Article

Reduced Data Volumes through Hybrid Machine Learning Compared to Conventional Machine Learning Demonstrated on Bearing Fault Classification

Simon Walther [1,†,*] and Axel Fuerst [1,†,*]

Institute for Intelligent Industrial Systems I3S, Department Engineering and Information Technology, Bern University of Applied Sciences, Pestalozzistrasse 20, 3400 Burgdorf, Switzerland
* Correspondence: simon.walther@bfh.ch (S.W.); axel.fuerst@bfh.ch (A.F.)
† These authors contributed equally to this work.

Abstract: In some real-world problems, machine learning is faced with little data due to limited resources such as sensors, time, and budget. In this case, the conventional machine learning approach may fail or perform badly. To develop a well-functioning model with a small training set the hybrid machine learning approach, the combination of different methods can be applied. Especially in the machine industry where Industry 4.0 is one of the most important topics - including condition monitoring, predictive maintenance, and automated data analyses - data are limited and costly. In this work, the conventional and hybrid approach are compared to the application of ball bearing fault classification. The dataset contains 12 different classes (11 with faults and 1 undamaged). For each approach, two different LSTM (Long Short-Term Memory) models are developed and trained on various training sets (different sensors). The hybrid model is realised by adding physical knowledge through applying fast Fourier transformation and frequency selection to the raw data. This study shows that the additional physical knowledge in the hybrid model results in a better performance of the hybrid machine learning than the conventional.

1. Introduction

In recent years, artificial intelligence has become part of our life. Big companies such as Google, Apple, Amazon, and Netflix use machine learning for different tasks such as image classification, speech recognition, and making suggestions for products or movies. For a well-functioning algorithm, a huge amount of data is needed.

In some cases, the available data are limited or costly and therefore the conventional machine learning approach may fail or perform badly. To ensure a well-functioning model with little data, hybrid machine learning can be applied. Hybrid means the combination of different methods, such as the combination of machine learning and physics—in form of a formula—or machine learning and methods from optimization theory.

Park et al. [1] analysed the power consumption of a chiller whose physics are well-known. In their work, a machine learning model developed with an artificial neural network (ANN) and a hybrid machine learning model are compared. The hybrid model is a combination of an artificial neural network and regression equations based on physics. Both models perform well in predicting the power consumption. The main difference is that the hybrid model requires fewer inputs (four inputs) than the conventional one (eight inputs). In addition to that, it is shown that the hybrid model can be applied where the conventional ANN is unable to provide accurate predictions, for example, even when the parameters are outside of the valid input space of the ANN. A combination of a deep learning model and a non-dominated sorting generic algorithm, used to solve multi objective optimization, is

developed by Heo et al. [2] to find the optimal setpoints of proportional integral controllers. This model is trained with five scenarios identified by using fuzzy c-means algorithms and physical ratios. The hybrid approach enables a cost-effective and sustainable operation of a full-scale wastewater treatment under varying inflow conditions. Liang et al. [3] combine two machine learning algorithms, K-means clustering and KNN classification, to reduce energy consumption by increasing the average utilization of physical machines in cloud data centres. The hybrid model is significantly better than the conventional ones.

Many real-world machine learning tasks are faced with little data. Forman and Cohen [4] compared different classifiers on small training sets. It is crucial to have meta-knowledge about which to apply in different situations. With little data, the question about the effect of the class distribution is important. Weiss and Provost [5] analyse the relationship between the class distribution of the training set and the performance of classification trees induced from these data.

In the machine industry, the monitoring of wear is a common research topic due to its potential for cost and resource savings. The fact that wear is in general not directly measurable requires the measure of a physical value that correlates with the state of wear to determine and interpret the condition of parts. Alonso and Salgado [6] measure the vibration of a tool longitudinally and transversely in the turning process with two accelerometers. The structure of the vibration measures is analysed using singular spectral analysis and clustering before presenting it to a neuronal network. Instead of using the vibrations, the acoustic emissions, recorded by a microphone, can be used. Li [7] shows methods based on acoustic emission sensing for tool wear monitoring in the turning process. Jia et al. [8] use deep neuronal networks for fault characteristic mining and intelligent diagnosis of rotating machinery. Methods based on artificial neuronal networks with a shallow architecture may need manually extracted features, much prior knowledge, and diagnostic expertise. Using deeper architectures, deep neuronal networks, and massive data enables fault characteristic mining and diagnosis direct from raw data without additional knowledge. Kong, Chen and Li [9] combine three machine learning algorithms, wavelet package decomposition, least square support vector machine, and the gravitational search algorithm, to monitor the tool wear states in the turning process. This method performs well even with small-sized training data sets and outperforms related methods for tool wear estimation such as neuronal networks and trees.

In the specific case of bearings, Widodo et al. [10] use a multi-class relevance vector machine and a support vector machine for fault classification. The data are acquired with a low-speed bearing test using acoustic emission and acceleration sensors. Mao et al. [11] compare different machine learning techniques such as support vector machines, artificial neural networks, and auto-encoders to detect and classify faults on the balls and the outer and inner rings. The vibration signals of ball bearings are highly complex, and for this case, Feng, Ma and Zuo [12] use spectral negentropy-based infograms. The faults are discerned according to the present peaks. Liu and Gryllas [13] built the feature space by using the indicators of the cyclic spectral analysis. The detection decision of the proposed method for rolling element bearings is made with a semi-supervised support vector machine. Saucedo-Dorantes et al. [14] developed a condition monitoring method for the detection of fault graduality in outer race bearings. The proposed method uses features of the current signals in the time and frequency domains in combination with a neuronal network classifier.

