engineering proceedings

MDPI

Proceeding Paper

Using Learned Health Indicators and Deep Sequence Models to Predict Industrial Machine Health †

Ido Amihai [1,*], Arzam Kotriwala [1], Diego Pareschi [2], Moncef Chioua [3] and Ralf Gitzel [1]

[1] ABB Corporate Research Center, 68526 Ladenburg, Germany; arzam.kotriwala@de.abb.com (A.K.); ralf.gitzel@de.abb.com (R.G.)
[2] ABB, 2629 JD Delft, The Netherlands; diego.pareschi@nl.abb.com
[3] Polytechnique Montréal, Montréal, QC H3T 1J4 QC, Canada; moncef.chioua@polymtl.ca
* Correspondence: ido.amihai@de.abb.com; Tel.: +49-6203-716041
† Presented at the 7th International conference on Time Series and Forecasting, Gran Canaria, Spain, 19–21 July 2021.

Abstract: In this paper, we describe a machine learning approach for predicting machine health indicators with a large time horizon into the future. The approach uses state-of-the-art neural network architectures for sequence modelling and can incorporate numerical-sensor and categorical data using entity embeddings. Moreover, we describe an unsupervised labelling approach where classes are generated using continuous sensor values in the training data and a clustering algorithm. To validate our approach, we performed an ablation study to verify the effectiveness of each of our model's components. In this context, we show that entity embeddings can be used to generate effective features from categorical inputs, that state-of-the-art models, while originally developed for a different set of problems, can nonetheless be transferred to perform industrial asset health classification and provide a performance boost over simpler networks that have been traditionally used, such as relatively shallow recurrent or convolutional networks. Taken together, we present a machine health monitoring system that can accurately generate asset health predictions. This system can incorporate both numerical and categorical information, the current state-of-the-art for sequence modelling, and generate labels in an unsupervised fashion when explicit labels are unavailable.

Keywords: neural networks; time series; sequence modelling; machine health monitoring; predictive maintenance

Citation: Amihai, I.; Kotriwala, A.; Pareschi, D.; Chioua, M.; Gitzel, R. Using Learned Health Indicators and Deep Sequence Models to Predict Industrial Machine Health . *Eng. Proc.* **2021**, *5*, 7. https://doi.org/10.3390/engproc2021005007

Academic Editors: Ignacio Rojas, Fernando Rojas, Luis Javier Herrera and Hector Pomare

Published: 25 June 2021

1. Introduction

Modern machine health monitoring systems (MHMS) owe much of their recent success to advances in machine learning algorithms, sensing technologies, and computational power [1–5]. Such systems make use of historical data collected from the monitored equipment, which are used to train machine learning (ML) models for evaluating their health and performance [1], in either a diagnostic or prognostic way, e.g., by remaining useful life estimation (RUL; e.g., [4,6]).

Historically, MHMS were based on ML algorithms that require hand crafted features. However, the utility of such models was limited due to the required domain expertise and inability to cover all spectrum effects, especially nonlinear dependencies in time and domain-specific effects [1]. A mitigation strategy for this problem is to use neural networks (NN), which do not require handcrafted features and can be trained using only the input data (e.g., [1,7–11]).

In the context of sequential data, several NN architecture-types have typically been applied based on their proficiency in learning the temporal dynamic behaviours of systems. In this respect, recurrent neural networks (RNNs) have been extensively used to model

sequential data [12]. Although different variants exist, an RNN is normally constructed as an NN with a feedback loop from the previous hidden layer of the network to the next:

$$h(t) = f(h(t-1), X(t); \theta), \quad (1)$$

where h(t) and X(t) are the hidden states and inputs to the network at time t, and θ is the network parameters.

Although RNNs are typically difficult to train due to issues with vanishing and exploding gradients [13], this can be mitigated by using gate functions that regulate the information that passes through the network. This is usually done through long short-term memory (LSTM) or gated recurrent units (GRU) [12], which, instead of the ordinary RNN transition function, involve more complex functions that incorporate gate structures that help regulate the information that passes through the network [14,15]. Other NNs used to model sequential data that are based on RNNs are echo state networks (ESN) [16,17]. ESNs mitigate the vanishing gradient problem by eliminating the need to compute the gradient for the hidden layers of the NN using a sparsely connected RNN called a "reservoir", where the weights are not learned via gradient descent [18].

