applied sciences

MDPI

Article

Vibration-Based Detection of Bearing Damages in a Planetary Gearbox Using Convolutional Neural Networks

Julia Scholtyssek *, Luka Josephine Bislich, Felix Cordes and Karl-Ludwig Krieger

Institute of Electrodynamics and Microelectronics, University of Bremen, Otto-Hahn-Allee 1, 28359 Bremen, Germany

* Correspondence: julia.scholtyssek@uni-bremen.de

Abstract: Tapered roller bearings are used partly in very rough and highly stressful environmental conditions. Therefore, the need for condition monitoring is increasing. This study is intended to provide an approach for monitoring bearings in a two-stage planetary gearbox based on vibration analysis. In total, the data of six damage phenomena and one healthy bearing are collected. A convolutional neural network (CNN) is trained and evaluated by using the balanced accuracy. Mainly, it is investigated how many damage severities can be detected. In addition, the robustness of the model regarding unknown speeds and damage phenomena should be proven. The results show a very good differentiation up to all of the presented damage phenomena. The classifier reaches an averaged balanced accuracy of 0.96. Also, samples collected at unknown speeds can be classified well for speed values within the known range. For unknown damage phenomena, the classifier shows limits so that a reliable classification is only applicable with a binary classifier, which differentiates between healthy and damaged. The investigations therefore show that a reliable detection of bearing damage is possible in a two-stage planetary gear. Furthermore, the transferability of the model is successfully tested and implemented for the binary classifier.

Keywords: planetary gear; bearing damage; vibration analysis; CNN; classification

check for updates

Citation: Scholtyssek, J.; Bislich, L.; Cordes, F.; Krieger, K.-L. Vibration-Based Detection of Bearing Damages in a Planetary Gearbox Using Convolutional Neural Networks. *Appl. Sci.* **2023**, *13*, 8239. https://doi.org/10.3390/app13148239

Academic Editors: Aleksandar Miltenovic, Franco Concli and Stefan Schumann

Received: 26 June 2023
Revised: 12 July 2023
Accepted: 15 July 2023
Published: 16 July 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the present application, the tapered roller bearings are installed in a two-stage planetary gearbox in port vehicles. Since the bearings are exposed to a highly stressful environment, the bearings can evolve the wearing phenomena. Particularly in the port application, unpredictable machine failures must be avoided, as this results in high costs. Therefore, the condition monitoring of the bearings is necessary to detect damage at an early stage. To implement this, acoustic signals are evaluated in this contribution.

There is already a large number of publications regarding the condition monitoring of bearings. The most important results are introduced briefly. In general, two diverging approaches can be distinguished, namely signal processing methods and intelligent methods [1,2]. In general, the signals in the time domain, frequency domain or time–frequency domain can be used for signal-processing methods. Typical methods to analyze the signals are statistical indicators, the detection of characteristic fault frequencies for each component or occurring sidebands and spectrum observations [3,4]. In addition to common methods like Fourier transformation, multiple decomposition methods like empirical mode decomposition or variational mode decomposition, have been developed [5]. The advantage of these methods compared to Fourier transformation is that they are applicable to non-stationary signals, which are frequently occurring in gearboxes. Therefore, the vibration signals are decomposed into mono-component modes. A new method for detecting bearing damage in this field is feature mode decomposition (FMD) [6]. Based on the use of a finite-impulse response filter, the signals are preprocessed. Afterwards, a mode selection based on a combined argument of correlated kurtosis and the correlation coefficient is

performed. FMD can successfully detect individual damages and, in contrast to previous decomposition methods, is also capable of extracting the modes in the case of combined damages. The database of all presented signal processing methods can be used as an input for a variant of intelligent methods, like support vector machines or neural networks [7].

Based on the literature, the presented methods are not applicable to every use case. Most of the publications concern laboratory investigations on bearings. Investigations of bearings integrated in a gearbox are rare. Planetary gearboxes are used in rough and variable conditions. Therefore, the evaluation of fault frequencies based on the fast Fourier transformation is not possible since it is designed for stationary signals [8]. Furthermore, planetary gearboxes, and, therefore, bearings in planetary gearboxes, show additional challenges like the superposition of vibration events, nonlinear transmission paths, low-speed components and noise. Also, most investigations are based on localized faults, like local pittings. Investigations on distributed wearing phenomena are rare.

This paper presents an investigation of localized and distributed defects. The investigations on the tapered roller bearings are being developed in the compound system of the two-stage planetary gearbox. In addition, this paper investigates the influence of unknown damage and is not limited to the detection of damage classes which the network has already used in the training process. For this, a convolutional neural network is developed, which takes spectral data as input. The first investigation addresses the class granularity. It is analyzed how many classes the classifier can reliably differentiate. Therefore, a binary classifier differentiating between two classes, healthy and damaged, is first developed. Afterwards, a classifier differentiating between three classes, healthy, localized and distributed damage, is used. Finally, a classifier differentiating between seven classes, which show different severities of the presented subclasses, is designed and evaluated. The second investigation addresses the impact of different speed rotations. The last investigation deals with unknown damage phenomena. For this, the classifier only differentiating in three classes is used. Single damage phenomena are held back during training and are only used in the test set to reveal if a correct classification is possible.