In the present work, a hybrid machine learning model is developed and compared to a conventional machine learning model. The idea is that the addition of, for example, physical knowledge to a machine learning algorithm performs better than the conventional machine learning approach. With the presented method, it is possible to have the same performance with less data inputs.

2. Background

2.1. Machine Industry

At present, Industry 4.0 is one of the most important topics in the Swiss machine industry. Research and development is moving in this direction. The main points of interest are connected and optimised systems, smart machines, and sensors. Regarding data science, there are tasks in condition monitoring, predictive maintenance, and automated data analyses to solve. Artificial intelligence is quite new to the machine industry and based on the missing understanding, the scepticism is quite high. To counter this, it is crucial to create confidence with understandable explanations, simple models, and a high reliability of the system.

The biggest difficulty for data science in the machine industry are the limited resources. The development time of a machine or a system can lie between months or a few years depending on the size and the complexity. In addition to that, the lifetime of the machine system can be up to 20 years or even more. During the whole lifetime, different components wear out and must be replaced. Therefore, a specific maintenance interval for each component is defined. The length of every interval is based on the experience and calculation of wear, including defined load spectra and operational conditions. Independent from the type of component, it is common that the end of life is not reached at the time of maintenance, and parts are replaced too early. An even worse scenario is when the component fails before the defined end of live. This results in downtime of the machines with a huge financial impact. In both cases, the scheduled maintenance costs money due to unnecessary or late maintenance. To prevent this, knowledge about the state of wear is essential. One possibility is to perform condition monitoring (estimating the state of wear) and another is to do predictive maintenance—in addition to the condition monitoring—that estimates the remaining lifetime and schedules the maintenance [15].

Apart from the whole topic of time and lifetime, there are other limiting factors that are faced during development of the system such as costs, space, and feasibility. Most of the sensors are not expensive, but each added sensor leads to a higher complexity of the system and makes the development costlier. The most expensive part is the data collection, where the system must be in operation and measurements over the whole lifetime are taken. The machines are designed as small as possible as necessary. The location of the sensors is given by the available space and the structure. A general advice is to use the sensors that are essential for the basic functions of the machine or the whole systems. These points are valid for existing machines as well as machines in the development phase. Apart from the technical difficulties, there are other concerns such as those about security and data protection. These are not relevant for this paper.

The main limitation that affects machine learning in the machine industry is the limited data based on the long lifetime and limited number of sensors that are relevant for the specific task. In most cases, artificial intelligence needs a huge amount of data to be trained and perform well. The lack of the required amount of data can lead the algorithms to fail. To obtain good results with little data, hybrid machine learning models can be used. These combine a conventional part with calculation or transformation with machine learning. With this method, additional knowledge is added and leads to more (specific) information that the algorithm can use and perform better. Furthermore, if physical knowledge is added, the algorithm may work for extrapolations as well, which is not the case for pure deep learning approaches.

2.2. Ball Bearings

The presented work uses data from ball bearings with different damages, types, and states of wear. All of them are created in the laboratory. To create different types and states of wear that can occur in normal operation, material is removed manually with a grinder to create pits. Bearings may corrode depending on the environment during operation. This is reproduced by degreasing the bearings with subsequent repeated dipping in saltwater. The result is shown in Figure 1. If a ball bearing is overloaded, this can lead to rolling body

marks. This case is reproduced with a static overload of the bearing. The used bearing type is 6203 from SKF. This is a single-row ball bearing without sealing that enables access to the running surface and the balls. The whole list of the tested damages is shown in Table 1.

Figure 1. Ball bearing with heavy corrosion.

Table 1. Classes of damage.

Class	Damage
0	Undamaged
1	Degreased
2	Light wear outer ring
3	Medium wear outer ring
4	Heavy wear outer ring
5	Heavy wear inner ring
6	Medium wear inner ring
7	Light wear inner ring
8	Light Corrosion
9	Medium corrosion
10	Heavy corrosion
11	Rolling body marks

The evaluation of bearings related to the VDI regulation 3832 [16] can be done by considering the envelope curve spectrum. Each defect has characteristic frequencies visible in the frequency spectrum. A ball bearing with an advanced damage on the outer ring has characteristic lines at rollover frequency (f_A) for the outer ring and its multiples (see Figure 2). The rollover frequency can be calculated with the formula below with the rotational frequency of the rotor, f_n; the number of rolling elements, Z; diameter of the rolling element, D_W; contact angle, α; and pitch circle diameter, D_{PW} [16]:

$$f_A = \frac{1}{2} f_n \cdot Z \left[1 - \frac{D_W \cdot \cos\alpha}{D_{PW}} \right]$$
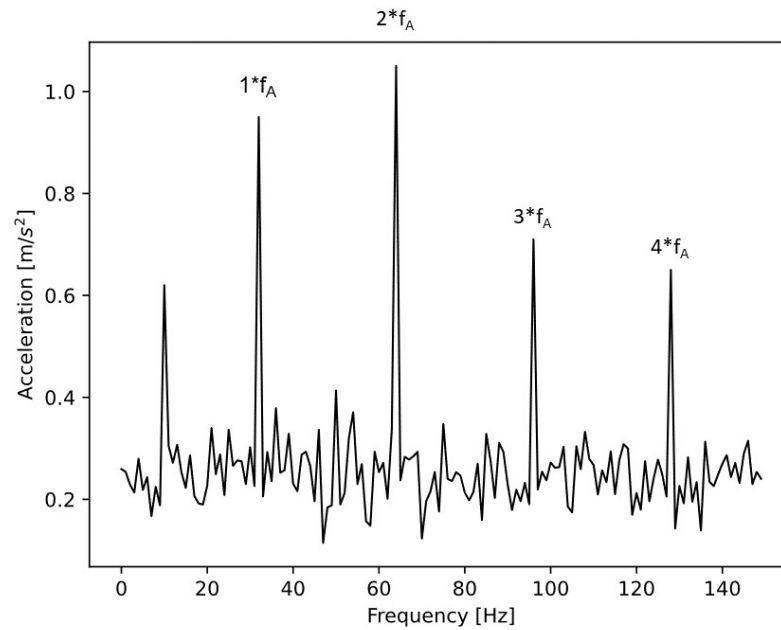
(1)

Figure 2. Envelope curve spectrum of a ball bearing with advanced outer ring damage.