In addition to RNN based architectures, convolutional NNs (CNNs) have also been used for sequence modelling. CNNs utilize convolutional operations, which are sliding filters that are applied over the data and enable the NN to extract time-invariant nonlinear features [19]. Recently it was demonstrated that CNNs coupled with residual connections, which are connections between an NN layer and a layer it is not directly connected to, can result in highly accurate models for sequential data [19]. An example of this type of architecture is the inception-time network [19], which is one of the architectures we implemented in this research and was inspired by the Inception-v4 architecture [20]. Crucially, it contains "Inception Modules", where the core idea is to simultaneously apply multiple convolutional filters of varying dimensions to the input [21].

Finally, the relatively new transformer architecture-type has also been successfully utilized for sequence modelling (e.g., [22]). These models rely on self-attention mechanisms to model temporal dynamics [23], the most common being the "scaled dot-product attention", "dot-product attention", and "additive attention" [23]. The scaled dot-product attention is computed via the following equation:

$$\text{Attention}(Q,K,V) = \text{softmax}((QK^T)/\sqrt{(d\_k)})\cdot V, \quad (2)$$

where matrices Q, K, and V are generated for each input, and where dk is the dimension of Q, and K. Dot-product attention is identical except that the scaling factor $\sqrt{(d\_k)}$ is not used, and additive attention is computed using a feed-forward NN with a single hidden layer [23]. Although transformers were developed for natural language processing (NLP) applications (e.g., German-English translations), they can be adapted for sequential numerical data, in the simplest case by replacing the embedding layers with fully-connected layers or other layer types that can transform numerical data (e.g., time delay embeddings [22]). Other approaches used for sequence modelling include large memory storage retrieval NNs [9], stacked denoising autoencoders [11], and deep belief networks [8].

Another important issue that arises when developing MHMS stems from the fact that they are typically developed using supervised learning, where ML models are trained to classify the health status of assets based on labelled training examples with a known health status. However, often the relationship between available data and asset health is not known in advance (i.e., the data is unlabelled) and must be determined using statistical, ML, or other methods. To address this issue, we developed an unsupervised approach, where sensor data from the training set was used to generate clusters that represent the asset health status [24].

Currently, the state of the art (SOTA) for processing sensor data are architectures for sequential data modelling such as Res-CNN [25], LSTM fully-convolutional NN [26], inception-time [19], and ResNet [18]. The models were shown to work well on many

sequence learning tasks (e.g., [19,23], see [18] for a review). Additionally, these new methods have already been applied in the field of predictive maintenance. For example, ResNet has been used on wind turbine data [27] and bearing data [28] to predict faults. Res-CNN has been applied to motor data [29], and fully-convolutional LSTM used on aircraft engine data [30]. However, to our knowledge, no paper has compared all of the above methods on a single dataset.

In this paper, we describe an ML approach that was used to predict machine health with a large time horizon. Due to the nature of our application, we used a two-week horizon, but the approach can be generalized to other horizons as well. To process the sensor data, we compare all the SOTA architectures named above. Moreover, we also describe the results obtained using a simpler NN baseline model based on bidirectional GRU cells (BiGRU) [24]. Finally, we compared these NN approaches to a random-forest (RF) model, which is a very popular ML approach not based on NNs that performs well on a variety of tasks and does not require special processing for categorical variables [31,32]. Additionally, the inputs to the model are both continuous sensor data and categorical metadata, and we use K-Means clustering to incorporate prior knowledge of the distribution of the predicted variable into our model and generate the predicted variable, as we first described in [24].

We first show that this approach can provide superior predictions of machine health in comparison to a similar model that only incorporates sensor data, similarly to what we previously reported [24]. Moreover, we demonstrate the superiority of SOTA networks over the simpler BiGRU architecture as well as a non-NN approach (RF) for classifying industrial asset health.

