Aalto University

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Miettinen, Jesse; Tiainen, Tuomas; Viitala, Risto; Hiekkanen, Kari; Viitala, Raine

Bidirectional LSTM-Based Soft Sensor for Rotor Displacement Trajectory Estimation

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2021.3136155

Published: 16/12/2021

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Miettinen, J., Tiainen, T., Viitala, R., Hiekkanen, K., & Viitala, R. (2021). Bidirectional LSTM-Based Soft Sensor for Rotor Displacement Trajectory Estimation. *IEEE Access*, *9*, 167556-167569. Article 9654210. https://doi.org/10.1109/ACCESS.2021.3136155

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 13, 2021, accepted December 8, 2021, date of publication December 16, 2021,
date of current version December 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3136155

Bidirectional LSTM-Based Soft Sensor for Rotor Displacement Trajectory Estimation

JESSE MIETTINEN[1], TUOMAS TIAINEN[1], RISTO VIITALA[1], KARI HIEKKANEN[2], AND RAINE VIITALA[1]

[1]Department of Mechanical Engineering, Aalto University, Espoo, Finland
[2]Department of Computer Science, Aalto University, Espoo, Finland

Corresponding author: Jesse Miettinen (jesse.miettinen@aalto.fi)

This work was supported in part by the Academy of Finland as part of the AI-ROT Research Project under Grant 335717, and in part by the Business Finland as part of the Reboot IoT Project under Grant 4356/31/2019.

ABSTRACT Constant rotor system monitoring enables timely control and maintenance actions that decrease the likelihood of severe malfunctions and end product quality deficits. Soft sensors represent a promising branch of solutions enhancing rotor system monitoring. A soft sensor can substitute a malfunctioning physical sensor and provide estimates of a quantity that is difficult to measure. This research demonstrates a soft sensor based on bidirectional long short-term memory (LSTM), and a training procedure for rotor system monitoring at high sampling frequency and varied operating conditions. This study adopts a large rotor and bearing vibration dataset. The soft sensor accurately estimates lateral displacement trajectories of the rotor from the bearing reaction forces over a large range of constant rotating speeds and constant support stiffnesses. The mean absolute error (MAE) of the LSTM-based soft sensor is 0.0063 mm over the test trajectories in the complete operating condition space. The soft sensor performance is shown to decrease significantly to a MAE of 0.0442 mm, if the training dataset is limited in the rotating speed range.

INDEX TERMS Long short-term memory (LSTM), recurrent neural network (RNN), rotor system, soft sensor, vibration.

I. INTRODUCTION

Measuring rotor vibration poses a common concern across numerous industries since excessive rotor vibration reduces machine lifetime and end product quality. For example, in paper production the paper thickness, gloss and basis weight are sensitive to vibration; thus the vibration tolerances are defined in micrometers. Furthermore, the rotor system components such as gears and bearings may wear down and fail before their expected end of lifetime due to rotor vibration. In order to control the vibration and to prevent such quality deficits and malfunctions, the machine vibration needs to be constantly monitored. Unfortunately, these vital vibration measurements may occasionally be prevented by a number of reasons. For example, machine geometry, processed materials or a hazardous environment might prevent the measurement of vibration. Furthermore, the financial burden and the high likelihood of sensor malfunctions associated with large sensor fleets might hinder rotor system monitoring in some large factories such as paper or steel mills.

Soft sensors appear a promising solution for the aforementioned problems related to rotor vibration monitoring. Soft sensors, or synonymously virtual sensors, are computational models estimating a physical quantity from other related information, such as secondary measurements. Soft sensors are often categorised either as model-based or data-driven soft sensors [1]. Model-based soft sensors are designed on the prior knowledge of the system. Some well-known techniques for designing model-based soft sensors include Kalman filters, [2], extended Kalman filters [3] and state observers [4]. Soft sensors based on these techniques have been shown to estimate for example torsional vibration and rotating speed of a thruster [5], eccentricity of steel mill rotors [2] and bioprocess states [6]. However, designing an accurate rotor system model for a soft sensor may be a difficult task due to highly non-linear interactions between the system components. These difficulties can be solved with some simplifying assumptions such as linearisation of the system dynamics.

The associate editor coordinating the review of this manuscript and approving it for publication was Li Minn Ang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

For example, Kalman filter for estimating lateral displacements of a drill collar employed a lumped model where the drill components were joined by linear torsional springs [7]. The disadvantage of these simplifying assumptions is that they may lead to a reality gap between the model and the real rotor system, which might hinder the model-based soft sensor performance on real data.

Data-driven soft sensors are fitted on historical data measured from the system [1]. They require little prior knowledge and no simplifying assumptions of the underlying system dynamics. Traditional data-driven soft sensors are based on techniques such as self-organising maps (SOM) [8], partial least squares (PLS) [9], multilayer perceptrons (MLP) [10] and support vector machines (SVM) [11]. Typically, the disadvantage of these machine learning based traditional soft sensors is that they require complex feature extraction techniques and manual feature selection. Although, SVM-based solutions estimating rotor displacement directly from the control coils of a hybrid magnetic bearing have been developed recently [12], [13], there still are few data-driven soft sensors that have been shown to generalise over a large range of operating conditions. For example, varied rotating speed or support structure stiffness are such operating conditions for rotor systems. Most likely, the learning capacity of these traditional data-driven soft sensors limits their applicability to problems requiring accuracy over a large range of many operating conditions.

Deep learning has been found to outperform other machine learning models in numerous benchmark tasks such as image classification [14], natural language processing [15] and reinforcement learning [16]. Unsurprisingly, these methods have also attracted the attention of soft sensor researchers. Numerous studies have indicated the applicability of different deep learning based soft sensors to many application areas [17]. For example, deep learning based soft sensors were adopted in monitoring industrial processes such as penicillin fermentation [18], polyethylene production [19] and hydrocracking [20]. However, deep learning based soft sensors for rotor systems have acquired less attention [17]. Although a few deep learning based soft sensors that estimate deformation of an air pre-heater have been proposed based on stacked autoencoder (SAE) [21], [22] and deep belief network [23], deep learning based soft sensors for rotor vibration are relatively scarce. Furthermore, most soft sensor applications function on data with low sampling rates, such as one sample per hour or day. Such soft sensors would likely be unsuitable for monitoring rotor vibration, such as rotor displacement trajectories, at the typical sampling rates of multiple times per second. A soft sensor for rotor vibration needs to learn dynamic changes in the system at high frequency instead of relatively static non-linear relations between numerous process variables.

This study demonstrates a data-driven soft sensor based on long short-term memory (LSTM) and a training procedure for rotor displacement trajectories. LSTM [24] is a recurrent neural network (RNN) capable of learning relations in sequential data over long time lags. LSTM can learn such relations due to the inbuilt memory that it learns to control at each time instant with gating mechanisms. This property has proven useful in previously developed soft sensors for a chemical process [25] and rotor speed [26]. This study employs the LSTM memory and gating mechanisms to learn the relation between the industrially available bearing vibration signals and rotor lateral displacement trajectories over a range of operating conditions. LSTM can encode this non-linear dynamic relation into its memory as latent hidden states with implicit state transition and observer functions parameterised by the gating functions. The rotor displacement values are accurately estimated at each time instant from the corresponding bearing vibration values at 2 kHz sampling frequency. The applicability of the LSTM-based soft sensor is demonstrated with a large and recently published rotor vibration dataset [27] including real rotor displacement trajectories and corresponding bearing vibration data under varied operating conditions. These conditions include rotating speed and horizontal support stiffness of the rotor.

The soft sensor performance is examined in three case studies. In the first one, the model is compared against other regression algorithms without in-built memory and shown to be superior in estimation accuracy. The second case study demonstrates LSTM-based soft sensor performance over a large range of operating conditions. The third case study highlights the importance of adopting the entire operating condition space in the training procedure. The results imply that the proposed LSTM soft sensor is accurate and applicable to rotor (lateral displacement) monitoring. Moreover, such a soft sensor can be useful for example in paper production. In paper production, the vibrational displacement is both, directly linked to the end product quality and difficult to measure due to the ongoing production process.

The contributions of this paper are the following:

- LSTM-based soft sensor for rotor displacement trajectory estimation under different operating conditions is demonstrated.
- A soft sensor training procedure for varied operating conditions is developed and shown crucial for accuracy.
- A LSTM-based soft sensor is shown to be superior in trajectory estimation against models without memory.

The paper has been structured as follows. Section II reviews LSTM architecture and relevant structural variations. Section III explains the required development and training procedure for a soft sensor based on LSTM. The experimental results of the LSTM-based soft sensor are shown in Section IV and discussed in Section V. Finally, the paper is summarised in Section VI.