The characteristic line of a damaged inner ring is completely different. The Sidebands are distributed around the rollover frequency (f_I) of the inner ring and its multiples with the distance of f_n (the rotor rotational frequency), see figure 3. The rollover frequency can be calculated with the formula below with the rotational frequency of the rotor, f_n; the number of rolling elements, Z; diameter of the rolling element, D_W; contact angle, α; and pitch circle diameter, D_{PW}[16]:

$$f_I = \frac{1}{2} f_n \cdot Z \left[1 + \frac{D_W \cdot cos\alpha}{D_{PW}} \right] \qquad (2)$$
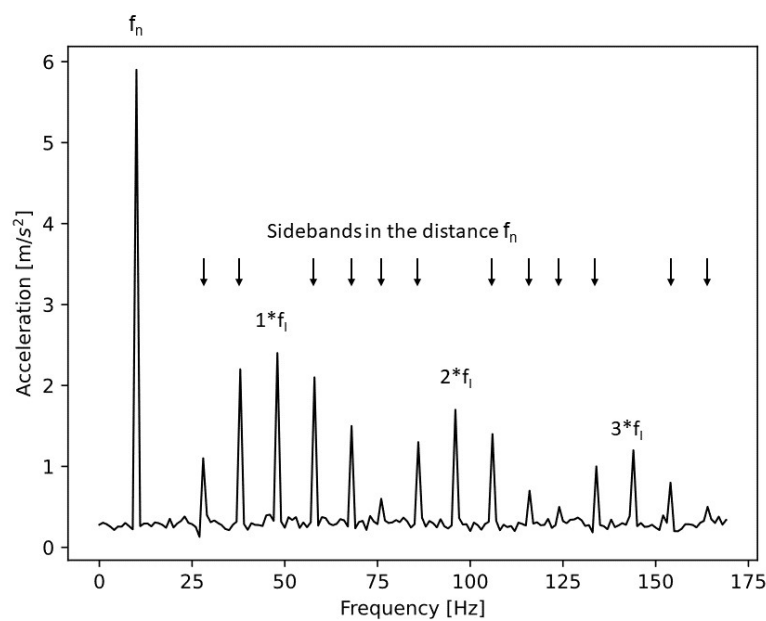

Figure 3. Envelope curve spectrum of a ball bearing with advanced inner ring damage.

Ball bearings with a rolling element damage have sidebands around the rollover frequency of the rolling element (f_w) and its multiples with the distance of f_k (the rotational frequency of the cage), see Figure 4. The rollover frequency of the rolling element can be calculated with the formula below with the rotational frequency of the rotor, f_n; diameter of the rolling element, D_W; contact angle, α; and pitch circle diameter, D_{PW} [16]:

$$f_W = \frac{f_n \cdot D_{PW}}{D_W} \left[1 - \left(\frac{D_W \cdot cos\alpha}{D_{PW}} \right)^2 \right] \tag{3}$$

The rotational frequency of the cage can be calculated with the formula below with the rotational frequency of the rotor, f_n; diameter of the rolling element, D_W; contact angle, α; and pitch circle diameter, D_{PW} [16]:

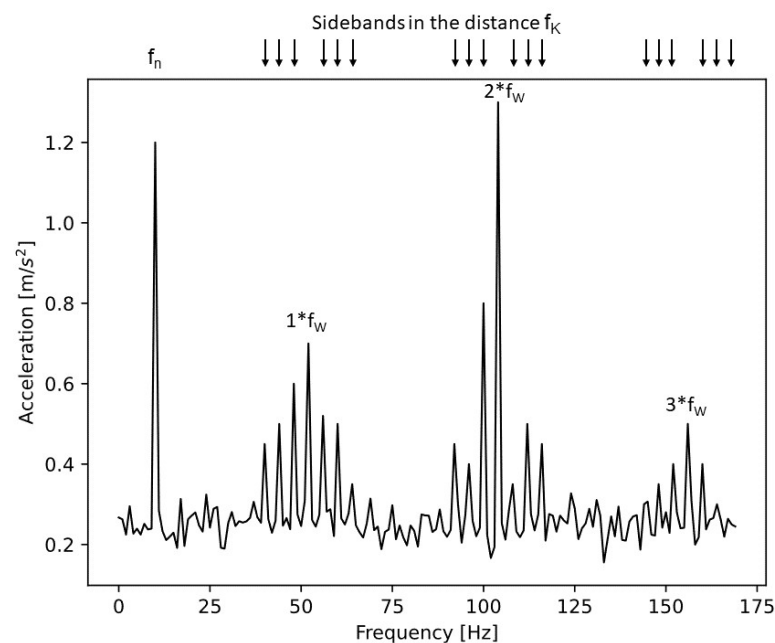$$f_K = \frac{1}{2} f_n \left[1 - \frac{D_W \cdot cos\alpha}{D_{PW}} \right] \tag{4}$$

Figure 4. Envelope curve spectrum of a ball bearing with advanced rolling element damage.

