

Article

Machine Tool Component Health Identification with Unsupervised Learning

Thomas Gittler ^{1,*}, Stephan Scholze ², Alisa Rupenyan ³ and Konrad Wegener ¹

¹ Institute of Machine Tools and Manufacturing (IWF), ETH Zürich, CH-8092 Zurich, Switzerland; wegener@iwf.mavt.ethz.ch

² Agathon AG, CH-4512 Bellach, Switzerland; stephan.scholze@agathon.ch

³ Inspire AG, ETH Zürich, CH-8005 Zurich, Switzerland; rupenyan@inspire.ethz.ch

* Correspondence: gittler@inspire.ethz.ch; Tel.: +41-(0)44-632-5252

Received: 31 July 2020; Accepted: 31 August 2020; Published: 2 September 2020



Abstract: Unforeseen machine tool component failures cause considerable losses. This study presents a new approach to unsupervised machine component condition identification. It uses test cycle data of machine components in healthy and various faulty conditions for modelling. The novelty in the approach consists of the time series representation as features, the filtering of the features for statistical significance, and the use of this feature representation to train a clustering model. The benefit in the proposed approach is its small engineering effort, the potential for automation, the small amount of data necessary for training and updating the model, and the potential to distinguish between multiple known and unknown conditions. Online measurements on machines in unknown conditions are performed to predict the component condition with the aid of the trained model. The approach was exemplarily tested and verified on different healthy and faulty states of a grinding machine axis. For the accurate classification of the component condition, different clustering algorithms were evaluated and compared. The proposed solution demonstrated encouraging results as it accurately classified the component condition. It requires little data, is straightforward to implement and update, and is able to precisely differentiate minor differences of faults in test cycle time series.

Keywords: condition monitoring; machine learning; prognostics and health monitoring; unsupervised learning; machine tools; manufacturing

1. Introduction

Failures and unplanned maintenance of machine tools cause severe productivity losses. As a remedy, Kusiak [1] proposes a vision of the smart factory, in which monitoring and prediction of the health status of systems prevent faults from occurring. A prerequisite for the monitoring of equipment is the synergy of operational technology (OT) and information technology (IT). It is often described as a cyber-physical system, which is a key research element of the smart factory [2,3]. For this cyber-physical manufacturing of the future, Panetto et al. [4] have identified four grand challenges, of which two relate to the operational availability of machine tools: resilient digital manufacturing networks, and data analytics for decision support. More precisely, the required applications in view of machine tools comprise tools for monitoring disruptions, prescriptive and predictive modelling, as well as risk analysis and control.

In this context, this study presents a new prognostics and health management (PHM) approach for machine tool components. It allows faults, critical states or deviations from a healthy behaviour to be detected. Most current approaches model the healthy states of the components. Deviations from the healthy states are then identified as potential failure causes. However, the breakdown reasons and their characteristics with respect to different failure types remain unknown. The proposed approach by

contrast identifies the type of fault that is present or likely to occur on a component. This is achieved by comparing a test cycle sensor signal with previously observed or recreated fault states of a machine component. To do so, the concept suggests transforming the sensor data time series of the test cycle into a representation of features. The features are different time series characteristics, such as e.g., Fourier or continuous wavelet transforms. To allow a generalist approach that can be applied to any type of component and test cycle data format, a large number of more than 700 features are calculated, before deciding which are retained. To detect differences in test cycles of different health or failure states, the features need to allow a clear distinction. All features with low significance, i.e., strong overlap of feature values for different conditions, are discarded. Based on this cleaned feature representation, previously recorded healthy or failure states can be grouped in clusters of their feature values. This model, consisting of selected features and grouped clusters of different healthy and faulty conditions, serves for the further predictive assessment of machine components in unknown conditions. To analyse a component in an unknown condition, it needs to execute an identical test cycle, for which the same features are calculated. The proximity of the feature values to previously recorded healthy or faulty conditions allows the state of the currently analysed component to be determined. As only the features with higher statistical significance are retained, even minor differences can be represented in the combination of multiple features. However, the larger the number of features, the higher the dimensionality of the clustering model, which introduces additional requirements for the selection of the clustering algorithm. Moreover, the clustering model needs to distinguish between healthy, faulty and previously unknown (neither healthy nor a known fault) conditions. To fulfil this aspect, different partitioning and clustering algorithms were evaluated, of which hierarchical density-based spatial clustering of applications with noise (HDBSCAN) managed to meet all requirements and showed the best performance. To obtain the necessary data for the different component conditions, faulty states were recreated artificially for model training by the head service technician expert of the machine Original Equipment Manufacturer (OEM) on which the tests and data collection were conducted. As the study is of exploratory nature to examine the feasibility of the proposed approach, the artificially introduced faults serve as the basis to evaluate its performance. Further research will be undertaken into a large-scale test and its applicability to a fleet of machines. The novelty of the proposed approach lies within (i) the representation of time series for condition monitoring as features for clustering, (ii) the fact that raw values of selected features are used rather than e.g., principal component analysis (PCA), (iii) the detection of both formerly known and unknown conditions of a component, and (iv) the universal applicability of the approach to different natures (constant, controlled-constant and varying) and types (linear, rotatory) of components. Advantages (i) and (iv) reduce the engineering effort in the implementation, (ii) retain the physical interpretability of the calculated features and the clustering results, and (iii) allow the proposed solution to be used with incomplete information and update it with growing data sets.

According to Choudhary et al. [5], the data-driven knowledge discovery process consists of domain understanding, raw data collection, data cleaning and transformation, model building and testing, implementation, feedback and final solution, and solution integration and storage. This study focuses on the steps related to domain understanding, raw data collection, and emphasizes especially data cleaning and transformation, and model building and testing.

2. Related Work

2.1. Failure Detection and Prognostics and Health Management (PHM) Applications in Machine Tools

Andhare et al. stipulate that more than 50% of common machine tool failures are due to component damage or looseness [6]. To prevent downtimes, PHM applications supervise, detect and anticipate machine and component behaviour. According to Tao et al. [7], increasing availability of both, measurement data and advanced algorithms stimulate the application of machine learning approaches in PHM. Equipping machines with the cognition to detect their health status autonomously follows

the paradigm of biologicalization, which seeks to mimic human and natural traits of intelligence in manufacturing systems, according to Wegener et al. [8]. Supervision of machine tool components is possible via modelling of their behaviour in the healthy state, and subsequently detecting anomalies during further operation, as shown e.g., by Sobie et al. and Ruiz-Carcel and Starr [9,10]. Often, faults and failures are typically not unidimensional, but the result of multiple colluding or simultaneous degradations. Most PHM approaches apply a binary distinction between health and failure states, without consideration of the various faults and their severity. These fault types have different impacts on the usability of the machine, depending on the process and the users' requirements. Therefore, not only does the presence of anomalies but also the different types and severities of faults on machine tool components need to be identified. A multi-dimensional health assessment allows the impact a degradation can have on a production process or a final product to be revealed. Besides an accurate assessment, challenges are the data gathering and modelling effort for different faulty states, as well as the reproducibility and applicability to different machine and component types.