The article is structured as follows. Section 2 presents the experimental setup. First, the two-stage planetary gearbox design is introduced. Afterwards, the investigated damages and lastly, the data acquisition and test rig are presented. Section 3 deals with the dataset and classifier preparation. Afterwards, the different investigations are presented in Section 4. In the first investigation the implementable class granularity is proven. Secondly, the influence of different speed rotations is observed. Finally, the robustness of the developed classifier is tested by classifying unknown damage events. In Section 5, the results are discussed and a conclusion is drawn including the optimization potential.

2. Materials and Methods

This section describes the fundamental setup of this contribution, which is used to realize the condition monitoring of bearing damages by using vibration analysis. First, the structure of the examined two-stage planetary gearbox is presented. Afterwards, the occurrence of bearing damages and the evaluated damages are described. Lastly, the experimental setup is outlined.

2.1. Two-Stage Planetary Gearbox

The most important specifications of the used two-stage planetary gear should be presented. The gearbox has a maximum input speed rotation of 4170 rpm and a maximum input torque of 450 Nm. The specifications of each stage are presented in Table 1.

Table 1. Specifications of the investigated two-stage planetary gearbox.

Property	First Stage	Second Stage
number of teeth sun gear	15	12
number of teeth planet gear	31	23
number of teeth ring gear	−78	−60
number of planets	3	4

The total gear ratio is $i = 37.2$. A planetary gear consists of three main components, namely the ring gear, sun gear and planet gears. In total, there are three types of planetary gearbox designs [9]. They differ in the component that is fixed and therefore not rotating. The presented design is based on a fixed ring gear. The sun gear is in the center, rotating around its own center. The planets are placed between the sun and the fixed ring gear and mesh with both components [10,11]. They rotate orbitally to the sun and, in addition, they rotate around their own center. The used planetary gearbox is a two-stage design. Hence, the planet gears of the first stage are connected to the sun of the second stage by using a planet carrier shaft. The construction design is shown in Figure 1.

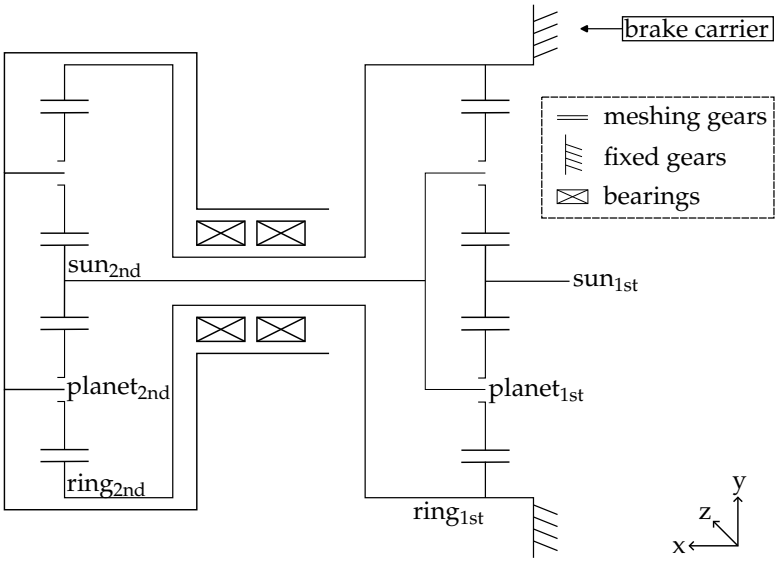


Figure 1. Construction draft of a two-stage planetary gear box.

In addition to the gearing components, the two-stage planetary gearbox has two sets of taper roller bearings. Both are mounted between the ring gear of the second stage and the ring gear carrier as depicted in Figure 1. The most important specifications of the applied taper roller bearings are listed in Table 2.

Table 2. Specifications of the investigated taper roller bearings.

Dimensions		Performance Data	
Inner diameter	150 mm	Basic dynamic load rating, radial C_r	455,000 N
Diameter between two opposite rolling elements	187.8 mm	Basic static load rating, radial C_{0r}	680,000 N
Outer diameter	225 mm	Limiting speed n_G	3550 1/min
Width	48 mm	Operating temperature range	−30–200 ° C

A taper roller bearing consists of an outer ring, taper rollers and an inner ring. In the presented application, the outer ring is rotating and the inner ring is stationary. Since both bearing sets are rotating with dependence on the wheel hub, they are investigated in conjunction. For each bearing component, an individual frequency can be calculated. The first factor is a characteristic constant for each component, which shows design dependencies. The second factor is the rotation frequency of the wheel hub f_n, which depends on the input rotation speed:

$$\text{Ball spin frequency } BSF = BSFF \cdot f_n = 5.06 \cdot f_n, \tag{1}$$

$$\text{Ball pass-frequency outer race } BPFO = BPFFO \cdot f_n = 13.58 \cdot f_n, \tag{2}$$

$$\text{Ball pass-frequency inner race } BPFI = BPFFI \cdot f_n = 16.42 \cdot f_n, \tag{3}$$

$$\text{Fundamental train frequency } FTF = FTFF \cdot f_n = 1.09 \cdot f_n. \tag{4}$$

Planetary gearboxes are gaining increasing attention. They are used in transportation systems, aerospace, automotive applications and wind turbines [9,12]. The main advantages are a compact structure, a stable transmission, a large transmission ratio and high efficiency. Due to rough conditions, the bearings and gearings often show faults like wear, fatigue cracks or pittings. The resulting faulty gearbox can cause operation hazards, with a shutdown as the worst-case scenario and the associated high costs. But some difficulties arise for vibration monitoring. For a two-stage planetary gearbox, the number of meshing components as well as that of the transmission paths increase. As a result, the effects of dispersion, absorption and superposition become more intense. This contribution focuses on the detection of bearing damages in a two-stage planetary gearbox.