By comparing Figures 2–4, it becomes obvious that the envelope curve is completely different for each type of damage. Each type and state of wear and damage has different characteristic frequencies and sidebands. This fact enables, in combination with experience and knowledge about the envelope curves, manual classification by considering the frequency spectrum.

2.3. Experimental Setup

While the mounted bearing is in simulated operation, the sound and the acceleration are recorded with an unidirectional microphone (Type 426A03 from PCB) and with a triaxial acceleration sensor (Type 356A16 from PCB). The idea to use a microphone in addition to the common acceleration sensor is to obtain additional information or to use it instead if possible because it must not directly be placed on the bearing. To record the signals, the sound and vibration modules NI 9232 (microphone) and NI9234 (acceleration sensor) from National Instruments are used. The measuring concept is shown in Figure 5.
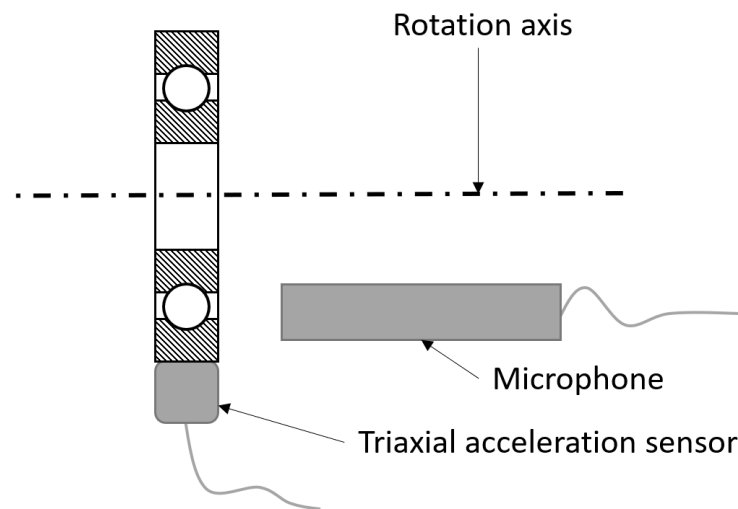
Figure 5. Scheme of the measuring concept.

The coordinate system of the triaxial acceleration is shown in Figure 6 on the real measuiring setup.

Figure 6. Coordinate system of the triaxial acceleration sensor.

Before measuring, the bearings are warmed up for 15 min at the evaluation speed of 500 rounds per minute given by the measuring setup. After this period, 10 measurements with a duration of 5 seconds each using a sampling rate of 51.2 kHz are recorded with no additional load applied to the bearing and directly stored in an HDF5 file. The collected dataset contains 12 classes with 10 measurements each.

3. Methodology

3.1. In General

3.1.1. Conventional Machine Learning

The methodology to develop a conventional machine learning model is visualised in Figure 7. The first step from the raw data to a dataset is called pre-processing. This can include different modifications and varies from project to project. Depending on the quality of the data and the measurements, a trend or outliers may exist and be removed. Additional filtering can increase the performance of the algorithm. This can be implemented and tested when the whole machine learning pipeline is set up. The data can be unbalanced due to

the measurement strategy or even by nature. For example, in a measuring campaign, the interesting sections can be rare. To avoid a biased model, the data must be balanced by removing or adding generating data. After this step, the data are split in the three datasets: training, test and validation.
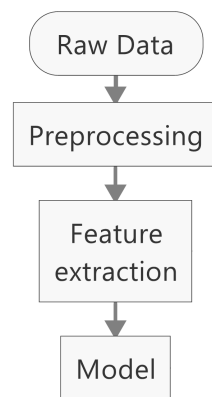
```
Raw Data
   ↓
Preprocessing
   ↓
Feature
extraction
   ↓
Model
```

Figure 7. Flowchart of conventional machine learning.

Before entering the dataset in the model, the features are extracted. For deep learning, this step is obsolete because the algorithm includes this part. The other algorithms have the need of feature engineering. In other words, the feature extraction must be programmed and the features (pattern, statistical values, etc.) chosen. The extracted features are the inputs of the model.

To complete the machine learning pipeline, the algorithm must be chosen. The experience of the engineer or a good literature research is crucial to find the most suitable algorithm, create the model, and tune the hyperparameters. In general, it is required to try different algorithms, compare the performance of them and choose the best for the final model.

3.1.2. Hybrid Machine Learning

The methodology of hybrid machine learning, shown in Figure 8, is slightly different compared to the conventional method. The additional step, in the presented case called "apply physical knowledge", can mean to apply a physical formula to the pre-processed data, pattern search and selection or using an additional algorithm. The modification of the data has the effect of additional knowledge and enables a better performance of the algorithm.
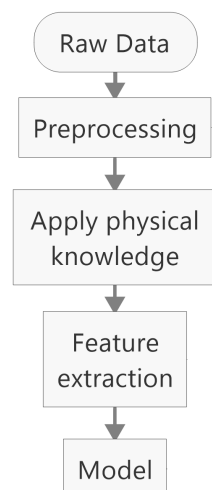
```
Raw Data
   ↓
Preprocessing
   ↓
Apply physical
knowledge
   ↓
Feature
extraction
   ↓
Model
```

Figure 8. Flowchart of hybrid machine learning.

If a deep learning algorithm is used, it seems strange to calculate or transform the data before entering the data in the model. However, the presented work is about machine learning with little data, and therefore, every additional piece of information, calculation or transformation is essential to increase the performance and the chance of success.