Machine tool failures depend on a multitude of influences. Internal variances (thermal and dynamic behaviour, manufacturing and assembly of components) and external factors (surrounding and environmental influences, usage and maintenance) make faults appear stochastic. These influences are cumbersome to reproduce in purely physical models approaches, wherefore many recently published PHM approaches in manufacturing incorporate statistical models. Prominent examples for the application of data driven models in monitoring are described by e.g., [10–15], relevant studies on data-based approaches for prognosis are described by [9,16–18]. Both the PHM approach, as well as the applied learning algorithm strongly impact the capabilities and performance of the application. Comprehensive overviews of learning and data mining techniques for manufacturing are provided by Wuest et al. [19] and Choudhary et al. [5], of which the described clustering approaches are used for this approach.

The field of prominent representatives of PHM applications in machine monitoring apply supervised learning algorithms, as described in comprehensive overviews by Gao et al. and Zao et al. [20,21]. As an example, Malhotra et al. [18] model the healthy state to subsequently detect anomalies with recurrent neural networks (RNN). Sequences of a healthy state are trained on a long short-term memory (LSTM) encoder-decoder, in order to obtain a degradation indication. The degradation curves are matched to other failure curves, in order to estimate the remaining useful lifetime (RUL). Reference [14] extracts features from volumetric errors (VE) on a five-axis machine tool via fractal analysis, to recognize changes in VEs as degradations. Duan et al. apply an auto-regression on multivariate numerical control (NC) signals of circular machine tool tests, where residuals due to anomalies are used to model the machine state as a semi-Markov Process [22]. Malhotra's and most other PHM approaches rely on simulated degradation for model training, as it is also the case e.g., for Sobie et al. [9] and Xing et al. [14]. They conclude that PHM models trained on simulated degradation data show an inferior performance to those trained on real machine data in a comparative study.

Overall, supervised algorithms allow differences from healthy behaviour of components in an unknown condition to be quantified. The indication of a deviation from a previously defined healthy state however lacks the description of the fault dimension or type. As each individual fault requires a corresponding data set for learning or classification, simultaneously designating the deviation and the fault type is a challenge. Moreover, component behaviour outside of the training or learned cases is challenging to detect and label for supervised approaches. Due to the inherent input–output relationship of supervised models, noise, outliers and inaccurate data have a strong adverse impact. Filling these gaps with simulated data has the disadvantage of inferior performance as pointed out by Sobie et al. [9]. Unsupervised algorithms can be applied to detect deviations from a collection of previously observed healthy states, and equally consider a priori known faulty states. The issue of incorrectly labelled data is irrelevant to unsupervised models, and they exhibit a higher robustness to noisy data, as outlined by Zhang et al. [23]. They published an unsupervised machining process supervision called AnomDB. It is an outlier detection framework for NC data, in which a PCA is

applied to a multivariate time series prior to feature extraction, followed by a density-based spatial clustering of applications with noise (DBSCAN). Zhang et al. showed a superior performance of their proposal compared to other unsupervised approaches.

In conclusion, unsupervised approaches show promising potential for machine tool supervision. However, their abilities to cope with noisy and multivariate data for PHM remains to be examined. Density-based clustering algorithms have shown superior outlier detection as compared to other clustering methods in these applications, as Zhang et al. demonstrated. On the downside, the anomalies were introduced synthetically, and their approach lacks an interpretability of the features due to the prior PCA performed on the features. Similarly, the distinction between known and unknown anomaly types, and the applicability of unsupervised algorithms to component supervision with real machine data needs to be proven.

2.2. Learning Algorithms for PHM Applications

Unsupervised learning algorithms differ significantly in view of clustering capabilities (e.g., accommodation of varying cluster shapes, sizes and densities, as well as the ability to cope with noisy data), and the amount of a priori required hyperparameters or assumptions for initialization. For the proposed approach, the following requirements need to be met: For performance, the algorithm must be computationally efficient. The attribution of samples to a cluster needs to be provided with an uncertainty measure, to detect and avoid false classifications. It needs to accommodate clusters of different shapes, which can be non-hyperspherical, or even non-convex. For the detection of unknown states and noise, the algorithm needs to distinguish if a sample belongs to an existing or a new, a priori unobserved cluster. To avoid heuristic tuning of hyperparameters, both the number of clusters, as well as other hyperparameters (e.g., maximum distance of neighbouring points) need to be inferred by the algorithm.

Finally, the number of samples per observed state will vary significantly, as observations of healthy axes typically dominate observations of failure states. Hence, the algorithm must be robust towards strong variance in cluster densities and sizes. Four state of the art clustering algorithms are compared in terms of their viability of PHM applications in machine tools: k-means [24], Gaussian mixture models (GMM) [25], DBSCAN [26] and hierarchical DBSCAN (HDBSCAN) [27].

2.2.1. k-Means

k-Means is a partitioning algorithm originally presented by MacQueen [28], which divides an n -dimensional space of data points into k distinct regions. Each partition k is defined by all points within the region and represented by its mean. The algorithm seeks to minimise the average squared distance between points in the distinct clusters. According to Arthur and Vassilvitskii [24], k-means can be designed in a computationally efficient way, but it has a number of disadvantages: (1) the algorithm attributes each data point to a cluster, it cannot designate noise or new clusters. (2) Following from its attribution rule, the cluster shape is assumed to be hyperspherical. (3) Attributed data points are provided without a measure of uncertainty for points lying further away from the cluster mean. (4) The number of clusters k has to be set in advance, it cannot be inferred by the algorithm. Some shortcomings can be overcome by modifications of the k-means algorithm, but the assumption of a globular cluster shape remains inevitable. Therefore, the predictive attribution of data points with high uncertainty or noise inhibits a risk of false positive classifications.

2.2.2. Gaussian Mixture Model (GMM)

Some of the shortcomings of k-means are addressed by GMMs, which model clusters as normal distributions around a mean, and expresses cluster attribution for a point as a probability. Hence, it inherently provides the uncertainty measure k-means lacks, and can identify points with low attribution probabilities as outliers. Through its probability-based cluster description, cluster shapes are not limited to globular shapes. While GMM addresses some issues of k-means, it still preserves

other disadvantages, according to McLachlan et al. [25]: (1) similar to k-means, the parameter k cannot be inferred by the algorithm itself. (2) The algorithm cannot represent more complex non-convex cluster shapes.

2.2.3. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN is a non-probabilistic algorithm and assumes clusters to be regions of high sample density [26]. It identifies clusters of any shape, as no prior shape assumptions are maintained. Moreover, it is able to infer the number of clusters itself and, therefore, resolves the downsides of k-means and GMM. Unfortunately, DBSCAN performs poorly on clusters with varying density, as the neighbour count threshold is a fixed parameter. McInnes et al. extended DBSCAN to a hierarchical algorithm (HDBSCAN), retaining the advantages of DBSCAN by inferring cluster sizes via the union of neighbouring clusters sharing a similar hierarchical. This detaches the cluster attribution from its shape and points distribution, resolving the problem of handling varying cluster densities [27]. Moreover, outliers, lying by definition in sparse regions, are not clustered by HDBSCAN. They are identified and marked as so-called *noise-points*, which are not attributed to any existing cluster. Overall, HDBSCAN performs well with outliers and noisy data sets, and has the ability to handle varying cluster densities, making it a suitable candidate for time-series feature based component state identification. An overview of the requirements for the proposed approach and the degree of fulfilment of the presented algorithms is shown in Table 1.