2.2. Bearing Damage

Generally, occurring bearing damages can be categorized into localized and distributed damages [13,14]. Localized faults can result from material defects. Those faults can be spread over time through particles in the oil circulation system or contact between components like meshing gears. Therefore, it is necessary to detect those faults in an early stage to avoid a progressive fault and a premature failure of the whole system. Distributed faults mostly occur beyond the guaranteed lifetime. They are caused by material fatigue due to the applied load and the tough operating conditions.

This contribution considers three different kinds of damages for each subclass. A comprehensive overview is shown in Table 3. The three bearing damages which belong to the distributed subclass show severe pittings distributed over the outer race and tapper rollers. They only show slight differences in severity. All three damages occurred during real operation.

The three bearing damages, which belong to the localized subclass, are sequentially more severe. They are added with one local fault in each iteration. Therefore, the damage of the class *local 0* is included in classes *local 1* and *local 2*. All three damage phenomena are synthetically produced and are most likely to be compared to scoring. The first damage is severe local damage on the outer race. The second damage is light local damage on the outer race with 9 cm distance to the severe local damage. The last damage is severe damage on one tapper roller. Both the distributed and the localized classes do not have information about the inner race condition. This is due to the fact that an evaluation of the inner race would need a total destruction of the whole bearing set.

Table 3. Overview of the considered damaged bearings.

Subclass	Class	Damaged Components
healthy	healthy	
distributed	dist 0	severe pittings on outer race and tapper rollers
	dist 1	severe pittings on outer race and tapper rollers
	dist 2	severe pittings on outer race and tapper rollers
localized	local 0	severe scoring on outer race
	local 1	severe and light scoring on outer race
	local 2	severe and light scoring on outer race, severe scoring on tapper roller

2.3. Experimental Setup

For the acquisition of the vibration signals, a piezoelectric vibration transducer is used, which has a linear frequency range from 0.13 Hz to 22 kHz and a sensitivity of 100 mV/g. The sensor is an accelerometer. A signal acquisition at the bearings is not possible. Therefore, the sensor signal acquisition is carried out at the brake carrier of the wheel drive. The brake carrier is located at the first stage as shown in Figure 1. The sensor is mono-axial and orientated in x-direction. The IEPE signal of the sensor is logged through a data-acquisition system with a sampling rate of 100 kHz and a resolution of 24 bit.

The data acquisition is carried out with a load-free test rig. In total, seven measurement campaigns are executed. In the first campaign, an undamaged test gearbox is used. Afterwards, the damaged components listed in Table 3 are separately mounted to the test gearbox. For each campaign, multiple constant rotational speeds n are applied, namely 700, 1200, 1700, 2100, 2700, 3400, 4200 rpm. Each rotational speed is measured five times with a duration of 20 s, respectively.

3. Data Processing

In the present investigation, it is very difficult to derive precise predictions of the bearing condition based on signal-processing methods due to the noisy environment. Therefore, intelligent methods are used. A convolutional neural network (CNN) is tested for this purpose. Since the CNN processes image data, spectrograms are generated from the raw signals in the first step. This is explained in more detail in Section 3.1. The general functionality and the structure of the used CNN is presented in Section 3.2. Furthermore, since many classifiers show the overfitting and drawbacks of imbalanced datasets, suitable evaluation methods are identified in Section 3.3 to enable a reliable evaluation of the results.

3.1. Data Preparation

Each dataset is divided into data segments with the length of half a second, 50,000 samples respectively, which are transformed into the time–frequency domain and depicted as spectrograms. The spectrograms are scaled logarithmic and the color scale is normed with the minimal and maximal values over the whole dataset. The spectrograms are saved as pictures without an axis because they are normed. The pictures are then used as three-dimensional input for a CNN. In total, the dataset consists of 9800 spectrograms, which are evenly distributed over all seven classes. An example of the generated spectrograms for a rotational speed of 2700 rpm is shown in Figure 2. For the same operating conditions, the resulting spectrograms of three damage cases, respectively *healthy*, *dist 2* and *local 0*, are shown. Slight differences can be detected. In relation to the healthy condition, the signal amplitude in the lower frequency range increases in the case of distributed damage, while the middle frequency range obtains higher amplitudes only within a clearly definable frequency band. For local damage, on the other hand, the lower frequency band has higher amplitude than for the healthy state and is larger. In addition, this damage exhibits higher amplitude–frequency bands in the middle-frequency range. In the high-frequency range, all three damage phenomena behave similarly.

