European Research Studies Journal
Volume XXIV, Special Issue 2, 2021

pp. 189-197

Management of Early Failure Detection of Production Process: The Case of the Clutch Shaft Alignment using LSTM Deep Learning Algorithm

Submitted 15/04/21, 1st revision 09/05/21, 2nd revision 30/05/21, accepted 12/06/21

Bartosz Przysucha[1], Tomasz Rymarczyk[2], Dariusz Wójcik[3], Marcin Kowalski[4], Ryszard Białek[5]

Abstract:

Purpose: *In this paper the neural networks model based on long short-term memory (LSTM) for early failure detection of the clutch shaft alignment system is developed. This issue is of particular importance when assessing the condition of the tool and predicting its durability, which are keys to the reliability and quality of the production process.*
Design/Methodology/Approach: *Based on real fault data of the measuring system, 500 clutch fault runs were simulated. Then, the time of failure was modelled with two neural networks, the conventional neural network of the ANN and the LSTM deep learning network. The study examined and compared the effectiveness and quality of both networks in the context of fault prediction.*
Practical Implications: *In vibroacoustic diagnostics, we often deal with machines operating in various conditions, which makes it difficult to diagnose them using standard methods. In such cases, spectral methods require analysis of frequency bands, which may contain other components in addition to information about the diagnosed parameter. The algorithm for predicting impending failure gives the possibility to monitor the current degradation status of the device. This makes it possible to streamline planning processes in the areas of inspection, preventive replacement of parts, warranty, service, or storage of spare parts.*
Findings: *The objective of the paper is to introduce an improved computational method for failure detection based on a deep learning algorithm. It was proven that LSTM networks are suitable for successfully solving this scope of tasks.*
Originality/Value: *The research showed that the proposed LSTM algorithm is more effective and accurate than conventional artificial neural networks (ANN) based on the multilayer perceptron model.*

Keywords: *LSTM deep learning, predictive maintenance, ANN artificial neutral network.*

JEL codes: *C45.*
Paper Type: *Research study.*

[1]*Lublin University of Technology, Lublin, Poland, b.przysucha@pollub.pl*
[2]*University of Economics and Innovation, Lublin, Poland, and Research & Development Centre Netrix S.A, Poland, tomasz.rymarczyk@netrix.com.pl*
[3]*Research & Development Centre Netrix S.A, Poland, dariusz.wojcik@netrix.com.pl*
[4]*University of Economics and Innovation, Lublin, Poland, marcin.kowalski@wsei.lublin.pl*
[5]*University of Economics and Innovation, Lublin, Poland, ryszard.bialek@wsei.lublin.pl*

1. Introduction

Nowadays, there are lots of attempts to use artificial intelligence methods in automatic diagnostics systems (Mushiri and Mbohwa, 2018). However, the number of existing variants of these methods is large, which makes it difficult to make the right choice. When choosing an artificial neural network, its type, architecture, parameters and learning algorithm should be specified. Unfortunately, the literature does not provide ready-made solutions for this choice, and even indicates the need for an individual approach to each of the problems (Noh and Hong, 2011).

Detection of defects at early stages of development allows to avoid a serious failure, the consequences of which may result in large economic losses or even a threat to human health and life (Pawlik, 2019; Rymarczyk *et al.*, 2019; Elsheikh *et al.*, 2018; Scalabrini Sampaio *et al.*, 2019). Unfortunately, the diagnostic methods and measures described in the literature in most cases do not show sufficient sensitivity to early stages of damage.

This article deals with the issue of maintenance and reliability of production systems, in particular the problem of early recognition of failures in the coupling shaft alignment system. The reliability theory is based on probability theory which investigates, detects, and forecasts random events. As regards the clutch shaft alignment – it determines probability of damages of individual elements or the whole the clutch shaft alignment system. In particular, the bearing or clutch may be responsible for the failure.

To ensure the reliability of machines working in continuous mode, where each downtime generates high costs, it is necessary to use continuous monitoring systems. One way of diagnosing machines is to assess the technical condition based on vibration measurements. In vibration diagnostics, vibration acceleration signals measured with piezoelectric transducers and/or vibration velocities recorded contact less with a laser vibrometer are used. The signals recorded in this way should be properly filtered and subjected to one of the applied methods of signal analysis, and then the measure sensitive to the damage should be used. Such an approach to the problem of diagnosis of technical condition has been the most frequently used so far.