Table 1. Qualitative capability comparison of selected clustering algorithms.

	k-Means	GMM	DBSCAN	HDBSCAN
Computationally efficient	☐	●	●	●
Provision of uncertainty measure	☐	●	●	●
Non-hyperspherical clusters	○	●	●	●
Recognition of noise or emerging clusters	○	☐	●	●
Accommodation of non-convex clusters	○	○	●	●
Inference of number of clusters	○	○	●	●
Complete hyperparameter inference	○	○	☐	●
Accommodation of varying cluster densities	○	○	○	●

Meaning of symbol annotation: ○—incapable, ☐—capable with modifications, ●—capable.

3. Materials and Methods

As the method is designed according to a conventional data science approach, this section is structured as follows:

- (1) Data acquisition: the preparation of the machine component, the test cycle design and the necessary data to be acquired and their format are described.
- (2) Data pre-processing: after the data are acquired, their parsing, cleaning and treatment to prepare them for model construction and training are detailed.
- (3) Model creation: the cleaned and prepared data of the training set are fed to a clustering algorithm to train a model.
- (4) Model deployment: the constructed model is used evaluated on the test data set, and furthermore used as a predictor for prior unknown data sets. The update and maintenance of the model is outlined as well.
- (5) Advantages over the state of the art: the differentiation and novelty of the proposed approach are highlighted, in order to allow a comparison with related studies.

3.1. Data Acquisition

On an arbitrary machine tool component, a test cycle is conducted outside of machining times and without a work piece engaged. This ensures comparable preconditions for data generation and

acquisition. The test cycles for model training and the use of the model for predictions are identical. Each component of a machine is analysed separately, the measurement and modelling process remains the same for all machine components. In this study, the approach is demonstrated exemplarily for machine axes. For each axis, the data of test cycles of both healthy and different faulty states are collected. Faulty conditions can be recreated by artificially introducing mechanical or electrical faults that reproduce the dynamics of a critical behaviour. In an exemplary case, common faults like excessive friction, mechanical defects, pretension loss and wear are used as representative fault types to be detected. The component prepared in both healthy and faulty conditions executes a test cycle trajectory: A translatory axis is moved from one end to the other, and back to its initial start position. Similarly, a rotatory axis is turned from start to its outward movement limit and back to its start position. The trajectory consists of 4 segments in each direction: an acceleration ramp and its transient response, a constant velocity segment, a deceleration ramp until complete halt and its transient response, and the constant holding in the following position. All of these segments show different aspects of the component's dynamic behaviour, allowing it to incorporate a high information density in the test cycle data. As the segments are recorded for both (+) movement or clockwise direction and (−) movement or counter clockwise direction, a total of 8 different segments are recorded in each test cycle. They are referred to as regions of interest (ROI). The test cycles are executed with the common process dynamics and velocities of the machine component in operation, in order to recreate operating conditions for the detection and quantification of anomalies. Furthermore, the test cycles are repeated multiple times to minimize variance over the samples and to enable the detection of outliers in the recordings. The test cycle data are acquired directly by the component drive or the NC of the machine with high sampling rates. Higher sampling rates allow to detect faults with high-frequency oscillations of mechanics and control feedback loop signals while satisfying the Shannon–Nyquist theorem. This is especially important for highly rigid structures, short axis travels, low inertia of moving parts or high axis dynamics, in which faults tend to translate into higher frequency oscillations of mechanics and control loop feedback signals.

3.2. Data Pre-Processing

The resulting data set is split in a test set and a training set, in order to both train and evaluate the model. During model deployment for prediction, the model is applied to test cycle data of machine axes in unknown condition to assess their health status. The status is described as either healthy condition, similar to a known faulty condition, or unknown (neither healthy nor a known faulty state).

Figure 1 provides an overview of the solution structure, with a focus on data processing: For the analysis of the measurement data, the current signals of the component's control loop are used, as a representative for the resulting force or torque. Preliminary filtering for poor signal accuracy, for outliers of test cycle duration, for sampling rate inconsistencies and for other anomalies is conducted. (1) Since the axes exhibit different behaviours for different conditions, e.g., lag in force or position signal due to mechanical play, a precise synchronization of the test cycle data is crucial. The current signal is best synchronized on feed forward rather than feedback signals. The test cycle current signal time series are segmented into the ROIs beforehand for separate analysis. Each ROI represents different dynamics, responses and, therefore, potential fault characteristics of the component, whereby a separation is necessary. The ROIs are treated as independent time series for data analysis, their results are merged in a later step. (2) To make the sampled, synchronized and segmented force signal time series comparable, features describing the relevant time series characteristics are extracted. The considered feature extraction approaches are e.g., fast Fourier transform (FFT), continuous wavelet transform (CWT), autocorrelation, or approximate entropy, which are each calculated with various parameter sets. The feature extractions are calculated for all possible parameter sets for each ROI, before irrelevant and insignificant features are filtered and discarded. This allows to extract a different set of features for each ROI, as the significance of a single feature for a specific ROI is higher than that of the same feature for the entire test cycle. In practice, a component with a loose motor may exhibit a behaviour similar

to healthy axis when held still or moved at a constant velocity (ROIs 2,4,6,8), but it is significantly different during, acceleration, braking and inversion (ROIs 1,3,5,7). For a component with signs of excessive friction, the exact opposite may be the case. The extraction of n_m features of all of m ROIs transforms the time series into a higher dimensional feature space, with all features constructing a vector \underline{v} of rank $n_{Total} = \sum_1^m n_m$. The corresponding feature values $v_n \in \underline{v}$ describe the time series as a point in an n -dimensional space. After calculation of all features per ROI, the features are normalized (3). As some faulty components show an extreme behaviour e.g., in vibrations, their features would distort the scaled distribution when using a standard mean or a min-max scaler. Hence, a robust scaler less susceptible to outliers and variance is used. Subsequently, multiple filters are applied to retain only those features allowing conditions to be distinguished from one another, reducing the dimensionality of the feature vector \underline{v} . First, features are filtered for statistical significance by p -value. Second, a filter for variance and kurtosis of features within samples of the same condition is applied—the variance filter removes features of which the values for the same condition negatively impact clustering due to the broad distribution. The kurtosis filter allows outliers to be filtered for, by opting for features with a flat-tailed distribution. A third filter discards highly correlated features to avoid bias. Overall, the filters are intended to remove unwanted stochastic influences during test cycles, introduced both by variance in the execution of the test cycle, the behaviour of the component, and the data acquisition. As a result, each time-series is now described by a vector \hat{v} in a high-dimensional feature space. The dimensionality of \hat{v} is reduced by the filtered features compared to \underline{v} , as it comprises only significant and uncorrelated features. Moreover, each feature exhibits a low variance and a platykurtic distribution over all test cycles for each specific, measured condition—hence a high density with very few outliers.

