Application of convolutional neural networks to rotary machines state recognition

E P Kornaeva¹,³ and A V Kornaev²

¹ Department of Information Systems and Digital Technologies, Orel State University named after I.S. Turgenev, 95 Komsomolskaya Street, Orel 302026, Russia
² Department of Mechatronics, Mechanics and Robotics, Orel State University named after I.S. Turgenev, 95 Komsomolskaya Street, Orel 302026, Russia
³ E-mail: lenoks_box@mail.ru

Abstract. The paper deals with the application of convolutional neural networks to state recognition of rotary machines with fluid-friction bearings. The training and testing were based on the data from the multisensory measurement system. The influence of minor defects such as tightening of the fastening bolts as well as various cases of lubricant supply to the state of the rotor system was investigated. At the same time, the changes introduced had an insignificant effect on the measurement results, which made nontrivial the identification of states. The preprocessing of the measurement data included normalization and equality of means hypothesis testing to identify abnormal data. The developed diagnostic system testing demonstrated high accuracy compared to the accuracy of experts.

1. Introduction

The applications of machine learning methods to the diagnosis of the rotary machine have several stages [1]. The first stage is based on logistic regression and support vector machines implemented by feed-forward artificial neural networks. The second stage is based on deep learning and multi-layers neural networks, primarily convolutional neural networks (CNNs) [1-5]. The first stage is considered historical, the second is the most common at the present time, and the third is promising [1].

Artificial intelligence diagnostic systems usually include three main blocks [1]: the data acquisition block; the block for preliminary analysis and feature extraction; the state recognition block. The data acquisition block is based on measurements of vibration, temperature, acoustic and other data on the condition of a machine. Many researchers agree that the preliminary processing of measurement data can significantly increase the accuracy of diagnosis [1-4]. The preliminary analysis can be performed in the time domain, frequency, or time-frequency domain [1, 6]. Subsequent data processing can be associated with the selection of a significant part of the data by filtering [1, 7]. Another approach is based on the selection of subsets from a set of data that have a positive effect on the accuracy of a diagnosis system [1, 8]. The recognition algorithm can be based on an expert system using classic or fuzzy logical rules, on a classifier using logistic regression, support vector machines, Bayesian classifiers, as well as using models of Markov’s processes, decision trees, and other methods [1]. However, in most cases, recognition algorithms are implemented using artificial neural networks. The most common type of network is CNN. The data preprocessing can also be carried out using artificial neural networks. The authors of [3] proposed an adversarial neural network architecture that allows
processing noisy data of limited size. In such cases, it is possible to develop complex architectures of neural networks with separate sections that solve separate problems.

2. Materials and methods

The training and testing data were obtained during experiments on the test rig, presented in figure 1. The measuring system produces up to 8000 measurements per second. The measurement system includes 2 proximity probes (displacement sensors of the rotor in the bearing), 3 accelerometers located on the surface of the test rig, a microphone, a fluid pressure sensor in the bearing, and a lubricant film contact resistance sensor.

![Figure 1](image)

**Figure 1.** The rotor machine test rig with the fluid-film bearing (left) and its information and measurement system (right).

After the experiment, the measurement results are randomly fragmented and stored in the matrices of input and output values.

The first stage of data preprocessing is normalization [5]:

\[ \bar{X} = \frac{X - \min(X)}{\max(X) - \min(X)} \]  

where \( X = (x^{(1)}, ..., x^{(m)}) \) is a matrix of measurement results with \( m \) element, \( \max(X), \min(X) \) are maximal and minimal values of the matrix.

The next stage of data preprocessing is associated with testing the hypotheses of the equality of means, which helps to estimate the homogeneity of the data samples (the results of parallel tests). For large samples with unknown variances, the statistical criteria have the following form:

\[ z_{ij} = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{s_{\bar{x}_i}^2/n_i + s_{\bar{x}_j}^2/n_j}} \sim N(0; 1) \]  

where \( \bar{x}_i, \bar{x}_j \) are matrices of measurements on the parallel tests with indexes \( i \) and \( j \), \( s_{\bar{x}_i}^2, s_{\bar{x}_j}^2 \) are corrected sample variances.