This article presents an enhanced method of comprehensive machine diagnostics in industry. The described method allows a rational and effective maintenance of technological machines in industrial conditions. The aim of the research was to improve the way of failure detection in clutch shaft alignment. The concept of reliable and economic supervision of industrial processes was presented.

The novelty is the use of deep machine learning method to develop the failure detection algorithm. To increase the effectiveness of the method, the innovative Long Short-Term Memory (LSTM) method was used. The conducted research shown that the proper use of the LSTM neural network allows to gain better results than using ordinary neural networks. The utilitarian aim was to enable the withdrawal or repair of a worn part before its failure causes losses and additional costs.

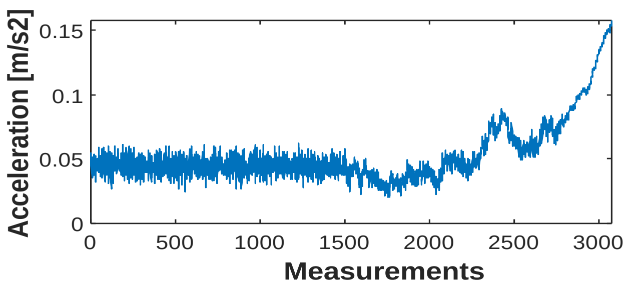
The article consists of 4 sections. The theoretical part contained in the Introduction is followed by the Materials and Methods section, which details the data, the concept of their use and used machine learning models. The results section describes the results obtained, comparing LSTM deep learning algorithm with standard artificial neural network (ANN). The final section contains conclusions regarding the usefulness of the method in practice.

2. Material and Methods

The research included analysis of the results of vibration acceleration measurements recorded on a bearing node in a propulsion engine connected by a shaft with a four-wheel clutch. The shaft with rolling bearings was in three bearing nodes. A fault was introduced in this system consisting in moving the middle bearing by 5 mm from the shaft axis. This fault caused the system to misalign. Vibration accelerations were measured by a piezoelectric sensor mounted on the bearing housing in horizontal and transverse directions to the shaft axis. The sampling frequency was 50kHz. As a result of shaft misalignment, overheating, displacement of the wheel coupling and a malfunction occurred. Damage consisting of overheating and displacement of the claw coupling due to misalignment of the system was recorded.

The analysed clutch system is one of the key elements of machine reliability. In the event of a faulty clutch shaft alignment, expensive machine components such as the engine or transmission may be damaged. Therefore, it is important to recognize the failure at an early stage, which will allow the machine to stop and replace e.g., a faulty bearing. 500 measurement sequences were used in the research. The number of sampling points in a single sequence ranged from 2410 to 3460. An example of the measuring sequence is shown in Figure 1.

Figure 1. *An example of the measuring sequence*



Source: *Own creation.*

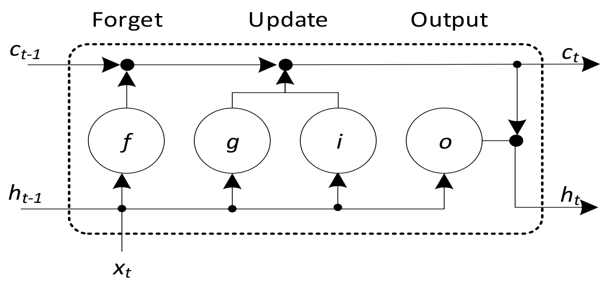
All measuring sequences were shortened to the shortest sequence length of 2410. They were shortened from the lower acceleration values. Then the sequences were divided into 10 element sets from which maximum, minimum, mean, and median values were selected. Because every technological machine has its own vibration characteristics of the clutch system, a classifier should be trained, which task is to select a predictor for a given device. This means that at the beginning several device classes should be identified, grouped due to similar vibration characteristics (range). Then, for each class of machines, predictors should be trained, whose task will be the status classification.

To solve the problem of early failure detection of the clutch shaft alignment, we propose the following workflow: 1-machine start, 2-recording of measurement sequences, 3-classification machine class, 4-selection of trained LSTM classificatory, 5-monitoring with the use of LSTM.