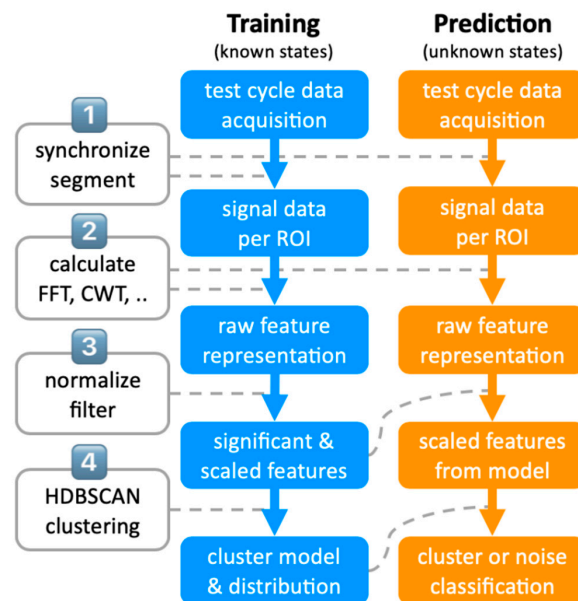


Figure 1. Solution approach for both model training (blue) and prediction of test cycle samples in unknown states (orange).

3.3. Model Creation

Based on the aggregated feature sets, a model can be trained to learn similarities or differences between feature set samples, which are high-dimensional ($n > 50$). Unsupervised algorithms are prone to perform worse with a growing dimensionality of the input vector, and therefore PCA for dense data, or singular value decomposition (SVD) for sparse data can reduce the dimensionality. In this case however, the significance, correlation, variance and kurtosis filtering already ensures that each element of the input vector explains a significant part of the overall variance. An additional dimensionality reduction negligibly increases the variance explained per vector element, and comes at the cost of detaching the input vector from their physical representation by the PCA/SVD aggregation.

Using unsupervised learning of the feature structures, the samples are clustered in agglomerations of similar feature sets. In this context, the notion of unsupervised learning refers to the fact that the actual conditions of the test cycle samples, commonly referred to as labels, are not fed into the model for training. The labels are merely used to determine the features to be retained for training the model in the initial model creation. Moreover, the labels of the test set are used to evaluate the performance of the approach. However, as the clustering approach only receives the feature values for each test cycle sample without labels, the actual training of the model is of an unsupervised nature.

Due to its ability to distinguish noise points from actual clusters, to accommodate varying cluster densities, as well as to infer the number of clusters, HDBSCAN is applied (4). For model training, noise points (i.e., samples with unknown conditions or failure states) are not relevant, as all samples definitely belong to a cluster (either healthy or one of the fault types). For the further analysis of unknown time series, however, a sample classified as noise reveals an unknown failure type, and therefore shall not be wrongly attributed to an existing cluster (false positive).

The results consist of a set of defined features and their normalization factors, as well as a model representing the distribution of the feature set samples. It enables time series of a test cycle performed on a component in an unknown condition to be processed, and a prediction on the component's current condition to be received. Future model updates can be performed similar to its initial training, where all n features are again extracted over all m ROIs, and subsequently normalized, filtered and clustered. With the measurement of a priori unknown failure types, the feature selection and filtering need to be repeated, as feature significance may have changed, i.e., previously insignificant features now serve as distinction between known failure type a , and new failure type b . Merely retraining the clustering model without recalculation of feature significance, therefore, neglects substantial information.

3.4. Model Deployment

For the prediction of a time series sample of an unknown machine condition, the following steps are conducted: (1) the time series is split into the defined ROIs, (2) the retained features of the model are selected and calculated, (3) the resulting features are normalized with the model scaler, and (4) the trained HDBSCAN model is applied to the unknown feature set. The return can yield two possible outcomes: either the sample of the test cycle is attributed to an existing cluster, which indicates that the component's condition corresponds to a prior measured and identified condition (healthy or a known fault type); or it is classified as a noise point, if the position of the sample vector \hat{v} lies outside of previously found regions with higher densities of samples in the feature space. The noise point classification occurs if the behaviour is different from any previously observed cluster of samples, meaning the component is either in an unknown faulty state, or neither in a healthy nor a known faulty condition. The latter may seem abstract, but could potentially happen if the boundaries of the healthy cluster are very dense, e.g., if only perfectly healthy machines were used for model training. Over time, intermediary states in a component lifetime (e.g., light, medium, strong wear) can be integrated and enable a more detailed clustering, ultimately allowing a RUL estimation when transition times between the different known conditions are measured or known.

3.5. Advantages Over the Current State of the Art

Compared to other approaches presented in the related work section, the proposed method detects not only the presence of failures. It also classifies the type of failure, given that it has previously been trained on and integrated in the model. Unknown conditions, which are neither a known fault or a healthy condition, are identified as such. This ability to cope with unknown failure types distinguishes it from conventional supervised classification approaches. It is applicable to various component and also machine types and natures: by the distinction of Gittler et al. [29], it can cope with test-cycle data of constant, controlled-constant and varying components. Moreover, the principle remains identical for translatory and rotary components. Given this versatility in the application of the method, it provides a high degree of automation in model construction and analysis. Moreover, updates

of the existing model require little engineering effort, as filtering and modelling require very few hyperparameters. The features retain the physical description of the signal samples, as the feature values without PCA or SVD transformation are used for clustering. In other related studies, large numbers of features or descriptive characteristics are usually reduced in dimensionality by PCA, e.g., as shown by Zhang et al. [23]. The training of the model can be performed on a small number of samples, enabling an application even with limited availability of test cycle samples. Therefore, it can serve both small and large installed bases and types of machines and components. The small number of hyperparameters and amount of data needed for the method reduce the engineering effort in its implementation, and lower the barrier of entry for machine and component OEMs. Furthermore, the model can be updated continuously with growing numbers of data samples and observed conditions. To the best of our knowledge, unsupervised approaches have not been demonstrated in machine tool component PHM applications.