II. LSTM AS IMPLICIT STATE SPACE MODEL

A. RNNs

RNNs are known for their capability to learn temporal patterns from sequential data such as time series data. RNNs learn these temporal patterns by processing each time instant separately and by communicating information of previously

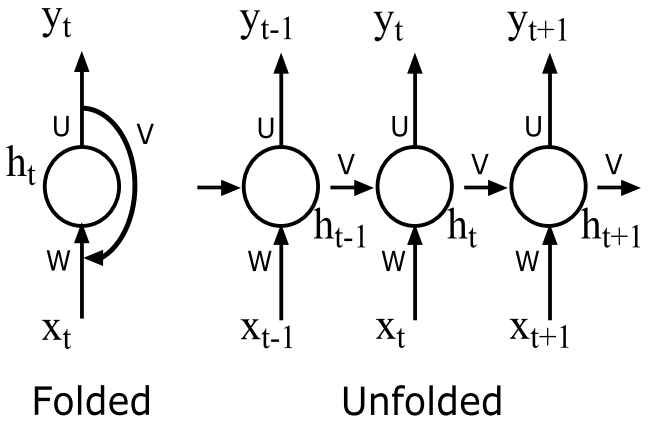

FIGURE 1. Recurrent neural network graph folded (left) and unfolded over 3 timesteps (right).

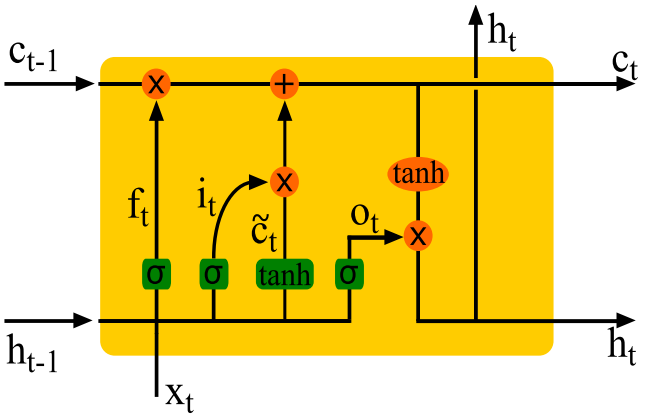

FIGURE 2. The computation graph of an unfolded LSTM over one timestep *t*.

detected patterns forward along the sequential process, and thus they are described as recurrent. This sequential process is visualised as generalised computation graphs in Fig.1. On the left, a folded RNN receives input x_t and a hidden state h_{t-1} from the previous time instant. For each time instant, the RNN also computes an output value y_t. In addition, these computations are expressed as an unfolded graph on the right in Fig. 1. The computations related to the hidden state are shown in (1), where a is an arbitrary non-linear activation function, W is the set of weights optimised for processing the inputs x_t, V is the set of weights optimised for processing the signals h_{t-1} from previous time step, and b_i is an additional optimisable bias term. Finally, the computations related to the output y_t are shown in (2), where b is an arbitrary non-linear activation function, U is the set of weights used to process the hidden state h_t, and b_o is an additional optimisable bias term.

$$h_t = a(Wx_t + Vh_{t-1} + b_i) \tag{1}$$
$$y_t = b(Uh_t + b_o) \tag{2}$$

B. LSTM STRUCTURE

Long short-term memory (LSTM) [24] is a recurrent neural network structure that can learn to perceive and memorize patterns over thousands of timesteps [28]. LSTM has been shown to perform accurately in many problem domains including sequential data [28]. Examples of these problem domains are speech recognition [29], translation [30], and rotor system fault diagnosis [31].

The exceptional temporal learning capabilities of LSTM build on the gating mechanism that allows a sophisticated control over the communication between the sequential computations. The communication between the sequential computations relies on two signals commonly known as the cell state c_t and hidden state h_t. The gating mechanism consisting of four gates can learn the sequential patterns in the input signals x_t. The gates are optimised to either change or not to change the values in the communication signals c_t and h_t. With timely changes to these recurrent signals, the signals may be passed forward through many sequential computations with little or no change. This property efficiently decreases the vanishing and exploding gradient problems related to the gradient based optimisation of RNNs.

The four LSTM gates process inputs at each time instant with a shared number of parameters. The computations related to the forget gate f_t, input gate i_t, output gate o_t and update gate $\tilde{c}_t$ are shown respectively in (3), (4), (5) and (6). In these equations, the sets of weights W_f, W_i, W_o and W_c multiply the concatenated input of the current time instant x_t and the previous hidden state h_{t-1}. Additionally, bias terms b_f, b_i, b_o and b_c may be added to the gate values. Finally, a non-linear activation function is applied to all computed values. The non-linear activation function σ in (3), (4), (5) is commonly the sigmoid function, which scales all the gate values between 0 and 1. The non-linear *tanh* activation function in (6) is the commonly known hyperbolic tangent, which scales the candidate values $\tilde{c}_t$ to the range between -1 and 1. Fig. 2 shows these gates as the green blocks with the corresponding activation labeled on top of each block. The orange circles and ellipse correspond to point-wise operations, where $\times$ is multiplication, $+$ is summation and *tanh* is the hyperbolic tangent. These point-wise operations are shown in (7) and (8), where the previous cell state c_{t-1} is first updated, and then the hidden state h_t is computed with the output gate o_t and the updated cell state c_t. The updated cell state c_t and the hidden state h_t are then passed forward to the next computation step $t + 1$. A copy of the updated hidden state h_t moved upwards in the LSTM computation graph in Fig. 2 is sometimes adopted as the output of the cell at the current timestep t.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$
$$\tilde{c}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{7}$$
$$h_t = o_t \odot tanh(c_t) \tag{8}$$

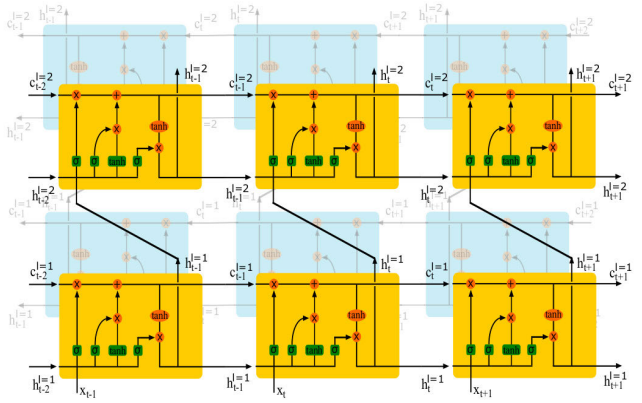

FIGURE 3. Computation graph of an unfolded bidirectional LSTM with two stacked layers over three consecutive time steps.

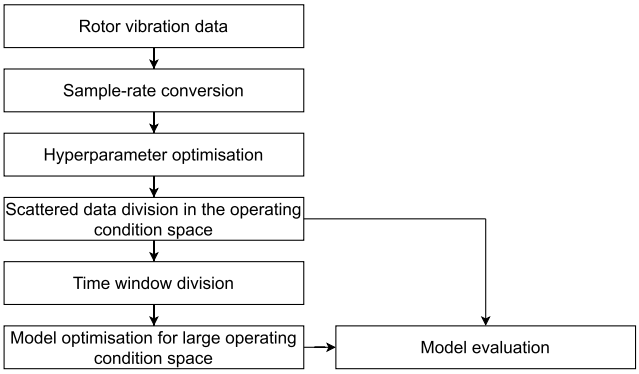

FIGURE 4. Flowchart of the developed optimisation procedure for LSTM-based soft sensor for rotor vibration.

C. TYPICAL VARIATIONS TO LSTM NETWORK STRUCTURES

The learning capabilities of LSTM can be enhanced by increasing the amount of adjacent LSTM cells at each time instant. The number of adjacent cells in an LSTM network can be increased by stacking LSTM cells and by introducing a bidirectional path [32]. A stacked LSTM network can learn more non-linear relations between input and output sequences, since the network computes the output h_t for each time-instant with multiple cells. The stacked cells consider the output h_t from the previous cell as their input x_t. A bidirectional LSTM has a broader perception on the sequence that it processes than a uni-directional LSTM. The bidirectional LSTM network processes the input sequence in two directions, i.e. forward and backward. A stacked and bidirectional LSTM network architecture consisting of two layers is shown in Fig. 3.

D. RNNs AS IMPLICIT STATE SPACE MODEL

Model-based soft sensors typically rely on state space models. State space models are based on physical equations of the system dynamics. With these physical equations, a relation between states and system inputs can be constructed. An example of such state space models is a linear, time-invariant discrete time system shown in (9). The state space model consists of the system equation and the output equation. In the system equation, the new state $x(k+1)$ at time step $k+1$ is computed with the state transition matrix A, current state $x(k)$, input matrix B and input values $u(k)$. In the output equation, the system output $y(k)$ of the current time step is computed with the output matrix C, current state $x(k)$, feedthrough matrix D and system input values $u(k)$.

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned} \tag{9}$$

RNNs can learn to function similarly as state space models from the input-output perspective. After the model optimisation, RNNs can compute the time series inputs deterministically. That is, the hidden state and the output values are computed for each time step based on the order and the values in the input time series data. More specifically, an RNN as shown in Fig. 1 computes an output y_t and a hidden state h_t from the input x_t and the previous hidden state h_{t-1}. These values correspond to the state space model output $y(k)$, system state $x(k)$, input $u(k)$ and previous system state $x(k-1)$ respectively. Furthermore, the weights W, V and U can be considered to parameterise the state space model matrices A, B, C and D. However, showing this correspondence is difficult since inferencing RNN computations and the hidden state h_t is hindered by the stochasticity of the weight optimisation algorithm. That is, each RNN has likely a unique set of weights after every optimisation run. Due to the difficulty of inference and similarities between these two approaches, RNNs can be considered as implicit state space models. For some soft sensor applications requiring output values for every timestep, for example trajectory estimation, employing RNNs as implicit state space models may be the preferable choice. Many RNN-based architectures employing non-deterministic time series input processing, such as encoder-decoder structures or attention mechanisms, may hinder the soft sensor performance due to excessive model complexity.