2. Methods

2.1. Data

For a more detailed account, see [24]. Briefly, the data consisted of both sensor data collected approximately every 6 h and categorical metadata, over a period of approximately 2.5 years from 51 vibration sensors. The data were divided into training, validation, and test sets, so that approximately the first 2 years of data were used for training and the final 0.5 years of data was split between the validation and test datasets through stratified random shuffling based on the distribution of the predicted variable (defined below). Note that due to important data privacy concerns specified by the owner of the data, some aspects of the data were transformed to maintain data privacy.

2.2. The Predicted Variable

The predicted variable was determined based on the distribution of the sensor data of the training set, as well as practical specifications provided by the owner of the data and only very basic domain knowledge. Specifically, the data owner requested predictions of the systems' health status two weeks into the future. The full method is described in [24], but in brief, we integrated prior knowledge of the predicted values into the architecture of our model so that instead of predicting its value directly, we computed a set of clusters based on its distribution in the training set. We then labelled all our predicted variables based on the nearest cluster centroid calculated through the K-Means algorithm. Since our training data distribution resembled a bimodal distribution, suggesting 2 distinct types of behaviour (see Figure 1), we used the nearest cluster centroid of two possible clusters as the predicted variable.

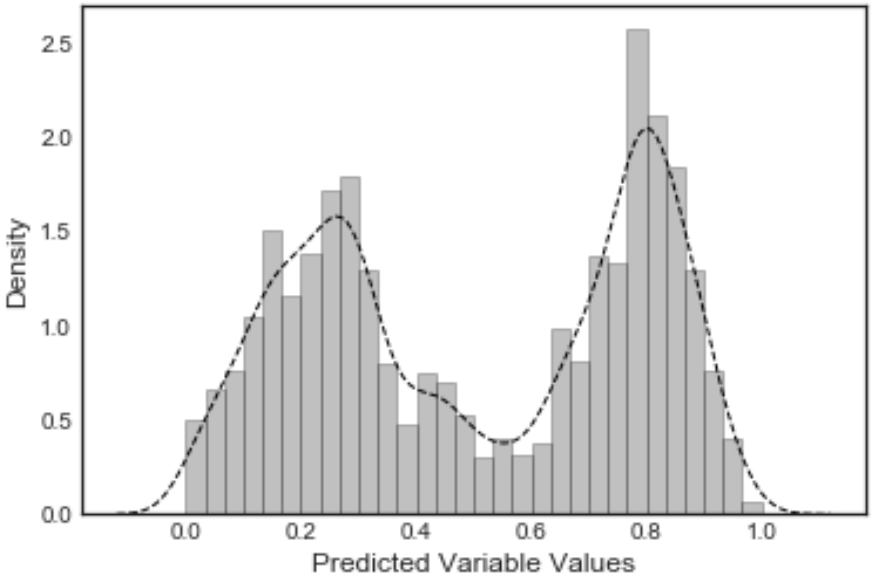

Figure 1. Distribution of the predicted variable in the training set. The dashed line represents a Gaussian kernel density estimation of the distribution (reproduced from [24]).

2.3. Modelling

In the current research, we tested several deep NN architectures for modelling the sensor data (i.e., sequence models). The first was a BiGRU, which we used as a baseline for comparison to different model architectures, and which we also used in a previous study [24]. We compared this relatively simple but popular architecture to several SOTA algorithms as well as a non-NN based approach (RF). First, we trained a transformer model that was slightly modified from [23], where it was used for English to German translation tasks so as to be suitable for sequential numerical data, mainly by replacing its embedding layers with fully connected layers. This stresses the notion that deep learning models that are developed to solve a certain task can often be rather straightforwardly adapted to solve a different task, even when the similarity between the tasks is not apparent. Additional SOTA algorithms that were used were Res-CNN [25], LSTM fully-convolutional NN [26], inception-time [19] and ResNet [18]. The hyperparameters of the models were selected by examining the loss function value on the validation set, and the models were tuned using the logistic loss-function, which is the most commonly used loss-function for binary-classification problems and is almost universally applied [33]:

$$L = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij}\log\left(p_{ij}\right), \quad (3)$$

where p is the predicted class and y is the true class label.