4. Results

As a demonstration component, a translatory axis of a grinding machine is measured in different states—healthy state, and different faulty states. The tests are conducted on an Agathon DOM 4-axis grinding center typically used for the grinding of indexable inserts. The Agathon DOM has two translatory axes (X, Y) and two rotatory axes (B, C), of which the X axis is used exemplarily for the collection of data and the implementation of the approach described. The data collection is carried out in a controlled environment at constant 21 °C to ensure consistency and reproducibility of the results. The faulty states are artificially created, and reproduce the behaviour of defects that occur in operation. The faulty states include: (a) excessive friction (due to a lack of lubricant, contamination or debris in moving parts, collision), (b) a loose motor (tear and wear in the drive unit, involuntary release of screws due to vibrations), (c) a wrong commutation offset (due to a mechanical shift in the gearbox or along the cinematic chain), or (d) general signs of wear in the mechanics. The faulty states were recreated artificially for model training by the head service technician expert of the machine OEM. The selection of faults is based on the most frequent errors that have occurred on the entire installed base of machines in the field. The fault (a) was recreated by the insertion of a gasket between the moving parts of the axis and an adjacent wall, allowing an elevated friction and stick-slip effect to be created similar to that of a distorted or unlubricated axis. Fault condition (b) was recreated by losing screws in the coupling between the motor and the drive shaft. The commutation offset error in (c) was introduced by manipulating the encoder offset in the drive unit of the motor. The fault of general wear in the mechanics (d) was achieved by untightening the screws that connect the guiderails to the machine, allowing the axis to shift slightly during movements. Faults (b) and (d) correspond exactly to the type of error that potentially occurs on machines with a lack of maintenance, whereas fault (a) and (c) were recreations that approximate the behaviour of the axis under a real-world fault condition.

Overall, test cycles in 1 healthy and 4 faulty conditions are measured. For the different component conditions, 10 test cycle samples for healthy, and 6 samples each for faulty states are collected. For the model construction, 7 samples of the healthy state, and 5 samples of 3 faulty states are used. The remaining 3 samples of the healthy state and each sample of the faulty states are used as a test set to demonstrate and evaluate the functioning and the performance of the model. One faulty state is disregarded for the model, to test the model's capability to detect and classify a previously unknown faulty condition not used for prior model training, as neither healthy nor one of the known faulty states. The signals are sampled with 2×10^4 Hz, as some unhealthy vibrations are observable just below 10^4 Hz. The data are collected directly via the Agathon DOM's numerical control (NC), which is a Bosch Rexroth MTX with IndraControl L65. The NC has an integrated oscilloscope, allowing to record up to 4 signals on 4 channels in parallel, in addition to the monitoring of a trigger signal which can be configured separately. The oscilloscope can store up to 8192 values, wherefore a maximum test cycle duration of 4096 ms at 2×10^4 Hz can be recorded. As the test cycle for the entire outward (+) and return (−) movement exceeds this threshold, the test cycle is split into two parts, each covering one

direction of the movement. Figure 2 shows a section of the test cycle for different healthy and faulty state signals, in which the axis performs the (+) movement part of the test cycle. The plotted lines correspond to the sample data used for model training: green—healthy, red—faulty: excessive friction, blue—faulty: wrong commutation offset, yellow—faulty: motor loose. Of the entire test cycles, only the very relevant time segments are examined (orange shaded sections represent ROIs 1–4), to consider the different dynamic characteristics. It becomes clear that the different time segments (ROIs) exhibit significantly different aspects of the component behaviour, whereby the separate feature extraction per ROI is reasonable. Nonetheless, it is visible that some faults show only minimal differences, e.g., for the healthy condition (green) vs. the motor loose (yellow) fault. Figure 3 exhibits a small slice of ROI 2 in which the challenge becomes evident: whilst the excessive friction is simple to distinguish from the signal of the healthy axis, the motor loose fault behaviour is almost identical to healthy behaviour. The mere differences that can be spotted are in the vibrations and characteristics of the curve. This observation justifies the motivation to extract time series features to represent and classify the different test cycle measurements.

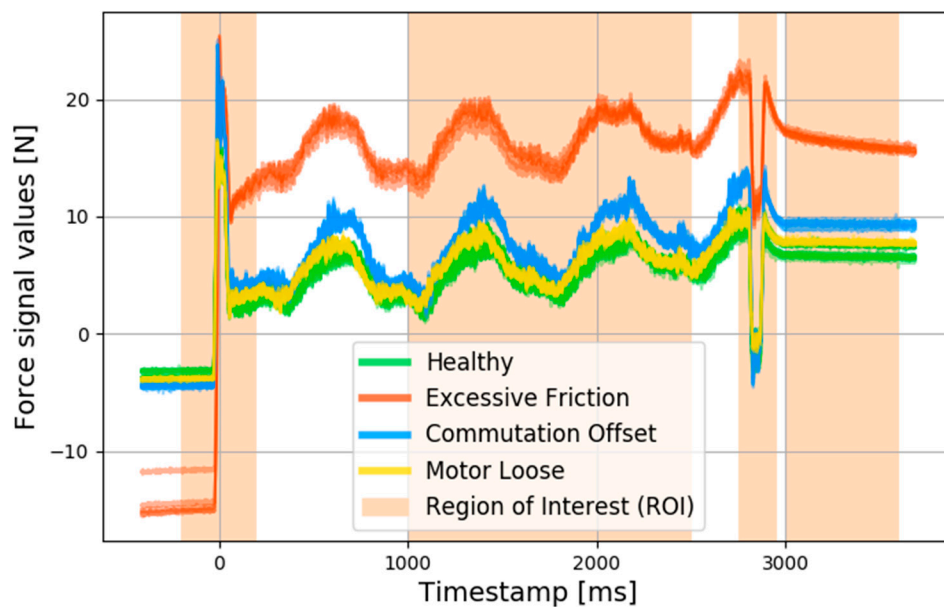


Figure 2. Force signal data of (+) direction test cycle of training samples.

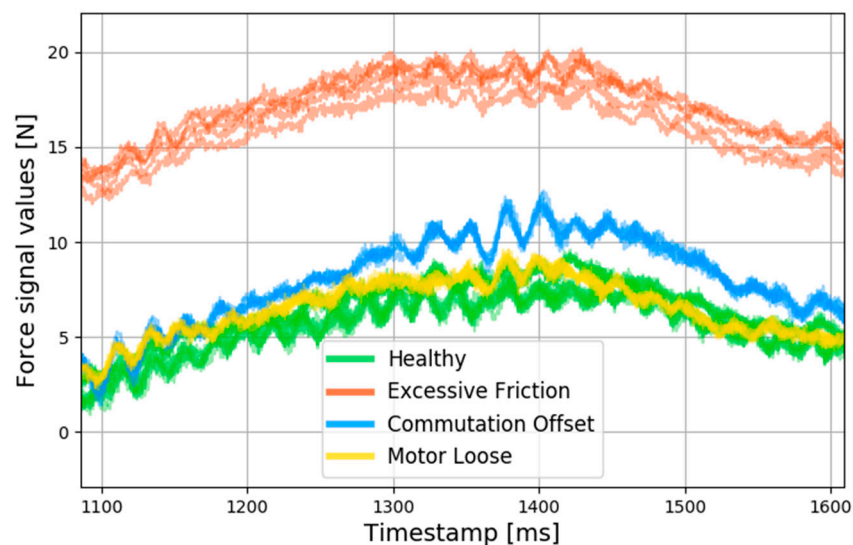


Figure 3. Zoom on region of interest (ROI) 2 displaying the characteristic behaviour of the different healthy and faulty conditions during constant velocity travel in (+) direction.