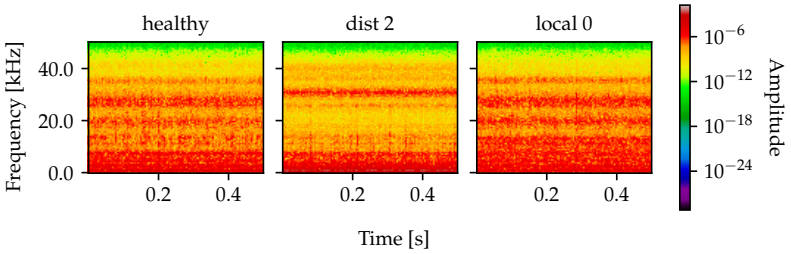


Figure 2. Spectrograms of three damage classes for 2700 rpm, which are used as input for the convolutional neural network (CNN).

3.2. Classifier

CNNs are used in many applications, which can be summarized mainly into three categories [15]. The first category is classification. Examples for classification are mainly medical image analysis and real-time target tracking. CNNs are also used for object detection, which is the second category. Object detection includes object localization and object classification. It is often used in medical context, for example, in radiology with X-ray images to detect fractured bones or in the case of tumor detection [16]. Lastly, the third main application is segmentation. In addition, CNNs can be used as a preprocessing step for feature extraction [17].

CNNs provide many advantages. They offer to extract features from the raw data with no need for preprocessing steps [17]. Furthermore, they are precise and result in high accuracy [18]. They have a high ability to work with variable or complex data, outperform other intelligent methods, are computationally efficient and only need few parameters [19]. Therefore, the application to noisy acoustic signals in the port environment is promising for the realization of a reliable classifier.

Generally, a CNN is a multi-layer neural network and used for image data [17]. CNNs use discrete convolution to extract information from the image data. The general structure of CNNs is presented shortly. CNNs basically consist of an input layer, hidden layers and an output layer. The hidden layers are convolutional, pooling or fully connected layers. First, the images are fed to the network using the input layer. Afterwards, a combination of multiple convolutional and pooling layers is used to extract features. One convolutional layer is built by a set of multiple kernels [15]. Each kernel has weights, which are window-wise convolved with the input array. This creates features and also reduces the data size [20]. The key parameters are the size and the number of kernels [16], whereas pooling layers are only used to reduce the feature size, while retaining the relevant information. They reduce the dimensionalities and downscale the matrix. Pooling layers, in addition, improve the net transferability to unknown data [18]. After multiple convolutional and pooling layers, the two-dimensional feature space is flattened to one dimensionality. This is performed through a fully connected layer and is required for the classification. At least one layer is necessary but multiple are possible. The last layer has the same size as the defined classes. In each fully connected layer, each input is connected and weighted to the next output. Lastly, a function of activation is needed, which normalizes the output to class probabilities.

CNNs have successfully been used in bearing condition monitoring [2]. Guo et al., for example, used an adaptive deep CNN to differ between four bearing health states [19]. Qian et al. used adaptive overlapping convolutional neural networks to increase the performance by using the one-dimensional raw vibration signal [21]. This method reached high accuracy for a public bearing dataset with ten health conditions, while using only 5 % of the data for training. In addition, the results were verified with a second dataset of a different bearing investigation.

For the presented application, a CNN is developed with tensorflow and tested for differentiating into the presented bearing damage cases. The architecture is shown in Figure 3. In addition to the layer configuration, the two-dimensional kernel size of each layer is provided. Also, the number of features is shown below each convolutional layer.

The first convolutional layer restructures the three-dimensional input, which is comprised of the RGB values of the spectrograms. The layer is combined with a relu activation layer and a max pooling layer. Afterwards, three convolutional layers, each combined with a relu activation layer, a normalization layer and a pooling layer using max pooling, are implemented. Finally, the data are flattened by global pooling, which is connected to a dense layer. Lastly, a second dense layer combined with a softmax activation layer is connected, which has the same size as the considered number of classes. Since the number of classes varies in different investigations, the last layer is variable.

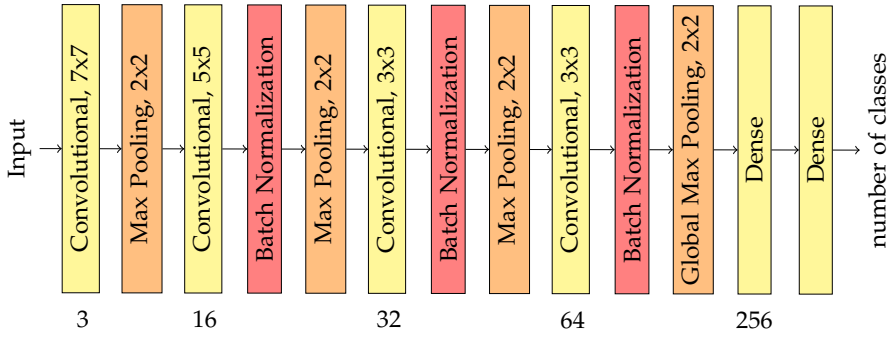


Figure 3. Applied CNN architecture.

3.3. Evaluation

To train the CNN, k-fold cross validation is used. Typically, in machine and deep learning, $k = 10$ is chosen as in the present investigation. First, the dataset is randomly shuffled. Afterwards, the dataset is divided into ten equal-sized parts, where each part is used once for testing [22,23]. The remaining data are used to develop the CNN-model, where eight parts are used for training and one part is used for validation. As consequence, the model is trained ten times. To evaluate the model accuracy, the mean result as well as the best and the worst iterations are considered. Cross validation ensures that all data points can be used for training and testing without resulting in overfitting [24]. Without a good evaluation, many models tend to classify the trained data perfectly but fail for new data. This is called the generalization problem, which is overcome by using cross validation.