The initial classification of the machine is a relatively easy task, so it can be implemented using simple, well-known methods, e.g., shallow artificial neural networks (ANN), elastic net, least angle regression (LARS), etc. The initial classification of the machine is a relatively easy task, so it can be implemented using simple, known methods, e.g., ordinary neural network. The classification of machine states based on measurement sequences is more challenging. Therefore, the research proposed the use of the LSTM network for this purpose. Figure 2 presents a diagram of the LSTM network structure.

An important feature that distinguishes LSTM from other methods is the ability to learn long-term relationships between time steps in time series or given sequences. The LSTM layer consists of 2 states - hidden (initial) and cell state. The output data of the LSTM layer for a given time step t is contained in a hidden state at t. The status of cell contains information learned based on previous time steps. At each time stage, the LSTM layer adds or removes information from the cell's state. Information updates are controlled by means of gates.

Figure 2. *LSTM network structure*



Source: *Own creation.*

Individual gates control levels of cell states: f – reset (forget), (i) – input gate controls level of cell state update, (o) - output gate, g – cell candidate, it adds information to cell state. Weights W, the recurrent weights R and biases b can be described by formulas (1).

$$W = \begin{bmatrix} W_i \\ W_f \\ W_g \\ W_o \end{bmatrix}, \ R = \begin{bmatrix} R_i \\ R_f \\ R_g \\ R_o \end{bmatrix}, \ b = \begin{bmatrix} b_i \\ b_f \\ b_g \\ b_o \end{bmatrix} \tag{1}$$

where i, f, o, g denotes the input, forget and output gates and cell candidate, respectively. The cell state in given time step t is descripted by $\boldsymbol{c}_t = f_t \odot \boldsymbol{c}_{t-1} + i_t \odot g_t$, where $\odot$ denotes element-wise multiplication of vectors. The hidden state at time step t can be described as $\boldsymbol{h}_t = o_t \odot \sigma_c(\boldsymbol{c}_t)$ where σ_c is the state activation function.

The equations (2-5) describe the components of LSTM layer in time step t.

$$i_t = \sigma_g(W_i \boldsymbol{x}_t + R_i \boldsymbol{h}_{t-1} + b_i) \tag{2}$$
$$f_t = \sigma_g(W_f \boldsymbol{x}_t + R_f \boldsymbol{h}_{t-1} + b_f) \tag{3}$$
$$g_t = \sigma_c(W_g \boldsymbol{x}_t + R_g \boldsymbol{h}_{t-1} + b_g) \tag{4}$$
$$o_t = \sigma_g(W_o \boldsymbol{x}_t + R_o \boldsymbol{h}_{t-1} + b_o) \tag{5}$$

In the above formulas, σ_g means gate activation function. In the LSTM network layers the sigmoidal activation functions was used. It can be expressed by $\sigma(x) = (1 + e^{-x})^{-1}$. LSTM network with 9 layers was used (Table 1). The first layer in LSTM model is Sequence Input. The task of the sequence input layer is to enter sequential data into the network. The next layer is LSTM. The presented model uses the LSTM bidirectional layer. Bidirectional LSTM layer learns long-term relationships between time series time steps or sequence data in both directions (forward with feedback).

These relationships are important when there is a need for the network to learn from full time series at each time step. To prevent overfitting the neural network, 2 layers of dropout were used. Dropout layers randomly sets the values of some input elements to zero. For this purpose, it uses random functions with a given probability.

The seventh layer of the LSTM network is fully connected layer. This layer multiplies the numerical input values by the weight matrix and adds vector of biases. In deep networks, one or more fully connected layers are introduced after convolution and down sampling layers. If the entry to the fully connected layer is a sequence (for example LSTM network), then the fully connected layer operates individually at each time step. If the output of the layer placed before the fully connected layer is e.g., an array A1 of the size X by Y by Z, then the fully connected layer output is an array A2 of the size X' (output size) by Y by Z. At the time step t, the appropriate input of A2 is $\mathrm{WA}_t + \mathrm{b}$, where A_t is the time step t of A and b is the bias. In these studies, Glorot and Yoshua initializer was the initiating algorithm for the weights of this layer (Glorot and Yoshua, 2010).