III. LSTM-BASED SOFT SENSOR FOR TIME SERIES DATA

This section explains the optimisation and training procedure of LSTM-based soft sensor for rotor vibration. The procedure is covered in three subsections. In addition, Fig. 4 shows the whole procedure. In Fig. 4, sample-rate conversion, scattered data division in the operating condition space and time window division relate to preprocessing the rotor vibration data. These preprocessing techniques are described in Section III-A. Hyperparameter optimisation concerns the time window division parameters, the model architecture parameters and model optimisation parameters. Section III-B explains the model architecture selection. Section III-C visits the model optimisation for large operating condition space and model evaluation.

A. PREPROCESSING TIME SERIES DATA FOR A DATA-DRIVEN SOFT SENSOR

Neural networks can learn directly from raw data without complex feature extraction techniques. However, there are a number of preprocessing methods for time series data that can enhance RNN-based soft sensor training convergence and test performance. Such data preprocessing techniques are sample-rate conversion, time window division and scattered division to training and test data. These methods increase the consistency between the input variable distributions without losing significant amount of information.

Sample-rate conversion increases the consistency of the training data distribution. Sample-rate conversion is useful, if the sampling rate varies between the training time series samples. Sample-rate varies in the time series data sampled from the rotor system, if the sampling is triggered with a rotary encoder and the rotating speed changes. Sample-rate conversion downsamples the data with higher sample-rates to the lowest sample-rate in the dataset. However, downsampling requires caution, because the downsampled time series signals should still cover the input and output trajectories with high resolution. In this work, all time series samples were converted to the 2 kHz sample-rate.

Time window division serves two purposes: increasing the amount of training data, and introducing more stochasticity in the training. The amount of training data can be increased by sub-sampling overlapping time windows from each time series data sample consisting of multiple time steps. The number of time windows extracted from a time series data sample depends on the number of the overlapping time steps between the time windows, the length of time windows and the length of the sample. Fig. 5 illustrates time window division of a generic time series data sample consisting of vibration data. In this work, the training dataset consisted of such sub-sampled time windows from multiple time series samples. These time windows were then drawn from the training dataset in a random order during training. Randomly drawing the training time windows ensures that the angular position of the rotor at the first and at the final time step changes uniformly. This way, RNN optimisation was subjected to less bias. Furthermore, shorter time windows may avert vanishing and exploding gradient problems related to RNN optimisation.

Training a data-driven soft sensor for a dynamical system, such as a rotor system, requires a large amount of trajectories sampled at varied operational conditions. For example, the rotor rotating speed should vary in the training dataset between the typically used minimum and maximum operating speeds. If the training data includes trajectories from sparse distribution of operating conditions, the model performance is likely unsatisfactory under some typical operating conditions.

In this research, the training and test sets were divided by randomly selecting time series samples from a grid of operating conditions. Fig. 6 shows a simplified training and test data division along a rotating speed grid. In Fig. 6,

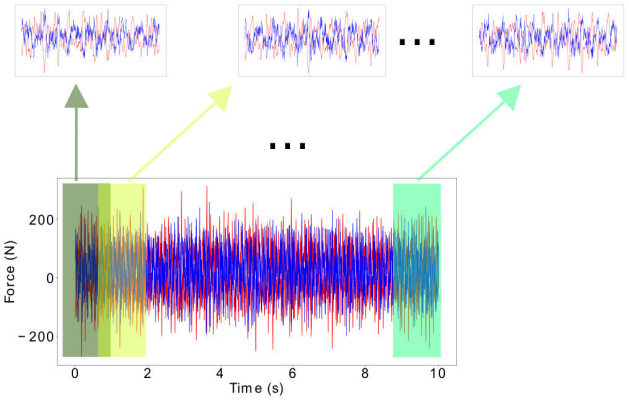

FIGURE 5. Overlapping time windows were extracted from a 10-second-long time series sample consisting of two vibration signals. The vibration signals were horizontally (red) and vertically (blue) directed bearing reaction forces.

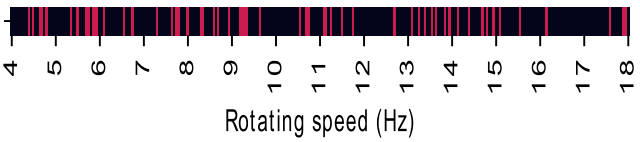

FIGURE 6. A set of time series samples randomly divided to the training dataset (red lines) and to the test dataset (black lines) based on the rotating speed of the system.

the red and the black lines are the time series samples, such as in Fig. 5 and 9, which would be included in the training set and the test set respectively. The advantage of this division technique is that the training data likely covers the operating condition space with satisfying resolution. That is, the model learns to estimate trajectories under many operating conditions, such as different rotating speeds, which leads to better performance under all operating conditions nearby the conditions in the training dataset.

B. LSTM NETWORK ARCHITECTURE

A suitable LSTM architecture for estimating the system output y_t at each time instant t has enough computational capacity, memory and nonlinear representational capability. These attributes relate to the length of the cell state and the number of layers. By increasing the amount of values the cell state c_t can hold to the next time step $t + 1$, the memory capacity increases. The number of parameters in each gate f_t, i_t, o_t and $\tilde{c}_t$ are linearly dependent on the number of values held in the cell state. Therefore, by increasing the memory capacity, the LSTM computational capacity also increases. A large cell state may be crucial, if many input variables in x_t show interdependence over multiple time steps. The nonlinear representational capability of LSTM network can be increased by stacking multiple LSTM cells and adding a bidirectional path to the network. After stacking LSTM cells, the superimposed cells each introduce another set of gates and memory to the function between the input variables x_t and the output y_t.

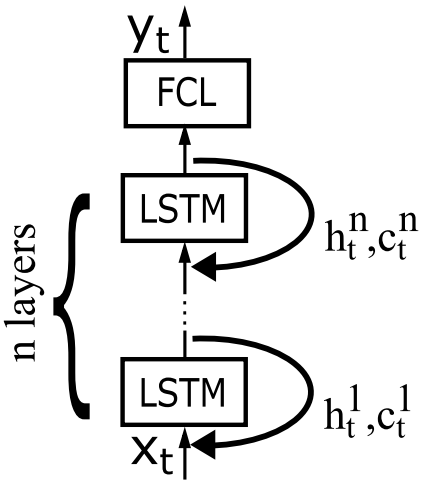

FIGURE 7. Folded LSTM architecture for soft sensors with n LSTM layers. Each LSTM layer is bidirectional. FCL is a fully-connected layer transforming the final LSTM layer output h_t^n to the correct output dimension.

Bidirectionality further increases the nonlinearity between the input variables x_t and the output y_t. The bidirectional paths of LSTM compute the output based on previous values from both directions in the sequence. Bidirectional LSTMs may increase the performance in computing output trajectories at the expense of introducing some time lag to the soft sensor outputs. The time lag is caused by the path processing input variables backwards in time, since it needs some inputs from the future to compute output y_t for the current time instant. This might hinder using bidirectional LSTMs in some real-time tasks.

Fig. 7 presents the LSTM-based soft sensor architecture used in this research. The architecture consists of n superimposed LSTM layers and a fully-connected layer (FCL). The bidirectional paths, as shown in Fig. 3, are not explicitly visible, but considered to be included in each LSTM block in the figure. The first LSTM layer processes input signals one time instant at time. I.e., the values of x_t correspond to the values sampled by the sensors at time instant t. The following hidden LSTM-layers compute more hierarchical representations of the system state. The final hidden LSTM-layer passes the hierarchical system state representation to the FCL, which computes the output values y_t corresponding to current time instant. A suitable number of LSTM layers is specific to the soft sensor application, and needs to be searched. In this research, random search was employed to find a suitable number of layers simultaneously with other hyperparameters, such as the size of the cell c_t.

C. LSTM-BASED SOFT SENSOR TRAINING AND EVALUATION

In this study, the data-driven soft sensor was trained with supervised learning. That is, the estimated displacement trajectories Y consisting of the displacement values $y_{t,k}$ at each time instant t were compared to the measured trajectories Y^{true}. For each corresponding trajectory Y and Y^{true}, an error term was computed with mean squared error (MSE). The error term is expressed in (10). In (10), k denotes rotor displacement in horizontal and vertical directions. The parameters of the LSTM-based soft sensor were updated based on these error terms with Adam optimiser [33].