3.2. The Case of Bearing Fault Classification

3.2.1. Conventional Machine Learning Model

As shown in Table 1, there are 12 different classes of damage. The goal is to build a model that can classify the measurements from the triaxial acceleration sensor and the microphone. To classify timeseries, the Long Short-Term Memory is a suitable algorithm. It is widely used for text and signal classification, as well as generation.

An LSTM network contains a number of LSTM cells. As the name says, these cells have a long- and short-term memory. This means that each cell has knowledge about the history and the present of the signal. For classification, prediction or generation of signals and text, the history is crucial. For example, to understand a sentence, it is important to have the knowledge about the previous words and not just look at the latest word. For text and signals, the history is important to understand the present, and the LSTM is able to do that. More information about machine learning in general and the LSTM network can be found in *Hands-On Machine Learning* [17].

The machine learning pipeline is set up as shown in Figure 9 with an LSTM network model. This method does not require a feature extraction step since this task is integrated in the algorithm. The raw data show no trend, have no outliers and no filter is applied. The dataset for the training is directly created from the raw data.
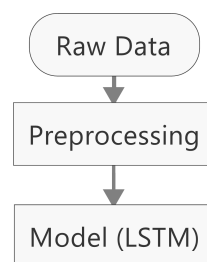

Figure 9. Flowchart of the LSTM model.

During the engineering phase, two different LSTM models are developed to classify the timeseries. The first model, shown in Figure 10, contains a single hidden LSTM layer for the feature extraction out of the data. To reduce the risk of overfitting the model to the training data, a dropout layer is added after the LSTM. The extracted features are interpreted by a fully connected dense layer. At the end, a fully connected dense layer is used to make the decision and provide the output.

The second model is very similar to the first one. Regarding the very little data and the variety of different classes, the extracted features are key for a good performance of the model. For more complex features, a second hidden LSTM layer is added. This additional layer makes the neural network deeper and can take the learnings form the previous layer and combine it to obtain features at a higher level of abstraction. The two hidden layers are followed by a dropout and two fully connected dense layers to interpret and make the decision as in the first model. The second model is shown in Figure 11.
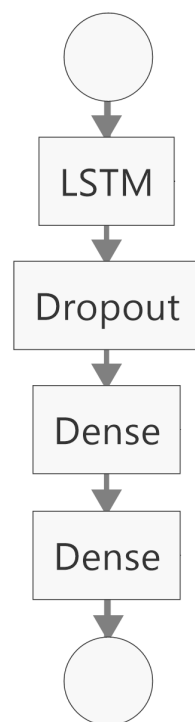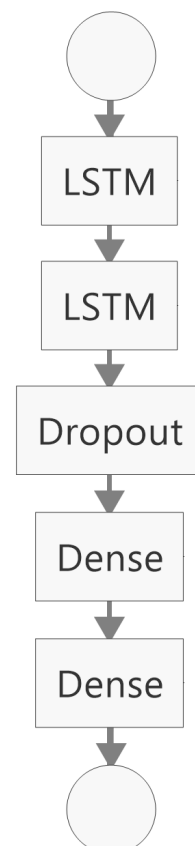
Figure 10. Model with one LSTM-layer.

Figure 11. Model with two LSTM-layer.

As input, there are the four following different measures available:

- Recorded sound from the unidirectional microphone;

- Three axes of the acceleration sensor.

The conventional machine learning model needs all four inputs to perform better than chance. This subject is treated in more detail in the chapter results. To classify the different cases, a class vector that represents all categories is needed. Every category is assigned to a number, see Table 1. A neuronal network needs multi-class classification, the classification of more than two classes, a binary class matrix where each column represents a class and each row represents a measurement and a sample. The category of each sample is assigned to its category by a one in the class-related column. The rest of the row are zeros.

3.2.2. Hybrid Machine Learning Model

The hybrid machine learning pipeline, see Figure 12, is set up like the conventional model with an additional step of adding knowledge.
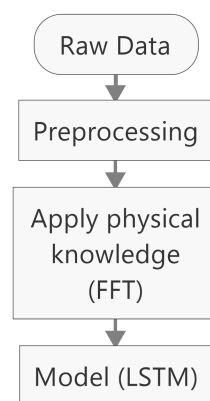

Figure 12. Flowchart of the hybrid LSTM model.

As already mentioned, ball bearings have characteristic frequencies such as the rotational frequency, the doubled rotational frequency caused by asymmetries of the rotor and the number of balls multiplied with the rotational frequency. The signal of a new and unused bearing has a specific spectrum of frequencies. Each frequency has a related importance (value of the amplitude). If the bearing becomes damaged or faces wear, the spectrum and the importance change. This means that, for example, sidebands become visible as an indication of a damaged inner ring. The strategy is to extract frequencies that correspond with a common damage of this specific ball bearing in a separate loop and sensitize the algorithm to these few frequencies to improve its performance and reliability.

For the presented work, the measurement data are transformed into the frequency space with the fast Fourier transformation (FFT). The FFT of the signal itself is not enough to obtain a better performance with the existing amount of data. As already mentioned, the characteristic frequencies and sidebands are indicators for different types of wear and damages. The positions of these peaks are known (visible in Figures 2–4), so it is easy to detect and extract them. The finally used number of peaks is optimised in different runs.

The used LSTM model for the hybrid machine learning model is the same as in the conventional machine learning pipeline, shown in Figures 10 and 11, to show the benefit out of the additional knowledge.

4. Results

As the chapter methodology mentioned, two different LSTM models with one and two LSTM layers are developed. The four following machine learning pipelines are built:

- Conventional one-layer LSTM;
- Conventional two-layer LSTM;
- Hybrid one-layer LSTM;
- Hybrid two-layer LSTM.