Prior to clustering, nearly 700 features for each of all $m = 8$ ROIs were extracted, resulting in a total of more than 5600 features. After filtering for relevance, statistical significance, variance, kurtosis and correlation, a total of 120 features for each sample were retained and used for clustering model construction. The discarded features are those, whose distribution does not allow samples of different conditions to be distinguished from one another at all. Some of the extracted and filtered features allow to distinguish clearly between all different kinds of faults, while others only permit us to distinguish between a pair of conditions, as show in Figure 4. Here, the exemplary distribution of 4 features extracted from ROI 2 in the slow test cycle (positive direction of axis travel) are shown, in which the histograms of the upper row show a distinct separation of feature values for all different conditions. The lower row shows two histograms of features that were retained, but that nonetheless have an overlap for some conditions. However, these features are nonetheless useful, as they still fulfil a viable function for the distinction of two or more conditions, and they potentially also permit to differentiate unknown conditions from those used to train the model. As the extraction and selection of features is the main determinant factor of the clustering result, this aspect is considered the most relevant in the described approach.

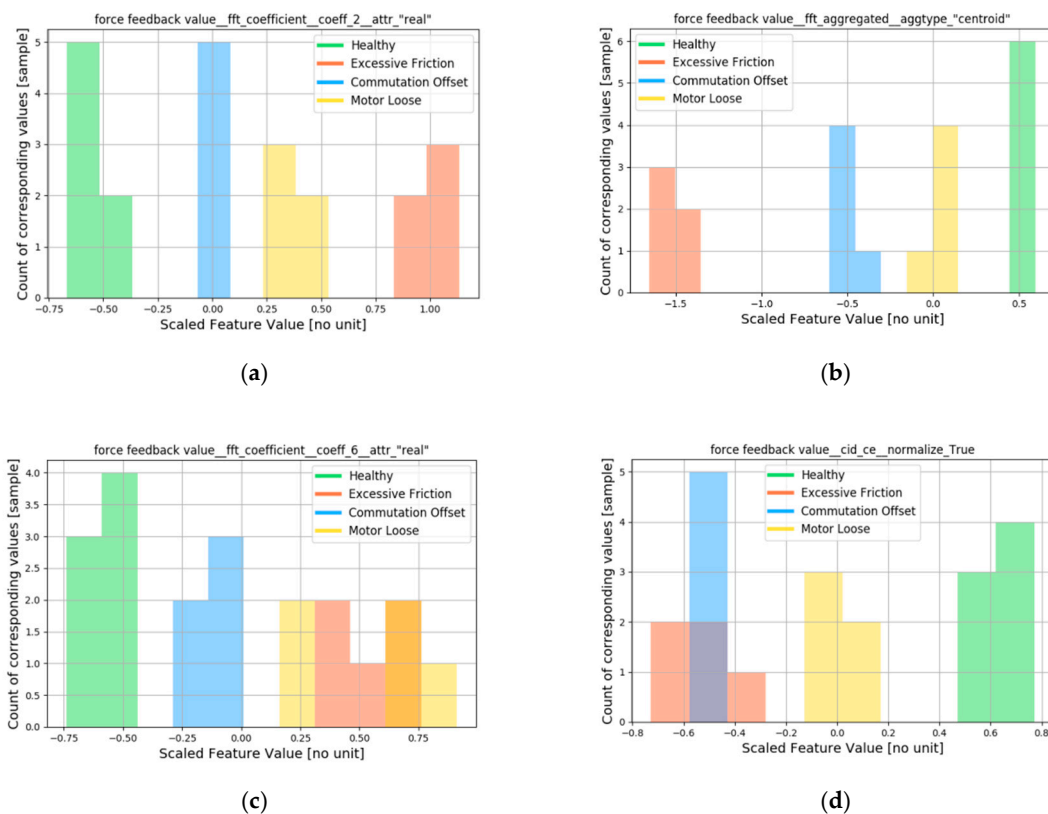


Figure 4. Examples of extracted and filtered features of ROI 2, in which the upper row [(a) and (b)] display high quality features allowing to distinguish all different conditions, whereas the lower row [(c) and (d)] contains features that overlap for some conditions: (a) Fourier-transform type; (b) Fourier-transform type; (c) Fourier-transform type; (d) complexity-invariant distance (CID) value.

To test the prediction precision, 5 samples of an unknown component condition representing mechanical wear are fed to the model for prediction. Figures 5 and 6 show the outcome of the different clustering approaches: The visualization is realized by transforming the multi-dimensional feature vectors of the samples into a 2D plane via T-distributed stochastic neighbour embedding (tSNE) for intuitive visualization [30]. The marker ‘O’ denotes a sample used for training, the marker ‘X’ designates a sample used as a prediction. The spatial location of the points represents the proximities of all points, wherefore neighbouring points have similar values of the feature vector \hat{v} . The colours

of the markers are assigned by the actual state of the training samples ('O'), or by the prediction of the test samples ('X'). As the prediction in clustering is an unsupervised process, the label for the predicted samples is assigned the label of the majority of points within the attributed cluster, e.g., if a sample is predicted to share a cluster with a large number of other healthy samples, it is assigned the condition healthy, and hence the colour green. To allow comparison of the engineering and tuning effort for all clustering algorithms, each was initialized with a minimum number of hyperparameters, i.e., without further modification. The optimal outcomes based on different initialization parameters were found iteratively. All results of a range of reasonable initialization parameters were evaluated and compared, of which the best results were chosen as a representative for the different algorithms. Figure 5 contains the k-means and the GMM clustering and prediction, in which both algorithms deliver identical results. k-Means was initialized with the parameter Number of Cluster n , with which the optimal result was found for $n = 4$. In a similar fashion, GMM was initialized with the Number of Components n , for which the optimum was also reached at $n = 4$. It is evident that the inability to handle noise points produces ambiguous prediction results, where all samples, regardless if outliers or noise points, are attributed to a cluster. In this case, a collection of points forming a proprietary cluster (red circle in Figure 5), corresponding to the unknown fault condition (mechanical wear), is wrongly attributed to the 'loose motor' cluster. Even though the distance between the two clusters is small, and the 'loose motor' condition shows similar physical properties and test cycle as results as the 'mechanical wear' fault, it is nonetheless a false positive prediction. Figure 6, depicting the model and prediction results of the HDBSCAN approach. In view of accurately classifying known healthy and faulty conditions, HDBSCAN performs identical to the k-means and GMM approaches. However, Figure 6 clearly shows that the samples of the prior unknown fault condition 'mechanical wear' are accurately identified as noise points, and therefore attributed to a new separate cluster. There is a pertinent notion in this context: the healthy condition, the motor loose and the mechanical wear faults show very similar behaviour considering the raw test cycle data. The faults are very minor and, therefore, do not differ greatly from the healthy condition. The fact that their distance and their delimitation from the other two similar conditions appears so clear demonstrates the effectiveness of the pre-processing, i.e., the feature representation and the subsequent filtering for significant features. All in all, the proposed approach allows us to concisely separate even minor differences and hence small faults from the optimal healthy condition of a component.