In addition, we were provided with metadata in the form of categorical variables that identify important aspects in the equipment, such as its specific type. To incorporate categorical variables in ML models, they are often transformed using one-hot encoding (OHE), where k new binary features are created for k different categories. However, as we stated in [24] when the cardinality of the features is high, OHE requires a large number of computational resources. Additionally, OHE treats the values of categorical variables as independent of each other and often ignores information about the relationships between them [34]. In order to circumvent these issues, we used the categorical metadata to learn entity embeddings, where each categorical variable is mapped to a fixed-size vector space, with parameters that are learned by the model (see [24,34,35]).

The overall modelling approach is presented in Figure 2. The embeddings were concatenated to the outputs of the sequence model component and fed to an FC layer with a rectified linear unit (ReLU) activation function. The outputs of this layer can then be fed to an additional FC layer with a Sigmoid activation function (i.e., the logistic function).

A constant learning rate of 0.001 was used with the Adam optimizer, and models were trained with early stopping, i.e., until we observed an error increase on the validation set [36].

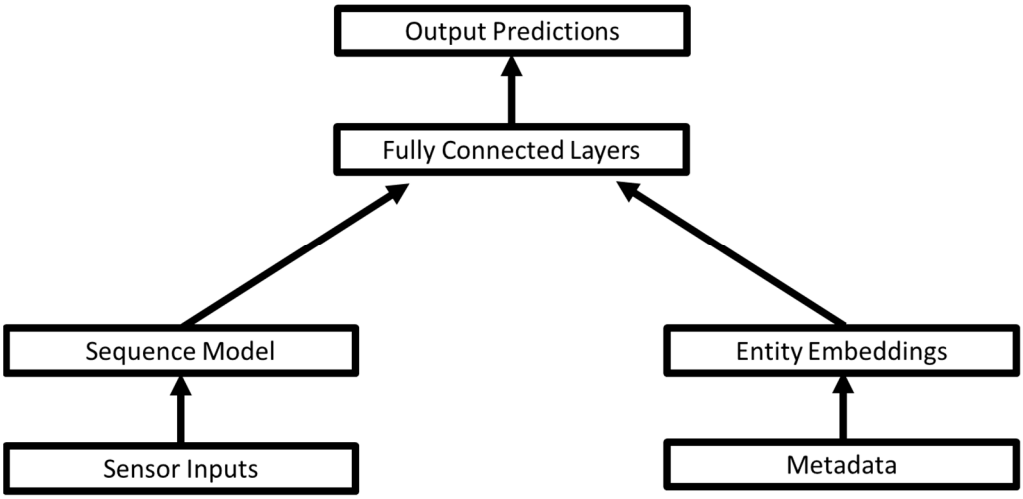

Figure 2. Overall model architecture.

The models were compared using two very popular classification metrics: the F1-score and the Matthews correlation coefficient (MCC) [37].

3. Results

All of the analyses were done using the Python programming language [38]. To assess the importance of the various model components, we performed an ablation study where we systematically removed the main components of our model and observed how it affected performance. In this respect, we compared our approach of using entity embeddings with the BiGRU model to the same model without the embedding inputs. Moreover, we tested a model where the penultimate FC layer was also removed (the first layer of the "fully connected layers" component in Figure 2). Finally, we compared the performance of various sequence models (sequence model component in Figure 2), including SOTA sequence models, as well as an RF model.

The performance of the experimental conditions is summarized in Table 1. The baseline BiGRU model generated an F1 score of 0.876 and an MCC score of 0.747. When entity embeddings were not included in the model, both F1 and MCC scores dropped. Similar results were obtained when the penultimate FC layer was removed, and the concatenated inputs from the BiGRU and embeddings were fed directly into the output layer of the model. Moreover, a model consisting only of the BiGRU component of the model achieved a similar performance, suggesting that the additional FC layer might not be needed when the additional metadata inputs are not included. When SOTA models were used instead of the BiGRU baseline, the model demonstrated an increased performance on both F1, $t(4) = 4.18$, $p < 0.01$, and MCC, $t(4) = 5.43$, $p < 0.01$. RF performed similarly to the BiGRU baseline on the F1 and MCC metrics. However, it also showed a strong bias towards predicting Class 1 (98.59% vs. 81.29% accuracy rates for Class 1 and Class 2, respectively). The F1 differences between SOTA algorithms and RF were marginally significant, $t(4) = 2.03$, $p = 0.056$, and statistically significant when considering only CNN based SOTA algorithms (e.g., Res-CNN, FCN, inception-time and ResNet), which performed best on our task, $t(3) = 4.93$, $p < 0.01$. MCC differences between CNN based SOTA algorithms and RF were marginally significant after correcting for multiple comparisons, $t(3) = 2.55$, $p = 0.04$.