Each of the above-mentioned models is trained on four different training sets. These sets contain the following sensor data:

- All sensor data (accX, accY, accZ, mic);
- Acceleration in direction of X (accX);
- Acceleration in direction of Z (accZ);
- Sound captured with the microphone (mic).

After the training and validation of the models on all sets, the accuracy of the training and testing can be compared to find the best model.

4.1. Conventional Machine Learning

A measurement (5 s) consists of 256,000 data points. This array is too big to be directly inserted into the machine learning model and would lead to a long-lasting training as well as a badly trained model. Thus, the dataset is split into smaller samples. This means that out of each measurement, 1000 samples with 256 data points are created. As result of this, the dataset contains 12 classes with 10,000 samples per sensor. The number of usable samples increased by a factor of 1000.

The results of the conventional machine learning model with one LSTM layer are shown in Table 2. Using all sensors for the training set the accuracy of the model is very good. However, this work has the focus on limited resources, thus the results of a single sensor in the set are of particular interest. The algorithm shows the best performance on the acceleration in the Z (radial) direction. The rollover of fault results in a radial vibration and acceleration, and this is the reason why the algorithm performs well with this sensor. The acceleration in the X direction is the lateral acceleration of the bearing position and its signal is similar to the radial acceleration but not equal. The change from directly radial to lateral measured acceleration results in a minor loss of information in the signal and performance of the algorithm. The training with the acoustic measurements from the microphone has the worst performance. The acoustic signal is very complex and can even contain ambient noise. In combination with little data, the algorithm is not able to extract the features belonging to the different faults.

Table 2. Acuracy of the conventional LSTM model (one LSTM-layer).

Used Sensor(s)	accX, accY accZ, mic	accX	accZ	mic
Train accuracy	96.34%	90.98%	94.38%	70.48%
Test accuracy	94.77%	87.95%	90.61%	68.17%

In general, adding a second LSTM layer has a positive effect on the performance of the conventional machine learning pipeline (see table 3) except the training with the lateral acceleration (accX). In this case, the model fails completely and has a performance similar to guessing because the features are too complex to interpret. For the microphone, the radial acceleration and the combination of all sensors the more complex features extracted by the second LSTM layer result in a small increase in accuracy.

Table 3. Accuracy of the conventional LSTM model (two LSTM-layer).

Used Sensor(s)	accX, accY accZ, mic	accX	accZ	mic
Train accuracy	98.54%	8.78%	96.62%	77.98%
Test accuracy	95.89%	10.11%	92.43%	72.75%

Figure 13 shows the confusion matrix of the two-layer LSTM trained on the data from the microphone. The most difficult category to predict is the number five (medium wear inner ring). There some are classified as two (light wear outer ring), three (medium wear outer ring) and six (medium wear inner ring). The classification as a six may be the result of a small difference in wear. A difference in the rotational frequency results in a shift in the characteristic frequencies. With this shift, it is possible that the wear of the inner ring can be classified as outer ring wear. This explains even the rest of the confusion between classes two to seven.
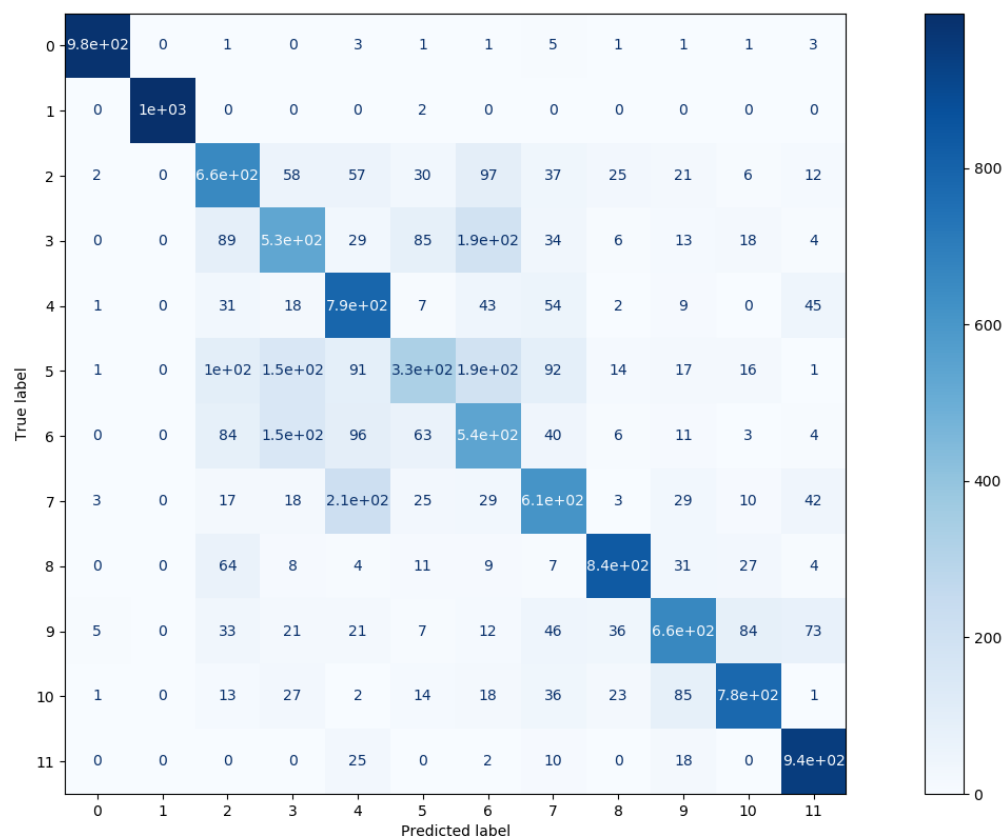

Figure 13. Confusion matrix of the conventional two-layer LSTM trained on mic.