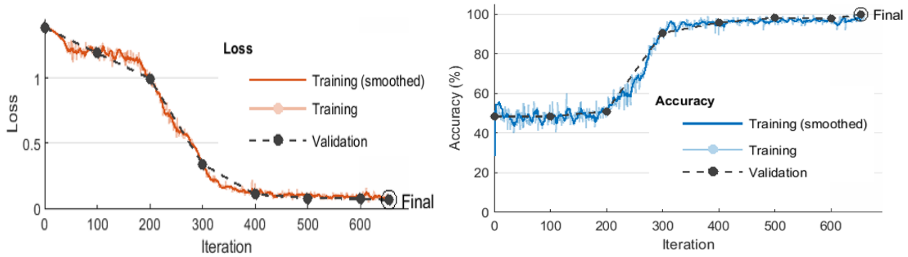
Table 1. *Layers of LSTM neural network*

#	Layer	Description
1	Sequence Input	Sequence input with 1 dimension
2	LSTM	LSTM with 100 hidden units
3	Dropout	10% dropout
4	LSTM	LSTM with 100 hidden units
5	Dropout	10% dropout
6	LSTM	LSTM with 100 hidden units
7	Fully Connected	fully connected layer with 4 neurons
8	Soft-max	Soft-max activation function
9	Classification Output	cross entropy function for 4 mutually exclusive classes

Source: *Own creation.*

The next layer is the soft-max. This type of layer is typical for deep classification neural networks. The soft-max layer is always preceded by a fully connected layer. Formula $y_r(x) = e^{a_r(x)}/\sum_{j=1}^{k} e^{a_j(x)}$ shows the soft-max activation function, where $0 \leq y_r \leq 1; \sum_{j=1}^{k} y_j = 1$. The last layer computes the cross-entropy loss for classification problem with mutually exclusive classes. To ensure learning sets of a sufficiently large number, aggregation of measurement data with similar characteristics was made.

Figure 3. *LSTM network learning waveforms based on (a) Loss and (b) Accuracy.*



Source: *Own creation.*

The criterion for network quality was cross entropy and accuracy. Accuracy is the percentage of correctly classified observations for all cases. Cross entropy loss between network predictions and target values is defined as $Loss = -\sum_{i=1}^{M} T_i log(X_i)/N$, where N – number of observations, M – number of responses, T_i – patterns, X_i – network outputs. Adaptive moment estimation (ADAM) training algorithm was used in the LSTM network. Testing the LSTM network on a set of validation cases showed Accuracy = 0.999 and Loss = 0.0796. Figure 3 presents LSTM network learning waveforms based on Accuracy and Loss. Sigmoid shapes of both curves show no overfitting (no fluctuation).

To fully assess the quality of the LSTM classifier, a shallow ANN neural network was trained based on the same data. A multi-layer classification perceptron with one hidden layer of 1-20-4 structure was used. The learning process was carried out by the Levenberg-Marquardt algorithm. The results obtained were worse than in the case of using the LSTM network.

3. Results

Table 2 presents a comparison of the classification results for the 2 types of neural networks tested. The table contains Loss values for ANN and LSTM networks divided into 4 aggregation functions (max, min, mean and median) for measuring sequences of length 10. Experiments were also carried out for functions aggregating measuring intervals longer than 10, i.e., 20, 30, ... 100, 125, 150, 175 and 200. In total, there were as many as 14 ranges tested. In all cases, better results were obtained using LSTM.

Table 2. *Cross entropy (Loss) comparison*

Aggregation function	ANN	LSTM
Maximum	0.0678	0.0796
Minimum	0.0615	0.0827
Mean	0.0697	0.0904
Median	0.0547	0.0828

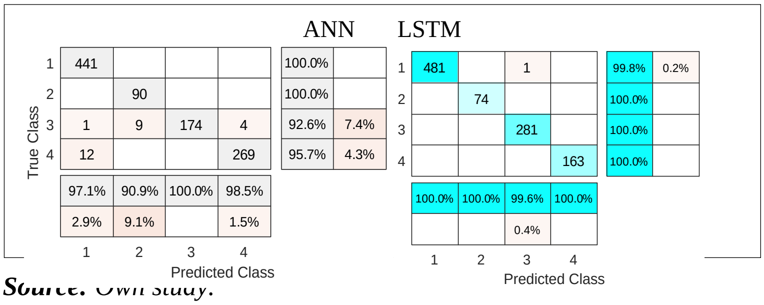
Source: *Own creation.*

Figure 4 shows a comparison of the ANN and LSTM classification results using a confusion matrix. The advantage of deep learning over the standard machine learning technique is very pronounced in the presented work.