The present investigation is a multi-class problem. The most common classifier metric is the accuracy (ACC). It represents the ratio of correctly classified samples to the whole dataset [25]:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}. \tag{5}$$

Therefore, the correctly classified samples of the positive class (TP) and the correctly classified samples of the negative class (TN) are set in relation to all sample points. This includes the correct classification as well as the false classification of the positive (FP) and negative (FN) class.

This metric is useful if there is either a balanced class distribution or only the probability of a correct classification is important, independent of the classes. In the present case, a balanced class distribution is only given when differentiating in seven classes. But this might not be the case for each cross validation. In addition, further information about the ability to detect every single class is needed. Therefore, another metric, handling imbalanced datasets, should be used.

A common solution for this is the balanced accuracy (BAC) [26,27]. The BAC is the average over each class dependent recall:

$$\mathrm{BAC} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{recall}_i. \tag{6}$$

The recall is the ratio of correctly classified samples of a defined class. Defining the considered class as positive and all others as negative, the class specific recall is the TP classified samples of a class in relation to all suggested class samples, which includes the TP but also the FN:

$$\text{recall}_i = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{7}$$

4. Results

The main aim of this contribution is to realize a reliable classifier to detect bearing damages. In this context, some specific investigations are presented for this application. First of all, the implementable class granularity should be tested. This is presented in Section 4.1. The investigations on the test rig are examined on defined speeds. In real operations, constant and defined speeds at any time cannot be assumed. Therefore, the effects of data collected at unknown speeds on the classifier results should be investigated. This is shown in Section 4.2. Lastly, based on the presented class granularities, it is investigated if unknown damages can be sorted into subclasses by the trained classifier. This is shown in Section 4.3. These last two tests are intended to provide an indication of the robustness of the designed classifier from Section 4.1.

4.1. Class Granularity

The first study examines whether the detection of bearing damage in the present application is implementable with the use of a CNN. At first, a binary classifier is developed, which distinguishes between healthy and damaged. The second class is composed of the distributed and localized classes. Then, the classification problem is extended to three classes. Thereby, the subclasses healthy, distributed and localized are differentiated. As shown in Table 3, the classes *dist 0*, *dist 1* and *dist 2* are combined for the distributed subclass. The subclass localized consists of *local 0*, *local 1* and *local 2*. In the last step, the class granularity is increased to seven classes. Therefore, all classes shown in Table 3 should be detected. The investigation of the class granularity should provide information as to which granularity should be preferred in the case of a suitable trade-off between high-quality and high-information content. The results are shown in Table 4. Each model is trained ten times using tenfold cross validation, with the averaged BAC over all iterations and the results of the best and worst iterations listed, respectively.

Table 4. Classifier results for different class granularities.

Class Granularity	BAC_{mean}	BAC_{max}	BAC_{min}
2	0.99	1	0.98
3	0.98	0.99	0.97
7	0.96	1	0.70

The result shows that the binary classifier achieves the overall best result with an average BAC of 0.99. The worse iteration is achieving a BAC of about 0.98. But the classifiers with higher granularity show very good results, too. With increasing class granularity, the BAC decreases. But the classifier with seven classes still reaches an average BAC of 0.96. The worst iteration produces a BAC of 0.70, which is far worse than the other granularities. However, this iteration could be seen as an outlier since the second-worst iteration reaches a BAC of about 0.98. In Figure 4, the result of the worst iteration of the classifier with seven classes is shown. A confusion matrix (CM) is used for visualization. On the horizontal axis, the predicted labels are shown, whereas on the vertical axis, the true labels are depicted. A good classifier should predict the true labels, and therefore most of the samples should be placed on the main diagonal. The CM shows that the distributed classes are classified nearly perfectly, while the healthy and all of the local classes show multiple misclassifications. It is noteworthy that all four classes are often predicted as class *dist 0*.

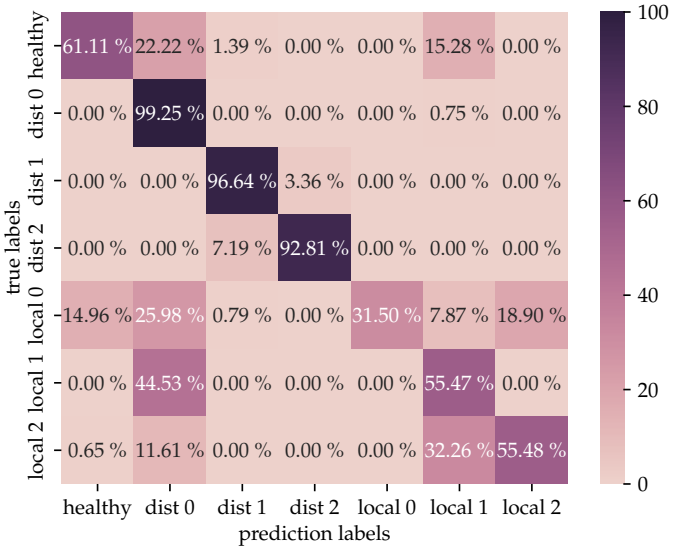


Figure 4. Confusion matrix for the worst iteration of the classifier differentiating seven classes.