Two different weight saving schemes were designed for the model optimisation. If many operating conditions were included in the training dataset, the model weights were saved every 500 batches, if the running MSE was smaller than the previously smallest running MSE. 500 batches likely included trajectories with largely varying operating conditions, and thus a validation dataset was not considered necessary. If only one operating condition was included in the dataset, a validation trajectory was excluded from the training trajectory. The weights were saved every time the validation MSE decreased below the previous best validation MSE. The set of weights saved with the lowest error were utilised in testing. The test performance of the model was evaluated with the mean absolute error (MAE), since it is more interpretable than MSE. The formula for this test performance metric is shown in (11).

$$MSE = \frac{1}{K} * \frac{1}{T} * \sum_{k=1}^{K} \sum_{t=1}^{T} (y_{t,k} - y_{t,k}^{true})^2 \tag{10}$$

$$MAE = \frac{1}{K} * \frac{1}{T} * \sum_{k=1}^{K} \sum_{t=1}^{T} |(y_{t,k} - y_{t,k}^{true})| \tag{11}$$

IV. VALIDATION OF LSTM-BASED SOFT SENSOR

The vibration of a rotor system is sensitive to the operating conditions, such as the rotating speed and the support stiffness. Rotor systems vibrate excessively, when excited at their natural frequency. The rotor vibration also significantly varies between different rotor systems, even in a case of two different realizations of a single rotor system design. The uncertainties tend to grow with the rotor system size. For example, the foundation stiffness of a rotor changes the natural frequencies of the whole system. This section investigates the performance of a LSTM-based soft sensor under two scenarios: varying rotor speed and horizontal foundation stiffness.

A. DATA DESCRIPTION

The dataset consists of multiple time series samples each including 100 revolutions of the rotor at constant rotating speed and at constant foundation stiffness. The rotating speeds in the dataset range from 4 Hz to 18 Hz at 0.05 Hz steps. For each constant rotating speed, the dataset also includes multiple time series samples measured at different horizontal foundation stiffnesses. A horizontal stiffness controller was designed for acquiring this dataset. More details of the controller and its effect on the rotor system are demonstrated in [34]. This work considers merely the resulting horizontal foundation stiffness in the range between 2.04 MN/m and 18.32 MN/m. For convenience, appendix A shows some

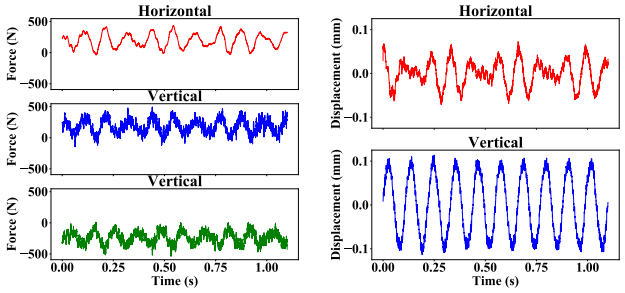

FIGURE 8. A subsample of measured forces from drive-end bearing housing (left) and corresponding measured center-point displacement of the rotor middle cross-section (right). The rotating speed was 9.15 Hz and the horizontal foundation stiffness was 3.2 MN/m.

examples of rotor vibration at different horizontal foundation stiffnesses.

Overall, a time series sample in the dataset consists of eight signals, which were acquired with encoder triggered phase locked sampling. Therefore, 1024 samples were acquired from every sensor during each revolution. All time series samples in the dataset have been resampled to 2 kHz sample-rate for this study. The data in each time series sample consists of two signals of center-point displacement (horizontal and vertical) of the rotor middle cross-section, three drive-end bearing housing vibration signals (forces) and three tending-end bearing housing vibration signals (forces). Fig. 8 shows examples of such signals. The forces measured at the drive-end bearing housing are on the left, and on right are the corresponding center-point displacements. Tending-end bearing housing vibration signals appear similar to the corresponding drive-end signals in the time domain. There are two vertical force sensor signals per bearing housing due to the measurement setup. More detailed information regarding the dataset can be found in [27].

B. EXPERIMENTAL SETUP

The LSTM-based soft sensor for rotor vibration was validated in three cases studies, which all followed a similar scheme. In each case study, the model computed the trajectory of the rotor middle cross-section center-point displacement in horizontal and vertical directions. These displacement trajectories correspond to the lateral vibration of the rotor. The inputs for the model were the horizontal and vertical vibration signals measured with three force sensors at both bearing housings of the rotor. The parameters for the LSTM-based soft sensor and model optimisation are detailed in Table 1. These parameters were found with random search. More details regarding the random search procedure are found in appendix B.

1) CASE STUDY I: LSTM-BASED SOFT SENSOR PERFORMANCE UNDER FEW OPERATING CONDITIONS

This case study compares the LSTM-based soft sensor to other regression algorithms without a memory and with a limited perception over the input trajectories. The baseline models are linear regression (LR) and three formulations of

TABLE 1. Model and training parameters.

Parameter	value
Training window length	2763
Training window overlap	74
Model input dimension	6 * 2763
Model output dimension	2 * 2763
Batch size	2
Cell size	81
Number of layers	4
Fully connected size	2*81
Dropout	0.14
Learning rate	0.001

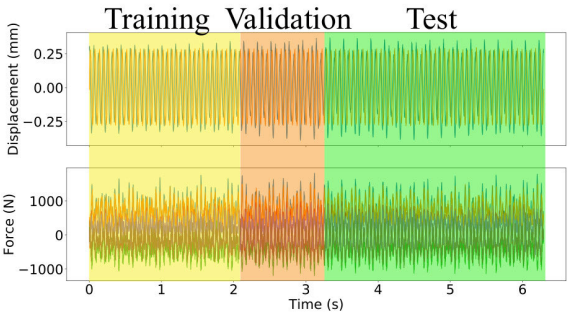

FIGURE 9. A time series sample consisting of 100 revolutions at 16 Hz rotating speed and at 18.32 MN/m horizontal foundation stiffness was split to training, validation and test data. Training data (yellow) was the first 33 %, validation data (red) was the middle 16 % and test data (green) was the final 50 % of the complete sample. Same division was followed in all tests in Section IV-B1. The upper graph shows horizontal and vertical displacement of the rotor middle cross-section and the lower graph shows the horizontal and vertical bearing reaction forces.

support vector machines (SVM). The model for LR consists of six weights per output trajectory (horizontal and vertical rotor center-point displacements). The weights for LR were optimised with the ordinary least squares method. All SVM-based models rely on the LIBSVM library implementations of ϵ-support vector regression (ϵ-SVR) [35]. The three (ϵ-SVR) models have different kernels. The kernel types were linear (LIN), radial basis function (RBF) and polynomial (POLY). The parameters ϵ and C for the SVR-models were grid searched. The grid search is explained in detail in appendix C.

This case study consisted of five independent tests. In each test, the training and test data shared constant rotating speed and constant horizontal foundation stiffness. Fig. 9 shows the training and test data division in all tests. The first 33 % was utilised in model optimisation and the last 50 % was employed in model evaluation. The middle 16 % was applied to LSTM validation. The first 33 % of every time series sample covered 33 full revolutions. This share of the sample holds enough data to omit the effect of random factors on the model optimisation. Similarly 50 % of the time series sample, corresponding to 50 rotor revolutions, holds enough data to omit the random factor effects on the model test performance. The validation data was considered necessary for LSTM optimisation alone, since other models were optimised deterministically.

TABLE 2. Case study 1: MAE (mm) of LSTM-based soft sensor and baseline models at different operating conditions.

Hz & MN/m	LSTM	LR	SVR-LIN	SVR-RBF	SVR-POLY
4 & 18.32	**0.0158**	0.0232	0.0228	0.023	0.0227
10.5 & 18.32	**0.0185**	0.0445	0.0437	0.032	0.0369
10.5 & 6.69	**0.0134**	0.0253	0.0253	0.0188	0.0201
10.5 & 3.96	**0.0116**	0.0241	0.0241	0.0158	0.0185
16 & 18.32	**0.0792**	0.1101	0.1118	0.1182	0.1113

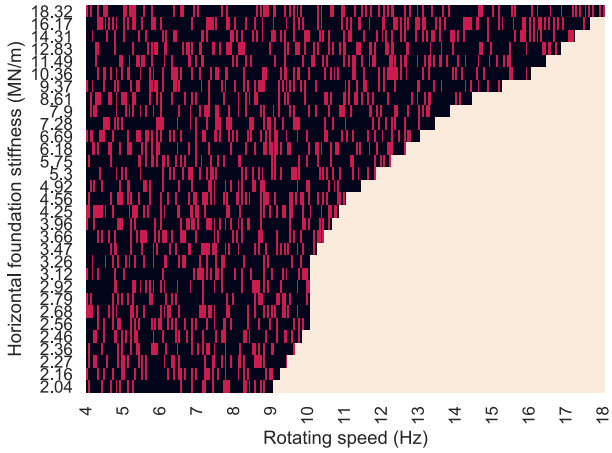

FIGURE 10. Dataset consisted of time series samples measured at varied constant rotating speeds and horizontal support stiffnesses. The dataset was divided to training and test data. The red blocks represent training samples and black blocks the test samples. Each sample includes 100 revolutions similar to Fig. 9. The white area in this operating condition space was not measured in order to avoid excessive rotor vibration at the rotor critical speed.