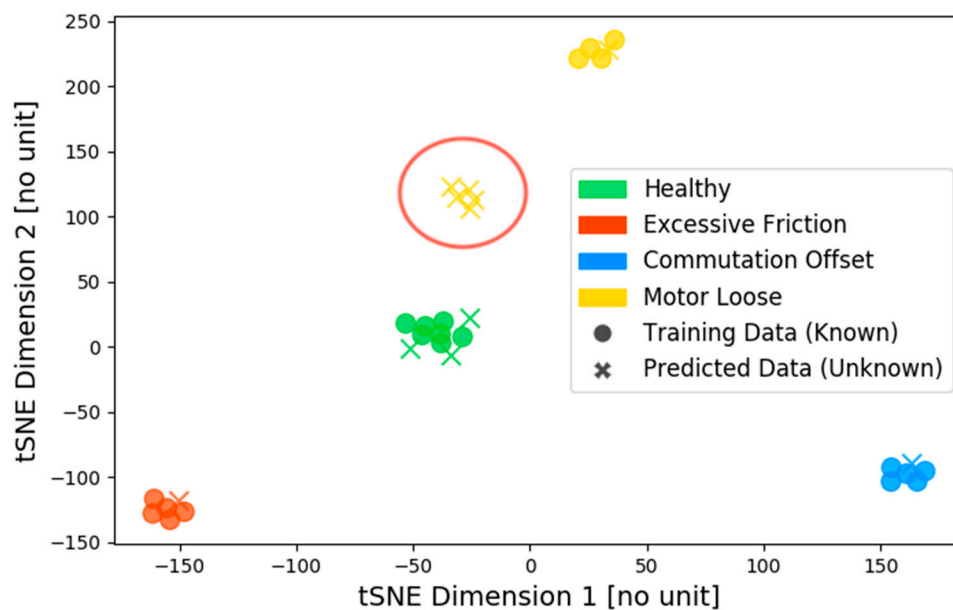


Figure 5. T-distributed stochastic neighbour embedding (tSNE) plot of training and prediction by k-means and GMM.

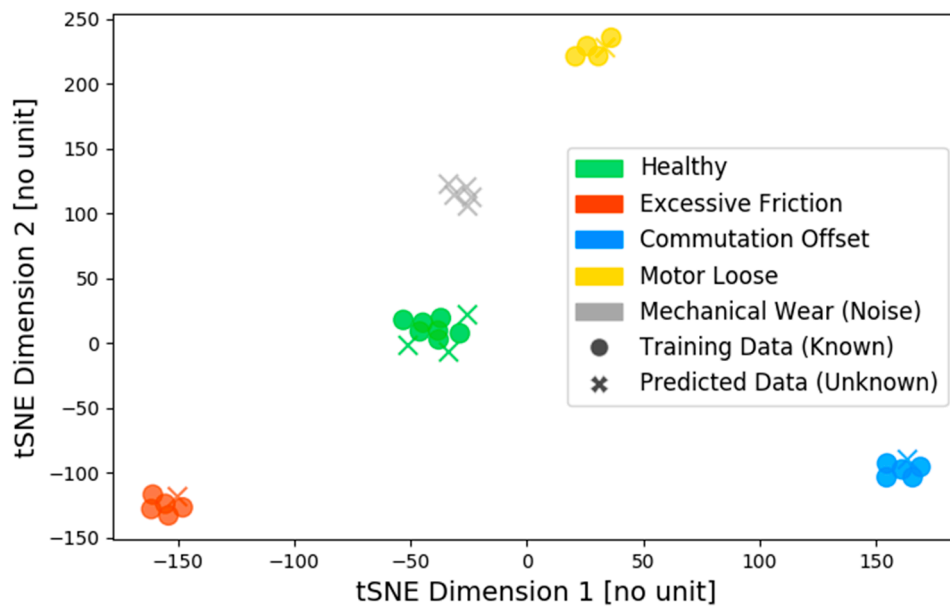


Figure 6. tSNE-plot of training and prediction with HDBSCAN.

After extensive testing of various parameter sets, only HDBSCAN was able to precisely cluster the training data, and accurately classify a cluster of unknown faults as noise. HDBSCAN was initialized with the only parameter Minimum Cluster Size k , for which the optimal results were achieved with $k = 3$. The results justify the selection of HDBSCAN as the optimal choice for unsupervised learning of machine component test cycle feature clusters. Its ability to accommodate varying cluster densities (i.e., more samples for the healthy vs. fewer samples for faulty states), the capability to classify a point or cluster of unknown condition samples, as well as the handling of non-convex cluster shapes in a high-dimensional space of feature vectors, make it a sound choice for the proposed approach. Table 2 shows the resulting best performances of all hyperparameter sets for each of the different algorithms. All initialization parameters were evaluated in sensible ranges to determine the optimal outcome, and hence the best possible performance for the underlying training and test data sets. For Figures 3 and 4, the visualization via t-SNE distorts the true noise and variance of some of the samples, as it warps the dimensions to accurately represent the distances of all points to one another. For this study, it is only meant as a visual reference to demonstrate the quality of the results. In reality, the clusters are of non-convex shape in the high-dimensional feature space.

Table 2. Result comparison of applied unsupervised approaches.

	Parameters	Optimum	True Positive	False Positive
k-Means	Number of clusters n	$n = 4$	85.3%	14.7%
GMM	Number of components n	$n = 4$	85.3%	14.7%
DBSCAN	Min samples per cluster k Epsilon ϵ	$k = 3, \epsilon = 0.7$	94.1%	5.9%
HDBSCAN	Minimum cluster size k	$k = 3$	100%	0%

5. Discussion

The proposed approach to assess the health of machine tool axes via time series feature extraction, filtering and unsupervised clustering has shown positive results. It has proven the applicability of unsupervised algorithms to component health identification, and demonstrated the advantages of unsupervised approaches over supervised models. It requires few data, and is straightforward to implement, maintain and extend for machine tool manufacturers. Unlike other PHM approaches, it allows for more than a binary distinction between healthy and failure states, including a priori

unobserved failure states. Therefore, not only can the presence of anomalies be identified, but different types and severities of faults on machine tool components. This multi-dimensional health assessment allows to reveal the impact a degradation can have on a production process or a final product. Besides an accurate assessment, the approach has proven to be applicable to real machine data rather than simulated data or anomalies. In the future, the performance with continuous model updates needs to be demonstrated. When new measurements of defects emerge, a model update with select measurements and subsequent model tuning is helpful. Moreover, the model tuning can be automated, as the multi-step approach is a complex optimization problem currently subject to heuristics and, therefore, non-deterministic. As most supervised approaches are able to quantify the degradation from the healthy state, this capability is yet to be delivered by the proposed approach. e.g., via distance or k-nearest neighbour calculation of actual test cycle samples. Additionally, the approach can be extended to components without control loop, by observing a stationary regime and applying the same solution scheme. Since the identification of a fault type yields an additional dimension, a future addition of a further dimension could be the evaluation of faults depending on the position of an axis. This allows for a more concise indication of where precisely a potential fault on an axis may develop or occur.