4. Conclusions

The concept of early failure detection in the clutch-shaft system is of great potential for improving the reliability of machines. There is a huge number of technological machines in the world equipped with this type of systems. Many of them are very expensive both in terms of purchase price and maintenance costs. The industrial revolution manifest itself with the implementation of technologies such as 5G, internet of things or Industry 4.0, drives the intelligence of machines.

Figure 4. *A comparison of the ANN and LSTM*



Source: *Own study.*

The presented concept of early detection of failures through ongoing monitoring of the state of the machine makes it possible to prevent situations in which a failure generates disproportionately high costs and other risks. The presented concept of early detection of failures through monitoring of the state of the machine makes it possible to prevent situations in which a failure generates disproportionately high costs and other risks. The conducted research proves that deep LSTM networks, due to their greater complexity and thus greater possibilities of analyzing and processing sequential information, enable the creation of more effective algorithms than using ANN.

It can be assumed that the advantage of LSTM over ANN will be the greater the more complex the problem will be. Therefore, the potential is created to create more sophisticated predictive and classification models based on multi-source data. This research direction will be explored in the future.

References:

Elsheikh, A., Yacout, S., Ouali, M.S., Shaban, Y. 2018. Failure time prediction using adaptive logical analysis of survival curves and multiple machining signals. Journal of Intelligent Manufacturing, 31, 1-13, doi: 10.1007/s10845-018-1453-4.

Gradzki, R., Kulesza, Z., Bartoszewicz, B. 2019. Method of shaft crack detection based on squared gain of vibration amplitude, Nonlinear Dynamics: An International Journal of Nonlinear Dynamics and Chaos in Engineering Systems, 98(1), 671-690. doi.org/10.1007/s11071-019-05221-0.

Homišin, J., Grega, R., Kaššay, P., Fedorko, G., Molnár, V. 2019. Removal of systematic failure of belt conveyor drive by reducing vibrations. Engineering Failure Analysis, 99, 192-202. doi.org/10.1016/j.engfailanal.2019.02.028.

Hossain, M.S., Ong, Z.Ch., Ismail, Z., Noroozi, S., Khoo, S.Y. 2017. Artificial neural networks for vibration based inverse parametric identifications, A review. Applied Soft Computing, 52, 203-219. doi.org/ 10.1016/j.asoc.2016.12.014.

Kim, D., Lin, Y., Lee, S., Kang, B., Han, S. 2018. A Hybrid Failure Diagnosis and Prediction using Natural Language-based Process Map and Rule-based Expert System. International Journal of Computers, Communications & Control, 13(2), 175-191. doi.org/10.15837/ijccc.2018.2.3189.

Mushiri, T., Mbohwa, Ch. 2018. Fuzzy Logic Dynamics and Machine Prediction for Failure Analysis, Hershey PA. Engineering Science Reference (eBook). doi.org/10.4018/978-1-5225-3244-6.

Noh, M.S., Hong, D.S. 2011. Implementation of remote monitoring system for prediction of tool wear and failure using ART2. Journal of Central South University of Technology: Science & Technology of Mining and Metallurgy, 18(1), 177-183. doi.org/10.1007/s11771-011-0677-7.

Pawlik, P. 2019. Single-number statistical parameters in the assessment of the technical condition of machines operating under variable load. Maintenance and Reliability, 21(1), 164-169. http://dx.doi.org/10.17531/ein.2019.1.19.

Rymarczyk, T., Przysucha, B., Pawlik, P. 2020 Analysis of multi-source data for monitoring and control of intelligent technological systems. Przegląd Elektrotechniczny, R. 96, 9, 95-98.

Scalabrini Sampaio, G., Vallim Filho, A., Santos da Silva, L., Augusto da Silva, L. 2019. Prediction of Motor Failure Time Using an Artificial Neural Network. Sensors, 19(19), 4342.

Uma Maheswari, R., Umamaheswari, R. 2018. Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train-A contemporary survey. Mechanical Systems and Signal Processing, 85, 296-311. doi.org/ 10.1016/j.ymssp.2016.07.046.