The input signals for all the models were the six bearing reaction force signals. The output from all the models were the horizontal and vertical rotor center-point displacement signals. The test error was computed as the mean absolute error (MAE) between the measured and estimated rotor displacement signals. Table 2 details the test performance of the LSTM-based soft sensor and the baseline models. LSTM-based soft sensor outperformed all the baseline models in terms of the MAE in these five tests.

2) CASE STUDY II: LSTM-BASED SOFT SENSOR PERFORMANCE UNDER MANY OPERATING CONDITIONS

Rotor systems are typically operated under more varied operating conditions than in the case study with few operating conditions in Section IV-B1. Although, a soft sensor capable of estimating trajectories at many rotating speeds could utilise an ensemble of models, a more desirable solution would require only a single model. Furthermore, more dynamic properties affecting the rotor system vibration exists. Neural networks offer an advantage for soft sensor development due to their capacity to learn from large data. Therefore, in this case study we explored the LSTM-based soft sensor performance over a large range of operating conditions. The operating conditions included constant rotating speeds between 4 Hz and 18 Hz at 0.05 Hz steps and a range of

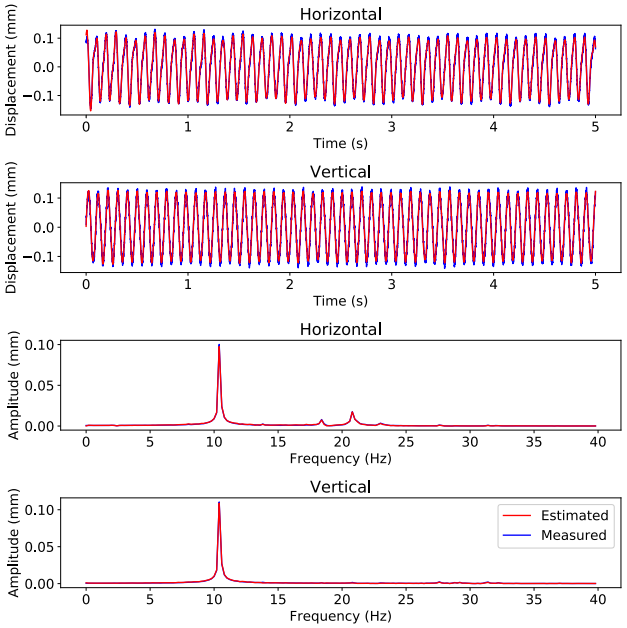

FIGURE 11. LSTM-based soft sensor estimated the first 5 seconds of rotor displacement trajectory from the time series sample shown in Fig. 25. MAE between estimated trajectory (red) and measured trajectory (blue) was 0.0064 mm. Two lower graphs show the trajectory and the estimate in the frequency domain.

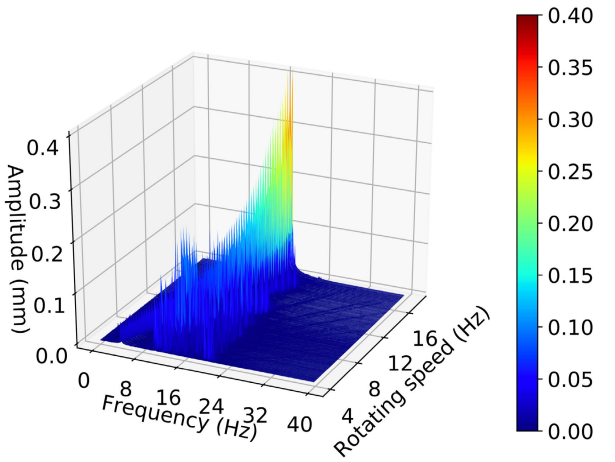

FIGURE 12. DFT collection of the 5-second windows of the measured horizontal displacement trajectories from every time series sample in the test set.

horizontal foundation stiffnesses between 2.04 MN/m and 18.32 MN/m.

The whole dataset of over 5000 time series samples, each such as the one shown in Fig. 8, was divided to training and test sets. The samples in the training set were chosen randomly based on the operating condition, and excluded from the test set. The resulting data division is shown in Fig. 10. The red blocks are the time series samples in the training set, which correspond to 25 % of the samples. The black blocks are the time series samples in the test set, which correspond to 75 % of the samples.

The LSTM-based soft sensor performance was evaluated with one 5-second window from every time series

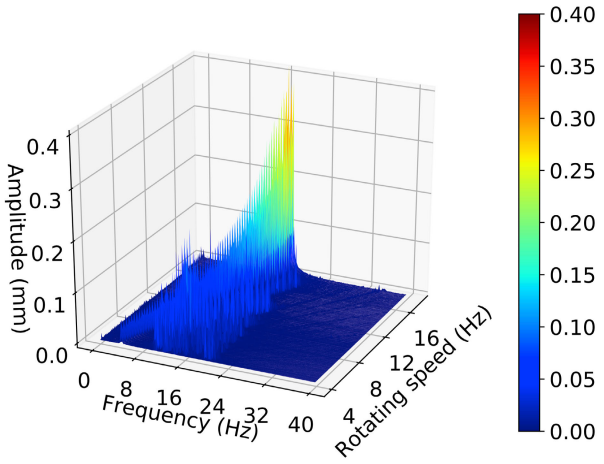

FIGURE 13. DFT collection of the estimated 5-second horizontal displacement trajectories corresponding to the DFTs of the measured horizontal trajectories in Fig. 12.

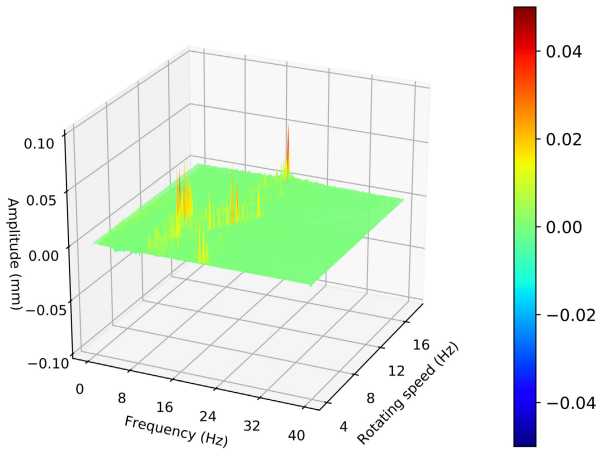

FIGURE 14. Collection of the differences between the DFTs of the corresponding measured and estimated horizontal displacement trajectories in Fig. 12 and 13.

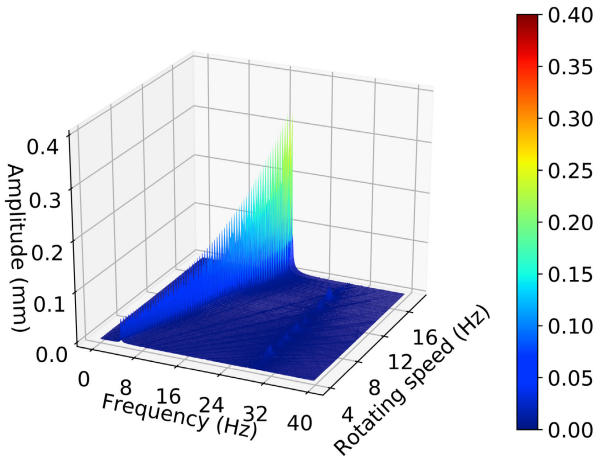

FIGURE 15. DFT collection of the 5-second windows of the measured vertical displacement trajectories from every time series sample in the test set.

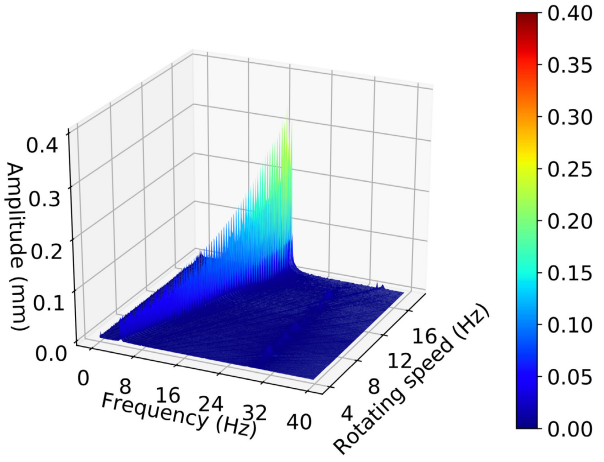

FIGURE 16. DFT collection of the estimated 5-second vertical displacement trajectories corresponding to the DFTs of the measured vertical trajectories in Fig. 15.

sample in the test set. The mean absolute error over these time windows was 0.0063 mm. Fig. 11 shows examples of 5-second windows of estimated and measured displacement trajectories in horizontal and vertical directions and the corresponding discrete Fourier transforms (DFTs) at 10.5 Hz rotating speed and 18.32 MN/m. Fig. 12 shows DFTs of the 5-second windows of horizontal displacement trajectories in the test dataset. Fig. 13 shows the corresponding DFTs of the estimated trajectories. Fig. 14 shows the differences between the corresponding DFTs. Fig. 15, 16 and 17 show respectively the DFTs and their differences in the vertical direction. The difference between the DFTs of the estimated and measured displacement trajectories is consistently less than 0.05 mm in the entire operating condition space, as Fig. 14 and 17 show.