For each pair of \( i \) and \( j \) parallel tests, the hypotheses of the mathematical expectations equality were tested. An alternative hypothesis is that the mathematical expectations are not equal. The critical domain was defined as two-sided. Figure 2 demonstrates the results of hypothesis testing in the form of a pairwise comparison matrix for the 10 parallel tests of 6 experiments. Decisions were made at the significance levels of \( \alpha = 0.05 \) and \( \alpha = 0.01 \) respectively. The elements of the matrix are the ratios of \( |z_{ij}| \) to the critical one, i.e. if the value of this ratio is less than 1, then the results of parallel experiments are recognized as homogeneous. Hypotheses about the equality of variances were also tested in pairs for all possible combinations of the results of parallel tests. Fisher's criterion was used for the two-sided critical region. The criterion value was defined as the ratio of corrected sample variances. As a result of such verification, it is possible to discard experiments that have atypical values.
Distinctive of CNNs are the input data in the form of one- two- or three-dimensional images and the use of convolutional layers instead of fully connected ones. In the case of using two-dimensional images (color or grayscale) the convolutional layers implement the cross-correlation operation of the matrix $X$ with the kernel $F$ ($S = X \ast F$) which is similar in structure to the mathematical convolution operation [1]. Components of the resulting matrix for the case of two-dimensional matrices $X$ and $K$ can be calculated as follows:

$$s_{i,j} = \sum_q \sum_r (x_{i+q,j+r} f_{q,r}).$$

(3)

where $s_{i,j}$, $x_{i+q,j+r}$, $f_{q,r}$ are the components of the resulting, the input and the kernel matrices relatively.

The CNNs have the following additional operations. The stride allows performing sliding with a shift of one or more steps. It helps to reduce the size of the resulting matrix [1, 4]. Another option for reducing the size of the resulting matrix is delated convolution [1, 4]. The padding operation expands the boundaries of the input matrix and pads it with some values, usually zeros (zero padding). This makes it possible to process the input matrix more completely by the convolution kernel. Usually, convolution operations are performed not on a single kernel, but on a set of kernels of the same size. The different kernels adapt to the various features of the object under study. Each convolution operation is followed by activation. The most frequently used activation is a rectified linear unit (“relu”) [1]. Another operation to reduce the size as computations are performed in convolutional networks is the pooling operation. Usually, it is used after convolutions. Fragments of the input matrix are replaced with a number of the maximum or average value in the selected fragment of the matrix.

The proposed CNN architecture includes normalization and dropout operations to reduce the effect of overfitting and increase the stability and accuracy of neural networks [1]. The architecture of the proposed CNN is presented in figure 3. CNN was designed using the MATLAB “Deep network designer” application [9].

The input of the CNN is an image in the form of a two-dimensional matrix $X$ (three-dimensional in the case of color images). The input layer is followed by blocks of convolutional and activation layers. A dropout layer randomly sets input elements to zero with a given probability. The last few layers of the network are usually fully connected. The output layer contains the number of neurons equal to the number of classes for recognition. An activation function of the output layer is the “softmax” [1].
The values on the output neurons are denoted as hypothesis $H$, the elements of the matrix of which should be compared during training, validation, and testing with the true values of $Y$.

The loss function with unknown values of the kernels $F^{(p)}$ and weights $\Theta^{(k)}$ is minimized while training and validation [1]:

\[
L(F^{(p)}, \Theta^{(k)}) = - \sum_{i=1}^{m} \sum_{j=1}^{n_l} y_j^{(i)} \ln(h_j^{(i)}) \Rightarrow \min.
\]

where $m$ is the number of samples, that is, the number of elements in matrices $X$, $Y$, $n_l$ is the number of output neurons.

The training process is implemented by the gradient descent method or one of its modifications [1]. All the data set is divided into training, validation and test subsets. The number of hyperparameters for deep CNNs is much larger than for the fully connected ones.