Author Contributions: Conceptualization, T.G.; methodology, T.G.; software, T.G.; validation, T.G., and A.R.; formal analysis, T.G.; investigation, T.G.; resources, S.S.; data curation, T.G.; writing—original draft preparation, T.G.; writing—review and editing, A.R. and K.W.; visualization, T.G.; supervision, S.S., A.R. and K.W.; project administration, A.R. and S.S.; funding acquisition, S.S. and K.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Innosuisse agency under Grant 2155002643. The authors would like to express their gratitude for the financial research support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kusiak, A. Smart manufacturing. *Int. J. Prod. Res.* **2018**, *56*, 508–517. [[CrossRef](#)]
2. Xu, L.D.; Xu, E.L.; Li, L. Industry 4.0: State of the art and future trends. *Int. J. Prod. Res.* **2018**, *56*, 2941–2962. [[CrossRef](#)]
3. Liao, Y.; Deschamps, F.; Loures, E.d.F.R.; Ramos, L.F.P. Past, present and future of Industry 4.0—A systematic literature review and research agenda proposal. *Int. J. Prod. Res.* **2017**, *55*, 3609–3629. [[CrossRef](#)]
4. Panetto, H.; Iung, B.; Ivanov, D.; Weichhart, G.; Wang, X. Challenges for the cyber-physical manufacturing enterprises of the future. *Annu. Rev. Control* **2019**, *47*, 200–213. [[CrossRef](#)]
5. Choudhary, A.K.; Harding, J.A.; Tiwari, M.K. Data mining in manufacturing: A review based on the kind of knowledge. *J. Intell. Manuf.* **2009**, *20*, 501–521. [[CrossRef](#)]
6. Andhare, A.B.; Tiger, C.K.; Ahmed, S. Failure Analysis of Machine Tools using GTMA and MADM method. *Int. J. Eng. Res. Technol.* **2012**, *1*, 1–11.
7. Tao, F.; Zhang, M.; Liu, Y.; Nee, A.Y.C. Digital twin driven prognostics and health management for complex equipment. *CIRP Ann.* **2018**, *67*, 169–172. [[CrossRef](#)]
8. Wegener, K.; Gittler, T.; Weiss, L. Dawn of new machining concepts: Compensated, intelligent, bioinspired. In Proceedings of the Procedia CIRP—8th CIRP Conference on High Performance Cutting (HPC 2018), Budapest, Hungary, 25–27 June 2018; Volume 77, pp. 1–17.
9. Sobie, C.; Freitas, C.; Nicolai, M. Simulation-driven machine learning: Bearing fault classification. *Mech. Syst. Signal Process.* **2018**, *99*, 403–419. [[CrossRef](#)]
10. Ruiz-Carcel, C.; Starr, A. Data-Based Detection and Diagnosis of Faults in Linear Actuators. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 2035–2047. [[CrossRef](#)]
11. Denkena, B.; Bergmann, B.; Stoppel, D. Reconstruction of Process Forces in a Five-Axis Milling Center with a LSTM Neural Network in Comparison to a Model-Based Approach. *J. Manuf. Mater. Process.* **2020**, *4*, 62.
12. Wuest, T.; Irgens, C.; Thoben, K.-D.; Wuest, T.; Thoben, K.-D. An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *J. Intell. Manuf.* **2014**, *25*, 1167–1180. [[CrossRef](#)]

13. Hiruta, T.; Uchida, T.; Yuda, S.; Umeda, Y. A design method of data analytics process for condition based maintenance. *CIRP Ann.* **2019**, *68*, 145–148. [[CrossRef](#)]
14. Xing, K.; Rimpault, X.; Mayer, J.R.R.; Chatelain, J.F.; Achiche, S. Five-axis machine tool fault monitoring using volumetric errors fractal analysis. *CIRP Ann.* **2019**, *68*, 555–558. [[CrossRef](#)]
15. Gittler, T.; Stoop, F.; Kryscio, D.; Weiss, L.; Wegener, K. Condition monitoring system for machine tool auxiliaries. In Proceedings of the Procedia CIRP—13th CIRP Conference on Intelligent Computation in Manufacturing Engineering (ICME 2019), Gulf of Naples, Italy, 17–19 July 2019.
16. Equeter, L.; Ducobu, F.; Rivière-Lorphèvre, E.; Serra, R.; Dehombreux, P. An analytic approach to the Cox proportional hazards model for estimating the lifespan of cutting tools. *J. Manuf. Mater. Process.* **2020**, *4*, 27. [[CrossRef](#)]
17. Ungermann, F.; Kuhnle, A.; Stricker, N.; Lanza, G. Data Analytics for Manufacturing Systems—A Data-Driven Approach for Process Optimization. *Procedia CIRP* **2019**, *81*, 369–374. [[CrossRef](#)]
18. Malhotra, P.; Tv, V.; Ramakrishnan, A.; Anand, G.; Vig, L.; Agarwal, P.; Shroff, G. Multi-Sensor Prognostics using an Unsupervised Health Index based on LSTM Encoder-Decoder. *arXiv* **2016**, arXiv:1608.06154.
19. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.-D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [[CrossRef](#)]
20. Gao, R.; Wang, L.; Teti, R.; Dornfeld, D.; Kumara, S.; Mori, M.; Helu, M. Cloud-enabled prognosis for manufacturing. *CIRP Ann.* **2015**, *64*, 749–772. [[CrossRef](#)]
21. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [[CrossRef](#)]
22. Duan, C.; Makis, V.; Deng, C. Optimal Bayesian early fault detection for CNC equipment using hidden semi-Markov process. *Mech. Syst. Signal Process.* **2019**, *122*, 290–306. [[CrossRef](#)]
23. Zhang, L.; Elghazoly, S.; Tweedie, B. AnomDB: Unsupervised Anomaly Detection Method for CNC Machine Control Data. *PHM 2019* **2019**, *11*, 1–12.
24. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.
25. McLachlan, G.J.; Lee, S.X.; Rathnayake, S.I. Finite Mixture Models. *Annu. Rev. Statist. Its Appl.* **2019**, *6*, 355–378. [[CrossRef](#)]
26. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; Volume 96, pp. 226–231.
27. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [[CrossRef](#)]
28. MacQueen, J.B. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 21 June–18 July 1965 and 27 December 1965–7 January 1966; pp. 281–297.
29. Gittler, T.; Gontarz, A.; Weiss, L.; Wegener, K. A fundamental approach for data acquisition on machine tools as enabler for analytical Industrie 4.0 applications. In Proceedings of the Procedia CIRP—12th CIRP Conference on Intelligent Computation in Manufacturing Engineering (ICME 2018), Gulf of Naples, Italy, 18–20 July 2018; Volume 79, pp. 586–591. [[CrossRef](#)]
30. Van Der Maaten, L. Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **2015**, *15*, 3221–3245.