3) CASE STUDY III: LSTM-BASED SOFT SENSOR PERFORMANCE UNDER LIMITED OPERATING CONDITIONS

This case study highlights the importance of training the LSTM-based soft sensor with trajectories drawn from the whole typical operating condition space. The operating condition space was constrained in the training set, as shown in Fig. 18. The model was first trained with time series samples from the rotating speed range between 4 Hz and 9 Hz. Then, the model performance was evaluated with time series samples from the rotating speed range between 9 Hz and 18 Hz. Following the procedures of the case study with many operating conditions in Section IV-B2, the test time windows were 5-second long trajectories from all test time series samples. The mean absolute error between the estimated displacement trajectories and true displacement trajectories was 0.0442 mm. Fig. 19 and 20 show the DFTs of these measured and estimated displacement trajectories in the horizontal direction, respectively. Fig. 21 shows the difference between these DFTs. Fig. 22, 23 and 24 show the corresponding DFTs and their difference in the vertical direction. The difference between the DFT amplitudes are much higher in the rotating speeds that are farther from the rotating speed

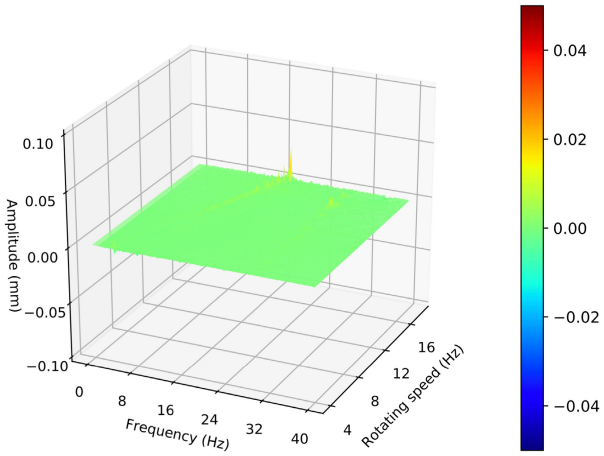

FIGURE 17. Collection of the differences between the DFTs of the corresponding measured and estimated vertical displacement trajectories in Fig. 15 and 16.

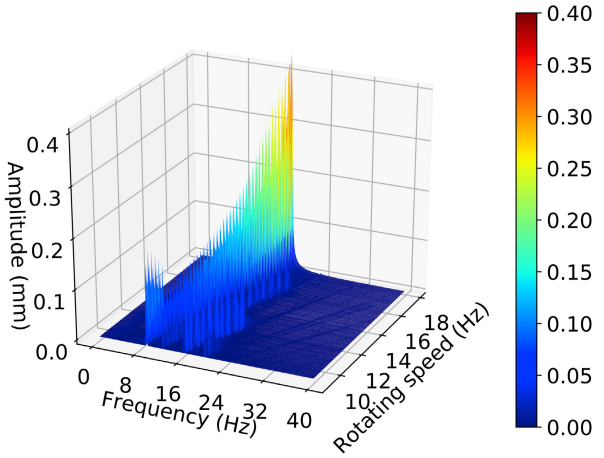

FIGURE 19. DFT collection of the 5-second windows of the measured horizontal displacement trajectories from every time series sample in the test set.

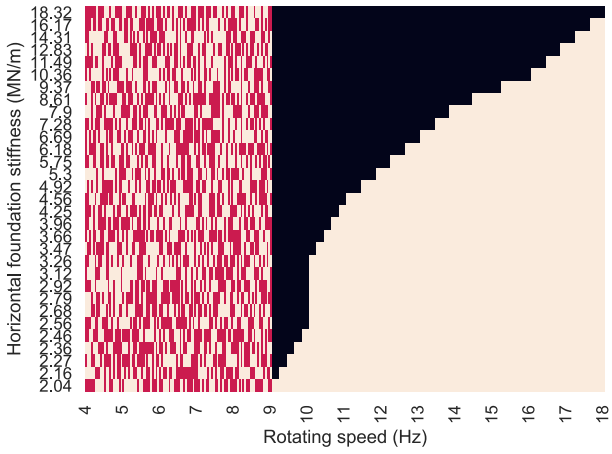

FIGURE 18. The dataset was divided to the training time series samples (red blocks) and to the test time series samples (black blocks). The training data was constrained to the rotating speed range between 4 Hz and 9 Hz. 50 % of the time series samples inside this range were randomly chosen to the training set and the rest were excluded from training. The test data was constrained to the rotating speed range between 9 Hz and 18 Hz.

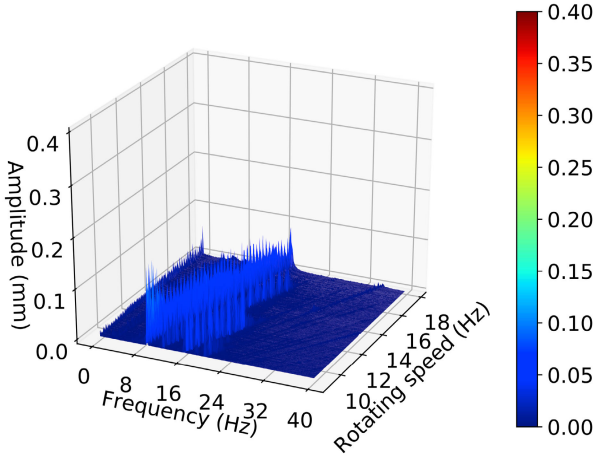

FIGURE 20. DFT collection of the estimated 5-second horizontal displacement trajectories corresponding to the DFTs of the measured horizontal trajectories in Fig. 19.

range utilised in training. These differences increase up to 0.3 mm in horizontal and vertical directions, as Fig. 21 and 24 show.

V. DISCUSSION

Based on the results in the case study with few operating conditions in Section IV-B1, LSTM is a better model for estimating rotor displacement trajectories from bearing housing vibration than linear regression and ϵ-SVRs. The results were consistent over five tests with significantly different rotating speeds and horizontal support stiffnesses. The better performance is likely due to the recurrent way of processing the input trajectories. That is, the model keeps the system state implicitly in the memory between the time steps, which is useful in determining the new system state and the output.

The results in the case study with many operating conditions in Section IV-B2 reveal the potential of the LSTM-based soft sensor. The model learned to estimate trajectories over a large range of operating conditions with a large amount of training data uniformly distributed in the operating condition space. The preprocessing techniques for building a large training set of the time series samples were shown applicable. By splitting the trajectories to shorter time windows and by scattered data division to training and test sets, the model achieved remarkable test performance based on the MAE of 0.0063 mm. Moreover, Fig. 14 and 17 show that the difference between the DFT amplitudes of the estimated and the measured displacement trajectories were small over all test operating conditions. The differences in horizontal and vertical directions were consistently under 0.05 mm. This indicates that the LSTM-based soft sensor generalised over the complete operating condition space with the proposed

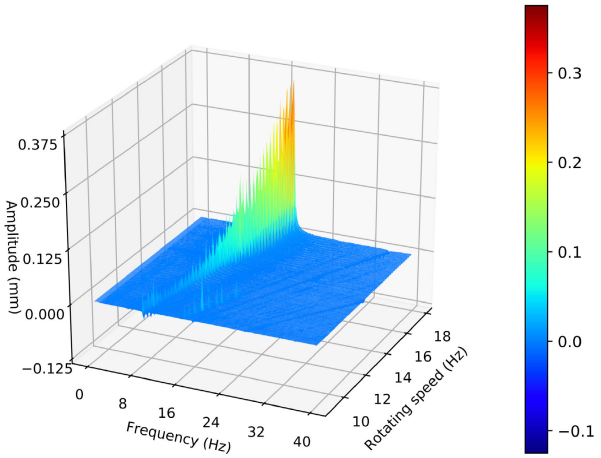

FIGURE 21. Collection of the differences between the DFTs of the corresponding measured and estimated horizontal displacement trajectories in Fig. 19 and 20.

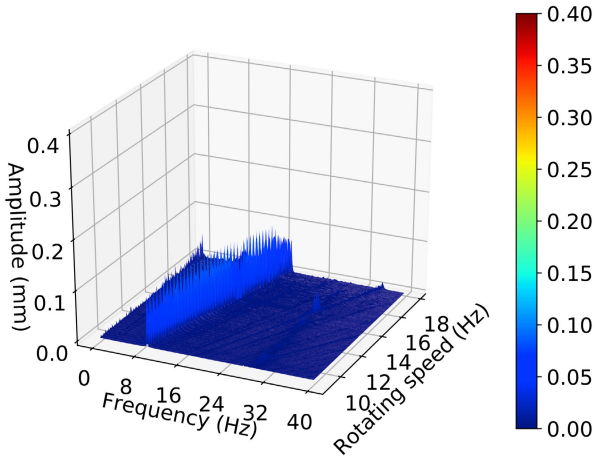

FIGURE 23. DFT collection of the estimated 5-second vertical displacement trajectories corresponding to the DFTs of the measured vertical trajectories in Fig. 22.