Table 1. Model Classification Performance.

Model	Class 1 Accuracy	Class 2 Accuracy	Overall Accuracy	F1	MCC
BiGRU	85.05	89.6	87.33	0.876	0.747
BiGRU, no entity embed-dings	78.06	92.7	85.4	0.864	0.715
BiGRU, no penultimate FC	78.2	91.8	85.0	0.860	0.707
Only BiGRU	78.36	91.1	84.7	0.856	0.7
Transformer	90.90	85.78	90.26	0.880	0.768
Res-CNN	94.10	87.38	93.26	0.904	0.817
FCN	93.87	90.24	93.42	0.919	0.842
Inception-time	94.63	87.76	93.77	0.909	0.826
ResNet	95.68	85.7	94.43	0.902	0.818
Random-forests	98.59	81.29	89.47	0.890	0.811

4. Discussion

Although fully connected deep learning models have been used in MHMS for many years [39–42], the use of NN approaches that are specialized for sequence models is a relatively recent research trend [43,44]. This is somewhat surprising considering that most industrial data are sensor data, which is by nature sequential. Notably, several studies used recurrent NNs to estimate RUL [45–49] or performance degradation [7,50–52]. Other studies have applied CNN models after transforming sensor data to 2-dimensional, similar to image data that are typically used by CNNs, in order to classify machine faults [53,54] or RUL [55,56]. Yet another research direction has been to transform the sensor signals to the frequency domain before applying CNNs for machine fault diagnosis [21,57–59], while other studies straightforwardly applied CNNs for monitoring the health status of industrial assets using the raw sensor data as the input [60–64]. Importantly, none of the previous studies compared several SOTA sequence models for MHMS on a single dataset [44,65–68], and the current study was the first to apply them in this context. Such models are significantly deeper and computationally more complex than those that were used in most previous studies and were originally developed for applications unrelated to machine health monitoring (e.g., NLP [23]).

The MHMS described in this paper can incorporate SOTA models as well as combine sequential and non-sequential inputs to obtain more accurate predictions, as when using each input type in isolation. Its effectiveness was verified through an ablation study where the main components of the model were systematically removed or altered. Moreover, the proposed MHMS makes use of the predicted variable distribution to derive classes for prediction using unsupervised clustering (see [24]). Such class derivation is especially important in applications where the theoretical variable, e.g., asset health distribution, is not known directly. Our proposed algorithm can be used to derive a proxy of the theoretical variable using a different variable for which the distribution in the training data can be estimated. Moreover, we tested the various SOTA algorithms on our data. While such models, e.g., with a single or few LSTM or GRU layer(s), can work relatively well on industrial tasks, we found that using SOTA models resulted in increased performance on the metrics that we measured, especially CNN based models. What this suggests is that while industrial data might contain important unique features, e.g., features that are representative of industrial asset health might only be discoverable in these data, the SOTA models developed for seemingly unrelated data and tasks are nonetheless also transferrable to these data. This is likely because SOTA sequential models are highly proficient at learning general temporal dynamic behaviour and hence can also be applied here.

In conclusion, we have proposed an MHMS that can handle both numeric and categorical data, can be used in conjunction with SOTA NNs and can be used to predict the health

status of industrial assets even when a health status variable is not explicitly provided. Such a system can serve as an integral component of full-fledged predictive maintenance software systems to provide increased automation for asset health inspection.