In summary, the classification of all presented granularities is possible with sufficiently good accuracy. Depending on the requirements and the use case, it should be decided individually which misclassifications are acceptable, and the largest granularity that meets these criteria should be selected accordingly. For the presented use case, the result of the classifier with seven classes is sufficient.

4.2. Speed Dependency

The second study examines the influence of speed dependencies. As presented, the data are collected at a constant speed with seven different rotational steps. In the real application, a constant speed at defined values is not possible. Therefore, the influence of the speed on the classifier accuracy should be investigated.

Two investigations are performed. In the first investigation, the classifier is trained with rotational speeds of 700, 1700, 2700 and 4200 rpm and tested with the other three speeds. In the second investigation, the classifier is trained with rotational speeds of 1200, 2100 and 3400 rpm. The following equations provide an overview of the investigations, where D is the whole dataset:

$$\text{Investigation 1: } D_{\text{train}} = D \setminus D_{\text{test}}, \quad D_{\text{test}} = \{1200, 2100, 3400\}\ \text{rpm}, \tag{8}$$

$$\text{Investigation 2: } D_{\text{train}} = D \setminus D_{\text{test}}, \quad D_{\text{test}} = \{700, 1700, 2700, 4200\}\ \text{rpm}. \tag{9}$$

Again, all investigations are performed using tenfold cross validation. The results are shown in Table 5. The highlighted data show the results of the tested data, which means the trained classifier does not know those rotational speeds. First of all, the BAC is decreasing for samples, if the present rotational speed is not used in the training process. In the worst case, data samples collected at 700 rpm only reach an averaged BAC of 0.20, and the worst iteration only reaches 0.06. Therefore, the classifier becomes worse than random prediction. Secondly, it can be derived from the results that, compared among rotational speeds not used in the training process, the intermediate rotational speeds can achieve the best classification accuracy. Therefore, the data collected at 2100 rpm still reach an averaged BAC of 0.83 in the first investigation, although this speed is not trained. The second and third best results are reached at 3400 and 1700 rpm. Also, it can be noted that the results from the first investigation achieve higher BAC. The main difference to the second investigation is that only intermediate speeds are tested and wrapped by trained rotational speeds. In addition, with the worst results of the lowest and highest rotational speeds and the best results for the middle rotational speeds, it can be derived that for a good classification of unknown rotational speeds, a suitable coverage of rotational speeds

during training is essential. Therefore, the rotational speeds in the training set should cover the whole range and be evenly distributed since the investigation shows that data of rotational speeds, which are wrapped by trained rotational speeds, result in good accuracy.

Table 5. Classifier results for withheld speeds.

n	Investigation 1			Investigation 2		
	BAC_{mean}	BAC_{max}	BAC_{min}	BAC_{mean}	BAC_{max}	BAC_{min}
700	0.99	1	0.99	**0.20**	**0.33**	**0.06**
1200	**0.63**	**0.76**	**0.44**	0.63	0.76	0.44
1700	1	1	0.99	**0.71**	**0.92**	**0.61**
2100	**0.83**	**0.90**	**0.73**	0.99	1	0.91
2700	1	1	0.99	**0.67**	**0.78**	**0.50**
3400	**0.75**	**0.92**	**0.53**	0.99	1	0.95
4200	1	1	1	**0.38**	**0.52**	**0.28**

It can, therefore, be assumed that the ranges in between the known rotational speeds from the present measurements are also classifiable.

4.3. Unknown damages

In the third study, the robustness of the classifier is evaluated with regards to unknown damage conditions. To verify this, we first use the classifier with three classes. Unknown damages should therefore be classified into the categories healthy, local and distributed damage. A distinction of the severity levels of the local and distributed damages cannot be tested due to the limited data. The three distributed damages are all defined as severe. In addition, there is a difficulty of quantifying the damage. A distinctive separation between light and heavy damage is non-trivial and subjective for each applicator. For these reasons, a classification of the severity is not provided.

The focus of the investigation is on the classification into the three presented subclasses. For the investigation, each damage class except *healthy* is withheld once in the training. The remaining damage phenomena are combined according to the subclasses and constitute the training data for the classifier. Subsequently, it is tested whether the withheld class can be classified into the correct subclass. The result is shown in Table 6. Here, the result for each withheld class is listed separately. Again, this test is performed with tenfold cross validation. The average accuracy and the accuracies of the best and worst iterations are listed.

Table 6. Classifier results for withheld classes with three subclasses.

Unknown Class	ACC_{mean}	ACC_{max}	ACC_{min}
dist 0	0.03	0.29	0
dist 1	0.99	1	0.98
dist 2	1	1	0.99
local 0	0.79	0.87	0.67
local 1	0.43	0.77	0.23
local 2	0.89	0.98	0.81

The result shows that classes *dist 1*, *dist 2* and *local 2* are classified almost perfectly. Class *local 0* shows moderate accuracy, whereas classes *dist 0* and *local 1* cannot be classified with sufficient accuracy. Class *dist 0* in particular performs very poorly, which is rather unexpected especially in comparison to the very good results of the other distributed damage classes.