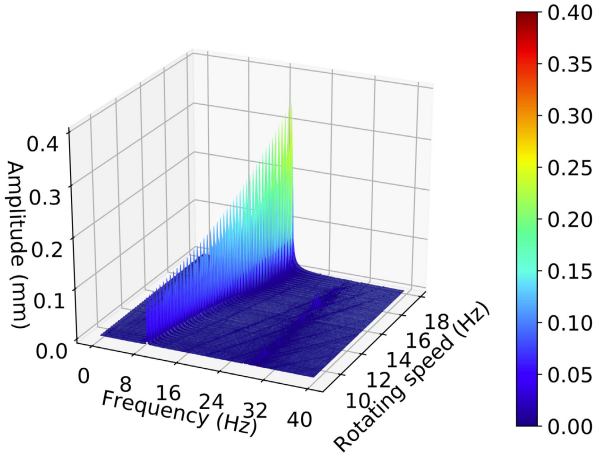

FIGURE 22. DFT collection of the 5-second windows of the measured vertical displacement trajectories from every time series sample in the test set.

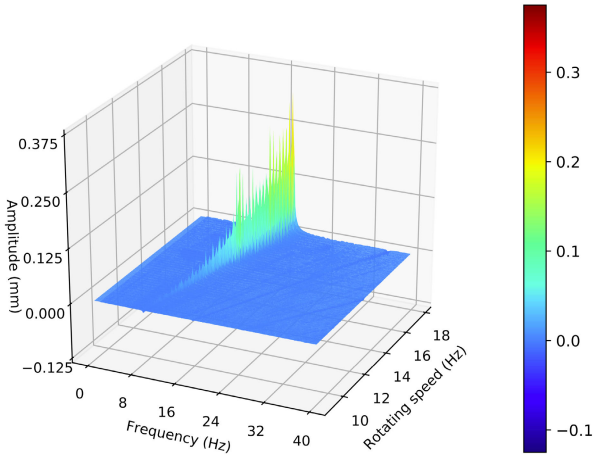

FIGURE 24. Collection of the differences between the DFTs of the corresponding measured and estimated vertical displacement trajectories in Fig. 22 and 23.

training procedure. Furthermore, merely 25 % of the time series samples was required for training.

The results in the case study with limited operating conditions in Section IV-B3 show the limitations of the LSTM-based soft sensor. The model performance drastically decreased in the operating condition space that was excluded from training data. The decrease in performance can be observed by comparing Fig. 14 and 17 to Fig. 21 and 24. The differences between the DFT amplitudes of the measured and the estimated trajectories increase close to 0.3 mm in horizontal and vertical directions at the higher rotating speeds. These differences are many magnitudes higher than the corresponding differences in case study with many operating conditions in Section IV-B2. It is evident from the figures in these case studies that the model underestimates the scale of amplitudes at the higher rotating speeds, if the training data includes only lower rotating speeds. These results confirm that the proposed training procedure is crucial for LSTM-based soft sensor generalisation over large range of operating conditions.

VI. CONCLUSION

A training procedure was developed for bidirectional LSTM-based soft sensor for rotor displacement trajectory estimation at high sample-rates. The first case study compared the LSTM-based soft sensor trained with time windows to other regression algorithms without a memory and showed that LSTM-based soft sensor was superior in estimation accuracy based on MAE in five different tests. The second case study showed that LSTM-based soft sensor trained with trajectories from randomly chosen operating conditions resulted in a high accuracy over a large range of operating conditions. The MAE between the estimated and measured test trajectories was 0.0063 mm. Moreover, the differences between the DFT amplitudes of the estimated and measured test trajectories were below 0.05 mm over the complete operating condi-

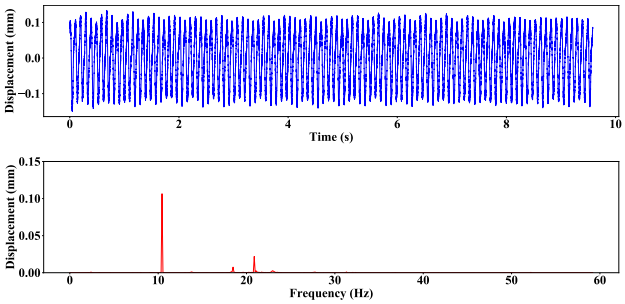

FIGURE 25. 100 rounds of measured rotor center-point displacement in the horizontal direction at the rotor middle cross-section. The nominal rotating speed was 10.5 Hz and the horizontal foundation stiffness was set to 18.32 MN/m.

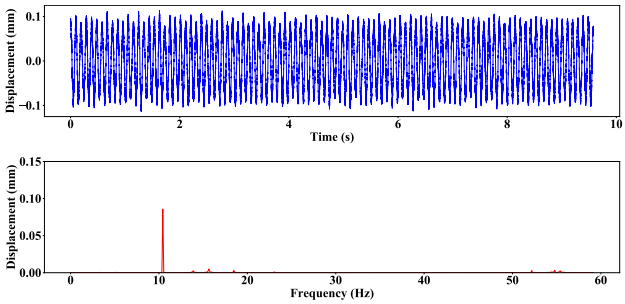

FIGURE 26. 100 rounds of measured rotor center-point displacement in the horizontal direction at the rotor middle cross-section. The nominal rotating speed was 10.5 Hz and the horizontal foundation stiffness was set to 6.69 MN/m.

tion space. The importance of adequate training data distribution for model reliability was shown in the third case study. The differences between the DFT amplitudes between the estimated and measured test trajectories were up to 0.3 mm, if the training data was limited in the operating condition space. Furthemore, limiting operating condition space in the training data resulted in MAE of 0.0442 mm between the estimated and measured displacement trajectories. The findings related to LSTM-based soft sensor generalisation over varied operating conditions are significant for further soft sensor research aiming for reliable industrial applications for rotor displacement trajectory estimation.

APPENDIX A
EXAMPLES OF ROTOR VIBRATION AT DIFFERENT FOUNDATION STIFFNESSES

Fig. 25, 26 and 27 show the effect of horizontal foundation stiffness on the horizontal rotor vibration at constant nominal rotating speed of 10.5 Hz. A higher horizontal foundation stiffness results in a higher rotor vibration amplitude at the same frequency that the rotor rotates. This change can be observed by comparing the DFTs (red and lower plot) in Fig. 25, 26 and 27. In the figures, the vibration amplitudes at 10.5 Hz are approximately $>$ 0.1 mm, $<$ 0.1 mm and 0.05 mm, when the horizontal foundation stiffness is 18.32 MN/m, 6.69 MN/m and 3.96 MN/m, respectively. Furthermore, the vibration amplitude at the known natural frequency of the rotor (21 Hz) is visible only in Fig. 25. This implies that the horizontal foundation stiffness has an effect on the natural frequencies of the system.

APPENDIX B
RANDOM SEARCH FOR LSTM-BASED SOFT SENSOR AND OPTIMISATION PARAMETERS

The motivation was to find such set of hyperparameters related to the model architecture and to the model optimisation, that the model could learn to estimate the rotor center-point displacement trajectories from the bearing housing force trajectories. Table 3 shows these hyperparameters and the respective search ranges for random sampling. All together 27 different sets of hyperparameters were sampled

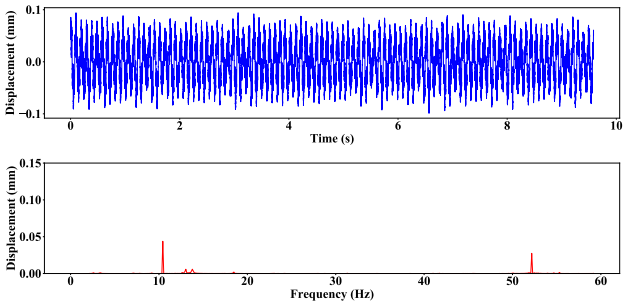

FIGURE 27. 100 rounds of measured rotor center-point displacement in the horizontal direction at the rotor middle cross-section. The nominal rotating speed was 10.5 Hz and the horizontal foundation stiffness was set to 3.96 MN/m.

TABLE 3. Model and training hyperparameter ranges for random search.

Parameter	range
Training window length	[500,4000]
Training window overlap	[10,5000]
Batch size	[2,10]
Cell size	[25,125]
Number of layers	[1,5]
Dropout	[0.05,0.2]
Learning rate	[0.01,0.001]

and compared against each other in tests conducted with small datasets. Each small dataset consisted of one time series sample, including 100 rounds of rotation. 30 such time series samples were selected for the basis of comparison between the hyperparameter sets. Each of the 30 time series samples was randomly drawn from the rotating speed range between 10 Hz and 16 Hz and from the whole horizontal stiffness range. That is, each hyperparameter set was trained and tested 30 times, once on every small dataset. The 30 small datasets were same for all the 27 hyperparameter set tests.

Each hyperparameter set was tested with the following scheme. First a small dataset was divided in to time windows. The time windows were randomly divided in to training and test window sets including 66 % and 33 % of the windows, respectively. The model was then optimised with the training window set to compute rotor displacement trajectories from bearing housing reaction forces. The trained model was then

evaluated on the test windows by computing the average of mean squared errors over all test windows. This test was repeated on each of the 30 small datasets. The set of hyperparameters that resulted in the lowest average test error over the 30 tests are shown in Table 1.