Author Contributions: Conceptualization, I.A., A.K., M.C., and R.G.; methodology, I.A.; software, I.A.; validation, I.A., A.K., M.C., and R.G.; formal analysis, I.A., A.K., M.C., and R.G.; investigation, I.A., A.K., M.C., and R.G.; data curation, I.A., A.K., M.C., and R.G.; writing—original draft preparation, I.A.; writing—review and editing, I.A., A.K., D.P., M.C., and R.G.; project administration, I.A. and D.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ABB.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data due to privacy concerns pertaining to the data source.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Zhao, R.; Wang, D.; Yan, R.; Mao, K.; Shen, F.; Wang, J. Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks. *IEEE Trans. Ind. Electron.* **2017**, *65*, 1539–1548. [CrossRef]
2. Lund, D.; MacGillivray, C.; Turner, V.; Morales, M. *Worldwide and Regional Internet of Things (IoT) 2014–2020 Forecast: A Virtuous Circle of Proven Value and Demand*; International Data Corporation: Framingham, MA, USA, 2014.
3. Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S.X. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3137–3147. [CrossRef]
4. Yin, S.; Li, X.; Gao, H.; Kaynak, O. Data-Based Techniques Focused on Modern Industry: An Overview. *IEEE Trans. Ind. Electron.* **2015**, *62*, 657–667. [CrossRef]
5. Chen, Z.; Fang, H.; Chang, Y. Weighted data-driven fault detection and isolation: A subspace-based approach and algorithms. *IEEE Trans. Ind. Electron.* **2016**, *63*, 3290–3298. [CrossRef]
6. Jardine, A.K.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [CrossRef]
7. Zhao, R.; Wang, J.; Yan, R.; Mao, K. Machine health monitoring with LSTM networks. In Proceedings of the 10th International Conference on Sensing Technology (ICST), Nanjing, China, 11–13 November 2016; pp. 1–6.
8. Liu, Z.; Jia, Z.; Vong, C.M.; Bu, S.; Han, J.; Tang, X. Capturing high-discriminative fault features for electronics-rich analog system via deep learning. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1213–1226. [CrossRef]
9. He, M.; He, D. Deep Learning Based Approach for Bearing Fault Diagnosis. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3057–3065. [CrossRef]
10. Janssens, O.; Van De Walle, R.; Loccufier, M.; Van Hoecke, S. Deep Learning for Infrared Thermal Image Based Machine Health Monitoring. *IEEE/ASME Trans. Mechatronics* **2017**, *23*, 151–159. [CrossRef]
11. Jiang, G.; He, H.; Xie, P.; Tang, Y. Stacked Multilevel-Denoising Autoencoders: A New Representation Learning Approach for Wind Turbine Gearbox Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 2391–2402. [CrossRef]
12. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
13. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks* **1994**, *5*, 157–166. [CrossRef]
14. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In Proceedings of the NIPS Workshop on Deep Learning, Montreal, QC, Canada, 8–13 December 2014.
15. Cho, E.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the Conference for Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734.
16. Gallicchio, C.; Micheli, A. Deep echo state network (DeepESN): A brief survey. *arXiv* **2017**, arXiv:1712.04323.
17. Jaeger, H.; Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **2004**, *304*, 78–80. [CrossRef]
18. Fawaz, H.I.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.-A. Deep learning for time series classification: A review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963. [CrossRef]
19. Fawaz, H.I.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.-A.; Petitjean, F. InceptionTime: Finding AlexNet for time series classification. *Data Min. Knowl. Discov.* **2020**, *34*, 1–27. [CrossRef]

20. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
21. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
22. Wu, N.; Green, B.; Ben, X.; O'Banion, S. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv* **2020**, arXiv:2001.08317.
23. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
24. Amihai, I.; Chioua, M.; Gitzel, R.; Kotriwala, A.M.; Pareschi, D.; Sosale, G.; Subbiah, S. Modeling Machine Health Using Gated Recurrent Units with Entity Embeddings and K-Means Clustering. In Proceedings of the IEEE 16th International Conference on Industrial Informatics, Porto, Portugal, 18–20 July 2018; pp. 212–217.
25. Liu, L.; Chen, S.; Zhang, F.; Wu, F.X.; Pan, Y.; Wang, J. Deep convolutional neural network for automatically segmenting. *Neural Comput. Appl.* **2020**, *32*, 6545–6558. [CrossRef]
26. Karim, F.; Majumdar, S.; Darabi, H.; Chen, S. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* **2018**, *6*, 1662–1669. [CrossRef]
27. Stetco, A.A. Wind Turbine operational state prediction: Towards featureless, end-to-end predictive maintenance. In Proceedings of the International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 4422–4430.
28. Duan, J.S. A novel ResNet-based model structure and its applications in machine health monitoring. *J. Vib. Control* **2021**, *27*, 1036–1050. [CrossRef]
29. Liu, R.; Wang, F.; Yang, B.; Qin, S.J. Multiscale Kernel Based Residual Convolutional Neural Network for Motor Fault Diagnosis Under Nonstationary Conditions. *IEEE Trans. Ind. Informatics* **2020**, *16*, 3797–3806. [CrossRef]
30. Zhang, W. Aero-engine remaining useful life estimation based on 1-dimensional FCN-LSTM neural networks. In Proceedings of the 2019 IEEE Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019.
31. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
32. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning*; Zhang, C., Ma, Y., Eds.; Springer: Boston, MA, USA, 2012; pp. 157–175.
33. Painsky, A.; Wornell, G. on the universality of the logistic loss function. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 936–940.
34. Cheng, G.; Berkhahn, F. Entity embeddings of categorical variables. *arXiv* **2016**, arXiv:1604.06737.
35. de Brébisson, A.; Simon, É.; Auvolat, A.; Vincent, P.; Bengio, Y. Artificial neural networks applied to taxi destination prediction. *arXiv* **2015**, arXiv:1508.00021.
36. Prechelt, L. Early stopping-but when? In *Neural Networks: Tricks of the Trade*; Montavon, G., Orr, G.B., Müller, K.R., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 55–69.
37. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [CrossRef]
38. Oliphant, T.E. Python for scientific computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20. [CrossRef]
39. Li, B.; Chow, M.-Y.; Tipsuwan, Y.; Hung, J. Neural-network-based motor rolling bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2000**, *47*, 1060–1069. [CrossRef]
40. Samanta, B.; Al-Balushi, K. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Syst. Signal Process.* **2003**, *17*, 317–328. [CrossRef]
41. Aminian, M.; Aminian, F. Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor. *IEEE Trans. Circuits Syst. II Analog. Digit. Signal Process.* **2000**, *47*, 151–156. [CrossRef]
42. Su, H.; Chong, K.T. Induction machine condition monitoring using neural network modeling. *IEEE Trans. Ind. Electron.* **2007**, *54*, 241–249. [CrossRef]
43. Samir, K.; Takehisa, Y. A review on the application of deep learning in system health management. *Mech. Syst. Signal Process.* **2018**, *107*, 241–265.
44. Toh, G.; Park, J. Review of vibration-based structural health monitoring using deep learning. *Appl. Sci.* **2020**, *10*, 1680. [CrossRef]
45. Zheng, S.; Ristovski, K.; Farahat, A.; Gupta, C. Long short-term memory network for remaining useful life estimation. In Proceedings of the IEEE International Conference on Prognostics and Health Management (ICPHM), Dallas, TX, USA, 19–21 June 2017; pp. 88–95.
46. Yuan, M.; Wu, Y.; Li, L. Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. In Proceedings of the IEEE International Conference on Aircraft Utility Systems (AUS), Beijing, China, 10–12 October 2016; pp. 135–140.
47. Malhotra, P.; Ramakrishnan, A.; Anand, G.; Vig, L.; Agarwal, P.; Shroff, G. Multi-sensor prognostics using unsupervised health index based on LSTM Encoder-Decoder. *arXiv* **2016**, arXiv:1608.06154.
48. Chen, Y.; Peng, G.; Zhu, Z.; Li, S. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. *Appl. Soft Comput.* **2020**, *86*, 105919. [CrossRef]