The best conventional model reaches a test accuracy of 95.9% (all sensors) and the worst, 10.1% (mic).

4.2. Hybrid Machine Learning

The extracted number of peaks (sidebands and characteristic frequencies) varies between the different faults, measures and sensors. After several tests, the number of used peaks is set to the 100 with the highest importance. The dataset contains 12 classes with 10 samples per sensor. One sample corresponds the FFT of one measurement.

The results of the hybrid machine learning model with one LSTM layer are shown in Table 4. The algorithm performs very good on all sets. It is conspicuous that the set using all sensors is not the best. The reason for this is the very limited amount of data. The algorithm struggles to fit a function to that number of features with the existing data.

Adding a second LSTM layer has different effects on the performance of the hybrid machine learning pipeline (see Table 5). To extract features with a higher complexity results in a better performance for the model trained on all sensors. The test accuracies of the two models trained on acceleration signals, radial and lateral, become slightly worse. This

indicates that there is no advantage in extracting more complex features. The additional layer does not affect the algorithm trained on the microphone data.

Table 4. Accuracy of the hybrid LSTM model (one LSTM layer).

Used Sensor(s)	accX, accY accZ, mic	accX	accZ	mic
Train accuracy	96.30%	99.07%	99.07%	98.15%
Test accuracy	91.67%	100.00%	100.00%	91.67%

Table 5. Accuracy of the hybrid LSTM model (two LSTM layers).

Used Sensor(s)	accX, accY accZ, mic	accX	accZ	mic
Train accuracy	100.00%	97.22%	100.00%	98.15%
Test accuracy	100.00%	91.67%	91.67%	91.67%

Using one single sensor signal ,the model with one LSTM layer performs better and only using all sensors at the same time the additional layer has a positive influence on the accuracy.

Figure 14 shows the confusion matrix of the two-layer LSTM trained on the data from the microphone. The wrong classifications of classes two and three can be explained with a small difference in the state of wear. All false classifications that include corrosion or rolling body marks are based on the fact that there is a damaged part on the inner and outer ring.
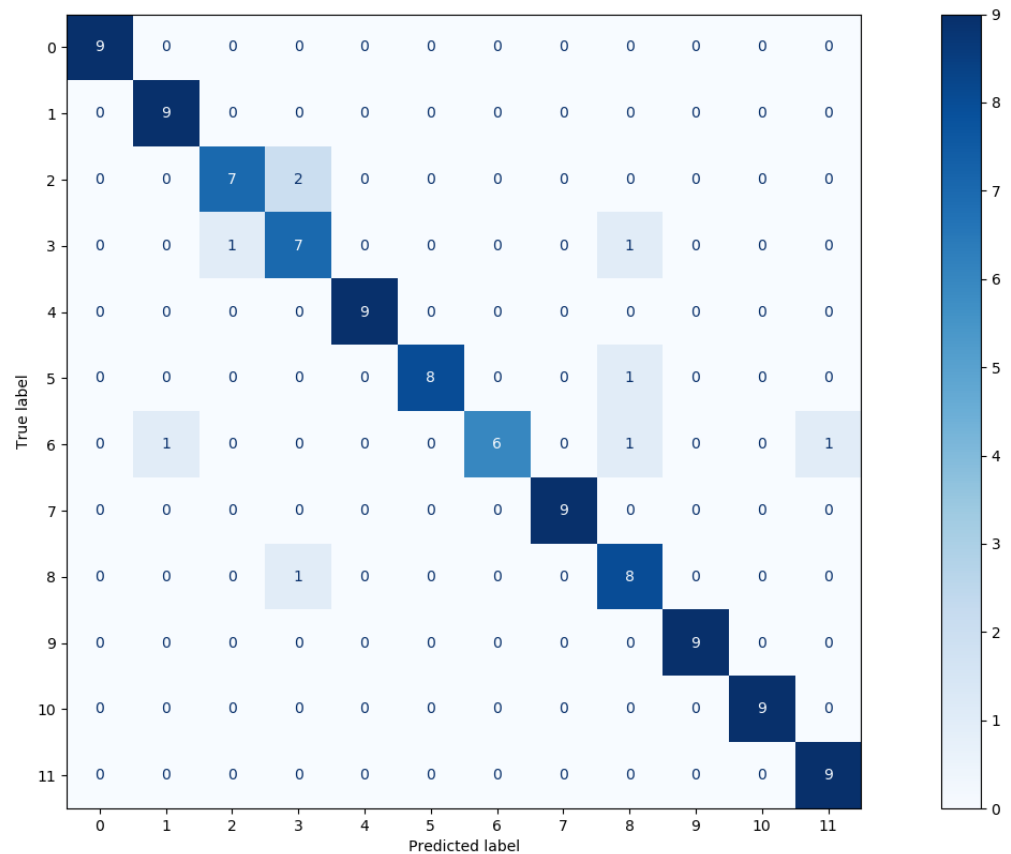

Figure 14. Confusion matrix of the hybrid two-layer LSTM trained on mic.

The best model achieves an accuracy of 100% (all sensors) and the worst, 91.7%. Compared to the conventional models the best hybrid is 4.1% better. The difference between the worst models is 81.6%.