APPENDIX C
SVR GRID SEARCH

The parameters C and ϵ have an impact on the SVR performances. Therefore, they were grid searched for each SVR -model. The grid for both parameters consisted of the following values: 0.1, 0.05, 0.01, 0.005 and 0.001. All combinations of these values were tested with all SVR -models. The grid search was conducted on the time series sample shown in Fig. 25. The first 33 % of the sample was utilised in fitting the models and the final 50 % of the sample was utilised in evaluating the model performances. The parameters resulting in highest test performance were used in all the tests in IV-B1. The searched parameters are listed in Table 4.

TABLE 4. Grid searched hyperparameters for SVR-models.

Model	C	ϵ
SVR-LIN	0.01	0.005
SVR-RBF	0.1	0.001
SVR-POLY	0.01	0.001

ACKNOWLEDGMENT

The calculations presented above were performed using computer resources within the Aalto University School of Science "Science-IT" project.

REFERENCES

[1] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, Apr. 2009.
[2] G. Goodwin, "Predicting the performance of soft sensors as a route to low cost automation," *Annu. Rev. Control*, vol. 24, no. 1, pp. 55–66, 2000.
[3] F. Auger, M. Hilairet, J. M. Guerrero, E. Monmasson, T. Orlowska-Kowalska, and S. Katsura, "Industrial applications of the Kalman filter: A review," *IEEE Trans. Ind. Electron.*, vol. 60, no. 12, pp. 5458–5471, Jan. 2013.
[4] J. Stephant, A. Charara, and D. Meizel, "Virtual sensor: Application to vehicle sideslip angle and transversal forces," *IEEE Trans. Ind. Electron.*, vol. 51, no. 2, pp. 278–289, Apr. 2004.
[5] M. Manngård, W. Lund, J. Keski-Rahkonen, J. Nänimäinen, V.-P. Saarela, J. Björkqvist, and H. T. Toivonen, "Estimation of propeller torque in azimuth thrusters," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 140–145, 2019.
[6] A. J. de Assis and R. M. Filho, "Soft sensors development for on-line bioreactor state estimation," *Comput. Chem. Eng.*, vol. 24, nos. 2–7, pp. 1099–1103, 2000.
[7] L. Hong, I. P. Girsang, and J. S. Dhupia, "Identification and control of stick–slip vibrations using Kalman estimator in oil-well drill strings," *J. Petroleum Sci. Eng.*, vol. 140, pp. 119–127, Apr. 2016.
[8] E. Alhoniemi, J. Hollmén, O. Simula, and J. Vesanto, "Process monitoring and modeling using the self-organizing map," *Integr. Comput.-Aided Eng.*, vol. 6, no. 1, pp. 3–14, 1999.
[9] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry," *Comput. Chem. Eng.*, vol. 32, nos. 1–2, pp. 12–24, Jan. 2008.
[10] J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, and R. M. Filho, "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Comput. Chem. Eng.*, vol. 33, no. 1, pp. 43–49, 2009.
[11] W. Yan, H. Shao, and X. Wang, "Soft sensing modeling based on support vector machine and Bayesian model selection," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1489–1498, Jul. 2004.
[12] H. Zhu and T. Liu, "Rotor displacement self-sensing modeling of six-pole radial hybrid magnetic bearing using improved particle swarm optimization support vector machine," *IEEE Trans. Power Electron.*, vol. 35, no. 11, pp. 12296–12306, Nov. 2020.
[13] T. Liu, H. Zhu, M. Wu, and W. Zhang, "Rotor displacement self-sensing method for six-pole radial hybrid magnetic bearing using mixed-kernel fuzzy support vector machine," *IEEE Trans. Appl. Supercond.*, vol. 30, no. 4, pp. 1–4, Jun. 2020.
[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," May 2019, *arXiv:1810.04805*.
[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," Dec. 2013, *arXiv:1312.5602*.
[17] Q. Sun and Z. Ge, "A survey on deep learning for data-driven soft sensors," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 5853–5866, Sep. 2021.
[18] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3168–3176, May 2020.
[19] Y.-L. He, Y. Tian, Y. Xu, and Q.-X. Zhu, "Novel soft sensor development using echo state network integrated with singular value decomposition: Application to complex chemical processes," *Chemometric Intell. Lab. Syst.*, vol. 200, May 2020, Art. no. 103981.
[20] X.-F. Yuan, L. Li, Y. Shardt, Y.-L. Wang, and C.-H. Yang, "Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development," *IEEE Trans. Ind. Electron.*, vol. 68, no. 5, pp. 4404–4414, May 2021.
[21] X. Wang and H. Liu, "Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction," *Adv. Eng. Inform.*, vol. 36, pp. 112–119, Apr. 2018.
[22] X. Wang and H. Liu, "A knowledge- and data-driven soft sensor based on deep learning for predicting the deformation of an air preheater rotor," *IEEE Access*, vol. 7, pp. 159651–159660, 2019.
[23] P. Lian, H. Liu, X. Wang, and R. Guo, "Soft sensor based on DBN-IPSO-SVR approach for rotor thermal deformation prediction of rotary air-preheater," *Measurement*, vol. 165, Dec. 2020, Art. no. 108109.
[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[25] W. Ke, D. Huang, F. Yang, and Y. Jiang, "Soft sensor development and applications based on LSTM in deep neural networks," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–6.
[26] M. Rao, Q. Li, D. Wei, and M. J. Zuo, "A deep bi-directional long short-term memory model for automatic rotating speed extraction from raw vibration signals," *Measurement*, vol. 158, Jul. 2020, Art. no. 107719.
[27] R. Viitala, J. Miettinen, T. Tiainen, and R. Viitala, "Rotor & bearing vibration dataset," Mendeley Data, Dept. Mech. Eng., Aalto Univ., Espoo, Finland, Tech. Rep., Aug. 2020. [Online]. Available: https://data.mendeley.com/datasets/pdrxyfprfk/1, doi: 10.17632/pdrxyfprfk.1.
[28] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[29] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *Proc. 31st Int. Conf. Mach. Learn.*, vol. 32, Jun. 2014, pp. 1764–1772.
[30] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. NIPS*, vol. 2, Dec. 2014, pp. 3104–3112.
[31] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.
[32] Y. Yu, X. Si, C. Hu, and Z. Jianxun, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.
[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, May 2015, pp. 1–15.

[34] R. Viitala and R. Viitala, "Method and device to investigate the behavior of large rotors under continuously adjustable foundation stiffness," *J. Vibroeng.*, vol. 22, no. 5, pp. 1037–1054, Aug. 2020.
[35] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.

JESSE MIETTINEN received the M.Sc. degree in mechanical engineering from Aalto University, in 2020, where he is currently pursuing the D.Sc. degree in mechanical engineering with the Department of Mechanical Engineering. In his master's studies, he focused on machine learning and mechanical engineering. In 2019, he started as a Research Assistant with Aalto University. His research interest includes deep learning applications for rotor system monitoring and fault diagnosis.

TUOMAS TIAINEN received the M.Sc. and D.Sc. degrees in mechanical engineering from Aalto University, in 2018 and 2020, respectively. In his master's thesis, he developed an automatic assessment system for mechanical engineering CAD exercises. He completed his Ph.D. dissertation on multi-probe roundness measurement of large rotors, in late 2020. In 2020, he visited Technische Universität Darmstadt researching magnetic bearings and rotating machinery. His current research interests include geometry measurement of large rotors, rotor dynamics, and data-driven models for large rotating machinery.

RISTO VIITALA received the M.Sc. degree in mechanical engineering from the University of Oulu, in 2018. He is preparing his dissertation on rotor dynamics. His work can be applied widely in industries where large rotating machines are used. In the future, he will focus on kinetic energy storages. His research interests include the effects of foundation stiffness on dynamic rotor behavior and bending stiffness variation.

KARI HIEKKANEN received the M.Sc. degree in computer science from the Helsinki University of Technology, in 1993, and the D.Sc. degree in computer science from Aalto University, in 2016. He currently works as a Research Fellow with the Department of Computer Science, Aalto University. Prior joining Aalto University, he has a long experience working in various management and executive roles in IT industry. His current research interests include enterprise systems and industrial internet, including the use of AI in various industrial domains. He lectures a number of courses on enterprise information systems, data strategy, and industrial internet on all academic levels.

RAINE VIITALA was born in 1992. He received the M.Sc. and D.Sc. degrees in mechanical engineering from Aalto University, in 2017 and 2018, respectively. His Ph.D. dissertation was acknowledged with the Aalto University School of Engineering Dissertation Award. In 2018, he visited Technische Universität Darmstadt researching rotating kinetic energy storages, such as flywheels. He has a solid background in experimental large rotor research, including vibration analysis and subcritical vibration, bearing excitations, roundness measurements, and manufacturing for operating conditions. His work can be widely applied in several industrial applications, such as electric motors and generators, turbines, and paper machines. He is involved in developing AI methods to measure rotor systems and rotor governed processes. He is instructing students at all academic levels. In addition, he lectures a bachelor's level course mechatronics basics.