3. Results and discussion

Two series of experiments were carried out to recognize faults and different conditions for hydrodynamic lubrication of the bearing.

The plan of the first series of experiments included the investigation of the normal state of the rotary machine and conditions with 5 different defects such as rotor imbalance and loose bolts of fasteners. The rotor speed was about 2600 rpm while the tests and the measurement system operated with a frequency of 1000 measurements per second. The plan included 6 experiments with 5 parallel tests each. All the tests were performed in a random order to reduce the drift effect. The results of the experiments were recorded in files for subsequent analysis and data preparation for training and testing the CNN.

The network testing results are presented in table 1. The recognition accuracy was 82.7%. Additionally, a survey of 4 experts was conducted in order to identify defects using the same experimental data. The accuracy of experts in determining 2 classes of states: normal and faulty was 67%.

| Fault diagnosis (6 classes) | Precision, % | Recall, % | Fscore  |
|----------------------------|--------------|-----------|---------|
|                            | 83.4         | 82        | 82.7    |
| Operating conditions (4 classes) | 93           | 92        | 92.5    |

The second series of experiments was devoted to the problem of recognizing the conditions of hydrodynamic lubrication by a temperature inhomogeneous fluid. The bearing was lubricated according to one of 4 cases of lubricant supply. Indirect measurements were used, namely, data from rotor vibration displacement sensors. The results of measurements in form or rotor trajectories are presented in figure 4. The accuracy of the CNN is presented in table 1. The accuracy of experts in determining 4 cases of lubricant supply was 53%.
Figure 4. The rotor vibrations trajectories (in microns) obtained from the second series of experiments for the 4 cases of hot and cold lubricant supply to the fluid-film bearing.

In both series of the experiments the data statistical preprocessing using equation (3) helped to increase the accuracy of the CNNs by 3-5%.

4. Conclusions
The application of convolutional neural networks to rotary machines state recognition makes it possible to achieve diagnostic accuracy that exceeds the accuracy of experts. One of the identified advantages of using images as input to neural networks is the ability to use color filters to label data with some additional information. For instance, with the information obtained using statistical analysis. Preliminary statistical analysis of the measurement data can improve the recognition accuracy by 3-5%.

Acknowledgments
The research was supported by the Russian Science Foundation, grant number No16-19-00186 (experiment with the rotor system fault diagnosis) and by the Grant of the President of Russian Federation, grant number MK-3394.2019.8 (experiment with the rotor system lubrication type recognition). The authors gratefully acknowledge this support. Authors also express graduate to Alex Fetisov and Alexey Rodichev for the help with the test rig design and manufacturing.

References
[1] Lei Y, Yang B, Jiang X, Jia F, Li N and Nandi A K 2020 Applications of machine learning to machine fault diagnosis: A review and roadmap Mechanical Systems and Signal Processing 138 39
[2] Liu R, Yang B, Zio E and Chen X 2018 Artificial intelligence for fault diagnosis of rotating machinery: A review Mechanical Systems and Signal Processing 108 33-47
[3] Zou L, Li Y and Xu F 2020 An adversarial denoising convolutional neural network for fault diagnosis of rotating machinery under noisy environment and limited sample size case Neurocomputing 407 105-20
[4] Tang S, Yuan S and Zhu Y 2020 Data Preprocessing Techniques in Convolutional Neural Network based on Fault Diagnosis towards Rotating Machinery IEEE Acess DOI: 10.1109/ACCESS.2020.3012182
[5] Goodfellow I, Bengio Yo and Courville A 2016 Deep Learning (Cambridge: MIT Press)
[6] Lei Y, He Z, Zi Y and Hu Q 2007 Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas Mech. Syst. Signal Process 21 2280-94
[7] Kira K and Rendell L A 1992 A practical approach to feature selection International Workshop on Machine Learning (Aberdeen) 249-56
[8] Liu H and Setiono R 1996 Feature selection and classification - a probabilistic wrapper approach International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (Fukuoka) 419-24
[9] Mathworks help center https://www.mathworks.com/help/matlab/help-and-support.html