49. Xia, T.; Song, Y.; Zheng, Y.; Pan, E.; Xi, L. An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation. *Comput. Ind.* **2020**, *115*, 103182. [CrossRef]
50. He, M.; Zhou, Y.; Li, Y.; Wu, G.; Tang, G. Long short-term memory network with multi-resolution singular value decomposition for prediction of bearing performance degradation. *Measurement* **2020**, *156*, 107582. [CrossRef]
51. Zhao, R.; Yan, R.; Wang, J.; Mao, K. Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors* **2017**, *17*, 273. [CrossRef] [PubMed]
52. Tao, Y.; Wang, X.; Sanches, R.V.; Yang, S.; Bai, Y. Spur gear fault diagnosis using a multilayer gated recurrent unit approach with vibration signal. *IEEE Access* **2019**, *7*, 56880–56889. [CrossRef]
53. Guo, L.; Gao, H.; Huang, H.; He, X.; Li, S. Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring. *Shock Vib.* **2016**, *2016*, 1–10. [CrossRef]
54. Janssens, O.; Slavkovikj, V.; Vervisch, B.; Stockman, K.; Loccufier, M.; Verstockt, S.; Van de Walle, R.; Van Hoecke, S. Convolutional Neural Network Based Fault Detection for Rotating Machinery. *J. Sound Vib.* **2016**, *377*, 331–345. [CrossRef]
55. Babu, G.S.; Zhao, P.; Li, X.L. Deep convolutional neural network based regression approach for estimation of remaining useful life. In Proceedings of the International Conference on Database Systems for Advanced Applications, Dallas, TX, USA, 16–19 April 2016; pp. 214–228.
56. Chen, Z.; Shang, L.; Zhou, M. A FP-CNN method for aircraft fault prognostics. In Proceedings of the 3rd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE), Dalian, China, 12–13 May 2018; pp. 571–579.
57. Wang, J.; Zhuang, J.; Duan, L.; Cheng, W. A multi-scale convolutional neural network for featureless fault diagnosis. In Proceedings of the 2016 International Symposium of Flexible Automation (ISFA), Cleveland, OH, USA, 1–3 August 2016; pp. 1–6.
58. Guennemann, N.; Pfeffer, J. Predicting defective engines using convolutional neural networks on temporal vibration signals. In Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications, Skopje, Macedonia, 22 September 2017; pp. 92–102.
59. de Oliveira, M.; Monteiro, A.; Vieira, F.J. A new structural health monitoring strategy based on PZT sensors and convolutional neural networks. *Sensors* **2018**, *18*, 2955. [CrossRef]
60. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5990–5998. [CrossRef]
61. Abdeljaber, O.; Avci, O.; Kiranyaz, S.; Gabbouj, M.; Inman, D.J. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J. Sound Vib.* **2017**, *388*, 154–170. [CrossRef]
62. Han, T.; Liu, C.; Yang, W.; Jiang, D. A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowl. Based Syst.* **2019**, *165*, 474–487. [CrossRef]
63. Dong, H.; Yang, L.; Li, H. Small fault diagnosis of front-end speed controlled wind generator based on deep learning. *WSEAS Trans. Circuits Syst.* **2016**, *15*, 64–72.
64. Lin, Y.; Nie, Z.-H.; Ma, H.-W. Structural Damage Detection with Automatic Feature-Extraction through Deep Learning. *Comput. Civ. Infrastruct. Eng.* **2017**, *32*, 1025–1046. [CrossRef]
65. 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]
66. Baur, M.; Albertelli, P.; Monno, M. a review of prognostics and health management of machine tools. *the international J. Adv. Manuf. Technol.* **2020**, *107*, 2843–2863. [CrossRef]
67. Alshorman, O.; Irfan, M.; Saad, N.; Zhen, D.; Haider, N.; Glowacz, A.; Alshorman, A. A Review of Artificial Intelligence Methods for Condition Monitoring and Fault Diagnosis of Rolling Element Bearings for Induction Motor. *Shock. Vib.* **2020**, *2020*, 1–20. [CrossRef]
68. Thoppil, N.M.; Vasu, V.; Rao, C.S.P. Deep learning algorithms for machinery health prognostics using time-series data: A review. *J. Vib. Eng. Technol.* **2021**. [CrossRef]