A closer look at the misclassifications of withheld class *dist 0* reveals that most of the predictions belong to subclass *local*. In contrast to the classification of a healthy gearbox if damage is present, this does not represent a safety risk. In Figure 5, the CM of the best iteration of *local 1* is shown. Although most of the misclassifications are at the distributed

damage and are therefore not critical, there are significantly more misclassifications at *healthy* than in the previous case. Therefore, although it yields better accuracy than the tests with *dist 0*, this could turn into a safety risk, and therefore the usage should be considered depending on the application.

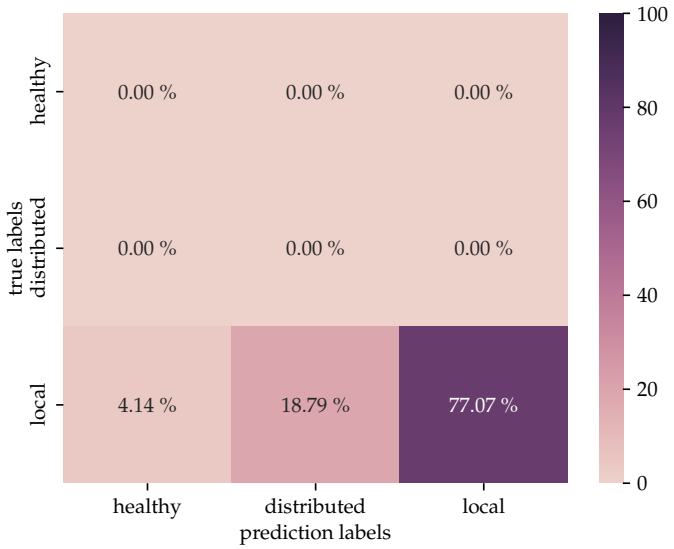


Figure 5. Confusion matrix for the classifier with three sub-classes tested on the withheld class *local 1*.

Because not all unknown damages can be classified in the defined subclasses, a second investigation differentiating only between healthy and damaged using the binary classifier is performed. Once again, every damage class is withheld one time and tested afterwards. The results are shown in Table 7. Like all other investigations, this is performed under tenfold cross validation, and therefore the average result as well as the best and the worst iterations are presented. Only class *local 0* shows slight difficulties, whereas all other classes exceed the average accuracies of over 0.95. For this reason, it can be derived that a reliable classification of unknown damage conditions into the classes healthy and damaged is achievable.

Table 7. Classifier results for withheld classes with two subclasses.

Unknown Class	ACC_{mean}	ACC_{max}	ACC_{min}
dist 0	0.96	0.99	0.90
dist 1	0.98	1	0.97
dist 2	1	1	1
local 0	0.88	0.98	0.72
local 1	0.98	1	0.97
local 2	0.95	0.98	0.93

5. Conclusions

In this contribution, an investigation of the detection of various bearing damages is presented. Despite the measurement of the bearings in the compound of the two-stage planetary gearbox, which results in the superposition of vibration events, nonlinear transmisson paths, low-speed components and noise, the results show that a reliable detection of the bearing damage is achievable. A distinction can be made between six presented damage phenomena. For this purpose, vibration signals are collected by using an accelerometer. The vibration signals are evaluated by creating spectrograms, which are used by CNNs as input data.

In addition to the investigation of damage classification, complementary investigations regarding the robustness of the developed system are performed. These are intended to

test the robustness of the system for field application as well as the expected variance in the damage cases. The results show that the classifier is robust for variations in the speeds within the trained speed ranges. Also in most cases, the classifier can achieve high accuracies in classifying unknown damages in the presented three sub-classes. However, due to difficulties with individual damage phenomena, the use of a binary classifier, which differentiates only between healthy and damaged, is recommended, especially for high-risk applications with an expected variance of damage phenomena to occur.

Since the classifier shows high robustness regarding unknown parameters, a transferability to planetary gearboxes with a different number of stages can be assumed. For this purpose, the network should probably be retrained. But due to missing access to equivalent data of a planetary gear with a different number of stages, a verification is not possible within this contribution. Therefore, the presented results are limited to the investigated two-stage planetary gearbox.

For future work, the classifier should be tested and optimized for bearing damages in the two-stage planetary gearbox installed in vehicles. Since the presented classifier can detect unknown damage phenomena, good transferability is assumed. Since unknown parameters, like additional noise or the influence of the load on the characteristics of the damage phenomena, are introduced in the field application, an adaptation of the classifier and additional filters are probably needed.

Author Contributions: Conceptualization, J.S.; methodology, J.S.; software, J.S. and L.J.B.; validation, J.S. and L.J.B.; formal analysis, J.S. and L.J.B.; investigation, J.S.; resources, J.S. and K.-L.K.; data curation, J.S.; writing—original draft preparation, J.S.; writing—review and editing, J.S., L.J.B., F.C. and K.-L.K.; visualization, J.S. and L.J.B.; supervision, K.-L.K.; project administration, K.-L.K.; funding acquisition, K.-L.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by the German Federal Ministry of Economic Affairs and Climate Action project KISS (19I21015D).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

ACC	Accuracy
BAC	Balanced accuracy
BPFI	Ball pass-frequency inner race
BPFFI	Ball pass-frequency factor inner race
BPFO	Ball pass-frequency outer race
BPFFO	Ball pass-frequency factor outer race
BSF	Ball spin frequency
BSFF	Ball spin-frequency factor
CM	Confusion matrix
CNN	Convolutional neural network
FMD	Feature mode decomposition
FN	False negative
FP	False positive
FTF	Fundamental train frequency
FTFF	Fundamental train-frequency factor
TN	True negative
TP	True positive

