Heidelberg, 2018. doi:10.1007/978-3-662-55729-7.

[4] F. Duscheck, R. Blameuser, S. Gehrmann, Consulting, Solutions and Ventures | BearingPoint, 2018. URL: https://www.bearingpoint.com.

[5] H. M. Hashemian, W. C. Bean, State-of-the-art predictive maintenance techniques*, IEEE Transactions on Instrumentation and Measurement 60 (2011) 3480–3492. doi:10.1109/TIM.2009.2036347.

[6] M. Küstner, Auswahl und bewertung eines sensorkonzepts für die implementierung einer predictive maintenance zur temperaturbehandlung unter automation, 2019.

[7] M. Gaust, Entwurf und implementierung von maschinellen lernalgorithmen zur klassifikation von maschinendaten für eine vorausschauende wartung von industriethermosystemen, 2019.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, Édouard Duchesnay, Scikit-learn: Machine learning in python, Journal of Machine Learning Research 12 (2011) 2825–2830. URL: http://jmlr.org/papers/v12/pedregosa11a.html.

[9] T. Thomas, A. P. Vijayaraghavan, S. Emmanuel, Applications of Decision Trees, Springer Singapore, Singapore, 2020, pp. 157–184. URL: https://doi.org/10.1007/978-981-15-1706-8_9. doi:10.1007/978-981-15-1706-8_9.

[10] A. Binding, N. Dykeman, S. Pang, Machine learning predictive maintenance on data in the wild, in: 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), 2019, pp. 507–512. doi:10.1109/WF-IoT.2019.8767312.

[11] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems, 1st ed., O'Reilly Media, Boston, 2017.

[12] T. Rashid, Make Your Own Neural Network, 1st ed., CreateSpace Independent Publishing Platform, North Charleston, SC, USA, 2016.

[13] M. Füllsack, A (partially) interactive introduction to Systems Sciences, 2013. URL: http://systems-sciences.uni-graz.at/etextbook/bigdata/confusionmatrix.html.

[14] C. Lin, Y. Hsieh, F. Cheng, H. Huang, M. Adnan, Time series prediction algorithm for intelligent predictive maintenance, IEEE Robotics and Automation Letters 4 (2019) 2807–2814. doi:10.1109/LRA.2019.2918684.

[15] H. Akaike, Maximum likelihood identification of gaussian autoregressive moving average models, Biometrika 60 (1973) 255–265. URL: http://www.jstor.org/stable/2334537.

[16] P. Dagum, A. Galper, E. Horvitz, Dynamic network models for forecasting, in: D. Dubois, M. P. Wellman, B. D'Ambrosio, P. Smets (Eds.), Uncertainty in Artificial Intelligence, Morgan Kaufmann, 1992, pp. 41–48. URL: https://www.sciencedirect.com/science/article/pii/B9781483282879500104. doi:https://doi.org/10.1016/B978-1-4832-8287-9.50010-4.

[17] J. Wang, Y. Liang, Y. Zheng, R. X. Gao, F. Zhang, An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples, Renewable Energy 145 (2020) 642–650. URL: https://www.sciencedirect.com/science/article/pii/S0960148119309371. doi:https://doi.org/10.1016/j.renene.2019.06.103.