5. Discussion

The presented work is based on a very simple case where a single bearing is observed and the environmental conditions are optimal. In general, the conventional machine learning approach is easier and more straightforward to implement. There is no need of deep knowledge and research about the system and its behaviour. However, as the results show in some cases, this approach can reach its limits very fast. If, for example, a signal with a complex content of information, in this case, the recorded signal of the microphone, is used, the algorithm has difficulties performing well.

The results show better statistical values for the hybrid machine learning approach in all cases. Especially for more complex signals, such as the sound, the difference is significant. This fact is easily explainable. The existing amount of data is too small for the algorithm to extract the needed information and classify it correctly. However, if the understanding of the system and the physical knowledge is used to prepare the data so that the algorithm can work with it, then the performance is going to be good. In the presented work, it is the extraction of the characteristic frequencies and sidebands. With this information, the classification could even be executed manually. Furthermore, this strategy has a huge advantage for selling and defending the model because the decision is based on known physical facts and behaviours that can be visualised and explained.

The small difference between the results of the hybrid models shows that this method is quite robust. However, for a deeper analysis, the amount of available data is too small.

6. Conclusions

This study shows that hybrid machine learning performs better than conventional machine learning if only limited data are available. Comparing the two approaches for each sensor and sensor combination, a hybrid pipeline is always best. In addition, by considering the different number of samples per sensor in the data set, 10'000 for the conventional using timeseries and 10 for the hybrid using frequencies, the advantage of the hybrid approach is very clear. This specific case shows that additional physical knowledge can improve the performance of machine learning models. The idea of the presented hybrid method is not new. In this work, a ready-to-use model is developed, and its simplicity is shown compared to the conventional one. The basis of this method is the knowledge about the physical behaviour. The limitation is the level of complexity that the engineer can handle and the know-how about the physics.

In general, it is requested to take account of hybrid machine learning especially when data and resources are limited. The knowledge about the system and its behaviour are crucial to succeed with little data and apply, for example, physical knowledge to the data, as is done in the presented work. In a future project, the hybrid machine learning approach is applied on a more complex task to proof the reliability and its robustness.

References

1. Park, S.; Ahn, K.U.; Hwang, S.; Choi, S.; Park, C.S. Machine learning vs. hybrid machine learning model for optimal operation of a chiller. *Taylor Fr. Sci. Technol. Built Environ.* **2019**, *25*, 209–220.
2. Heo, S.; Nam, K.; Tariq, S.; Lim, J.Y.; Park, J.; Yoo, C. A hybrid machine learning–based multi-objective supervisory control strategy of a full-scale wastewater treatment for cost-effective and sustainable operation under varying influent conditions. *J. Clean. Prod.* **2021**, *291*, 125853.
3. Liang, B.; Wu, D.; Wu, P.; Su, Y. An energy-aware resource deployment algorithm for cloud data centers based on dynamic hybrid machine learning. *Knowl.-Based Syst.* **2021**, *222*, 107020.
4. Forman, G.; Cohen, I. Learning from Little: Comparison of Classifiers Given Little Training. In *European Conference on Principles of Data Mining and Knowledge Discovery*; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3202, pp. 161–172.
5. Weiss, G.M.; Provost, F. Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction. *J. Artif. Intell. Res.* **2003**, *19*, 315–354.
6. Alonso, F.J.; Salgado, D.R. Analysis of the structure of vibration signals for tool wear detection. *Mech. Syst. Signal Process.* **2008**, *22*, 735–748.
7. Li, X. A brief review: acoustic emission method for tool wear monitoring during turning. *Int. J. Mach. Tools Manuf.* **2002**, *42*, 157–165.
8. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72–73*, 303–315.
9. Kong, D.; Chen, Y.; Li, N. Monitoring tool wear using wavelet package decomposition and a novel gravitational search algorithm–least square support vector machine model. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2020**, *234*, 822–836.
10. Widodo, A.; Kim, E.Y.; Son, J.D.; Yang, B.S.; Tan, A.C.; Gu, D.S.; Mathew, J. Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. *Expert Syst. Appl.* **2009**, *36*, 7252–7261.
11. Mao, W.; He, J.; Li, Y.; Yan, Y. Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study. *Proc. IMechE Part C J. Mech. Eng. Sci.* **2016**, *0* , 1–19.https://doi.org/10.1177/0954406216675896
12. Feng, Z.; Ma, H.; Zuo, M.J. Spectral negentropy based sidebands and demodulation analysis for planet bearing fault diagnosis. *J. Sound Vib.* **2017**, *410*, 124–150.
13. Liu, C.; Gryllias, K. A semi-supervised Support Vector Data Description-based fault detection method for rolling element bearings based on cyclic spectral analysis. *Mech. Syst. Signal Process.* **2020**, *140*, 106682.
14. Saucedo-Dorantes, J.J.; Zamudio-Ramirez, I.; Cureno-Osornio, J.; Osornio-Rios, R.A.; Antonino-Daviu, J. Condition Monitoring Method for the Detection of Fault Graduality in Outer Race Bearing Based on Vibration-Current Fusion, Statistical Features and Neural Network. *Appl. Sci.* **2021**, *11*, 8033.
15. Axel Fuerst. Nutzen schaffen mit Industrie 4.0. In *Industrie 4.0*; Swiss Engineering STV: Zürich, Schweiz, 2019.
16. VDI 3832:2013. *Measurement of Structure-Borne Sound of Rolling Element Bearings in Machines and Plants for Evaluation of Condition*; Engl. VDI-Gesellschaft Produkt- und Prozessgestaltung: Dusseldorf, Germany,2013.
17. Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems*, 1st ed.; O'Reilly Media: Sebastopol, CA, USA, 2017.