References

1. Van Hecke, B.; Yoon, J.; He, D. Low speed bearing fault diagnosis using acoustic emission sensors. *Appl. Accoust.* **2016**, *105*, 35–44. [CrossRef]
2. Zhang, S.; Zhang, S.; Wang, B.; Habetler, T. Machine Learning and Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review. *IEEE Access* **2020**, *8*, 29857–29881. [CrossRef]
3. Segla, M.; Wang, S.; Wang, F. Bearing fault diagnosis with an improved high frequency resonance technique. In Proceedings of the IEEE 10th International Conference on Industrial Informatics (INDIN), Beijing, China, 25–27 July 2012; Volume 10, pp. 580–585.
4. Smith, W.; Randall, R. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64–65*, 100–131. [CrossRef]
5. Feng, Z.; Zhang, D.; Zuo, M. Adaptive Mode Decomposition Methods and Their Applications in Signal Analysis for Machinery Fault Diagnosis: A Review With Examples. *IEEE Access* **2017**, *5*, 24301–24331. [CrossRef]
6. Miao, Y.; Zhang, B.; Li, C.; Lin, J.; Zhang, D. Feature Mode Decomposition: New Decomposition Theory for Rotating Machinery Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2023**, *70*, 1949–1960. [CrossRef]
7. Beering, A.; Döring, J.; Krieger, K.-L. Feature-based analysis of reproducible bearing damages based on a neural network. In Proceedings of the GMA/ITG-Fachtagung Sensoren und Messsysteme 2019, Nürnberg, Germany, 25–26 June 2019; Volume 20, pp. 451–456.
8. Yongbo, L.; Shubin, S.; Zhiliang, L.; Xihui, L. Review of local mean decomposition and its application in fault diagnosis of rotating machinery. *J. Syst. Eng. Electron.* **2019**, *30*, 799–814.
9. Lei, Y.; Lin, Y.; Zuo, M.; He, Z. Condition monitoring and fault diagnosis of planetary gearboxes: A review. *Measurement* **2014**, *48*, 292–305. [CrossRef]
10. Fischer, R.; Kücükay, F.; Jürgens, G.; Pollak, B. *Das Getriebebuch*, 2nd ed.; Springer: Wiesbaden, Germany, 2016; pp. 123–135.
11. Leaman, F.; Vicuna, C; Clausen, E. A Review of Gear Fault Diagnosis of Planetary Gearboxes Using Acoustic Emissions. *Acoust. Aust.* **2021**, *49*, 265–272. [CrossRef]
12. Zhang, X.; Wang, L.; Miao, Q. Fault diagnosis techniques for planetary gearboxes under variable conditions: A review. In Proceedings of the 2016 Prognostics and System Health Management Conference, Chengdu, China, 19–21 October 2016; pp. 1–11.
13. Mahgoun, H.; Ziani, R. Bearing Diagnostics Using Time-Frequency Filtering and EEMD. In *Rotating Machinery and Signal Processing*; Felkaoui, A., Chaari, F., Haddar, M., Eds.; Springer: Cham, Switzerland, 2019; pp. 44–55.
14. Tandon, N.; Choudhury, A. An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect. *J. Sound Vib.* **1997**, *205*, 275–292. [CrossRef]
15. Taye, M. Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. *Computation* **2023**, *11*, 52. [CrossRef]
16. Yamashita, R.; Nishio, M.; Do, R.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef] [PubMed]
17. Dhillon, A.; Verma, G. Convolutional neural network: A review of models, methodologies and applications to object detection. *Prog. Artif. Intell.* **2020**, *9*, 85–112. [CrossRef]
18. Zhang, A.; Lipton, Z.; Li, M.; Smola, A. Dive into deep learning. *arXiv* **2021**, arXiv:2106.11342.
19. Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* **2016**, *93*, 490–502. [CrossRef]
20. Ajit, A.; Acharya, K.; Samanta, A. A Review of Convolutional Neural Networks. In Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 24–25 February 2020.
21. Qian, W.; Li, S.; Wang, J.; An, Z.; Jiang, X. An intelligent fault diagnosis framework for raw vibration signals: Adaptive overlapping convolutional neural network. *Meas. Sci. Technol.* **2018**, *29*, 095009. [CrossRef]
22. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; Volume 2, pp. 1137–1143.
23. Nandi, A.; Ahmed, H. Classification Algorithm Validation. In *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machine*; Nandi, A., Ahmed, H., Eds.; John Wiley & Sons, Ltd.: New York, NY, USA, 2019; pp. 307–319.
24. Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-Validation. In *Encyclopedia of Database Systems*; Liu, L., Ed.; Springer: Boston, MA, USA, 2009; pp. 532–538.
25. Tharwat, A. Classification assesment methods. *Appl. Comput. Inform.* **2021**, *17*, 168–192. [CrossRef]
26. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**, arXiv:2008.05756.
27. Brodersen, K.; Ong, C.; Stephan, K.; Buhmann, J. The Balanced Accuracy and Its Posterior Distribution. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3121–3124.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.