

# Intelligent optical microresonator imaging sensor for early stage classification of dynamical variations

Anton V. Saetchnikov\* Elina A. Tcherniavskaia Vladimir A. Saetchnikov Andreas Ostendorf\*

Anton V. Saetchnikov

Chair of Applied Laser Technologies, Ruhr University Bochum, 44801 Bochum, Germany

Radio Physics Department, Belarusian State University, 220064 Minsk, Belarus

anton.saetchnikov@rub.de

Elina A. Tcherniavskaia

Physics Department, Belarusian State University, 220030 Minsk, Belarus

Vladimir A. Saetchnikov

Radio Physics Department, Belarusian State University, 220064 Minsk, Belarus

Andreas Ostendorf

Chair of Applied Laser Technologies, Ruhr University Bochum, 44801 Bochum, Germany

andreas.ostendorf@rub.de

**Keywords:** *microresonator, whispering gallery modes, sensor, deep learning, long short-term memory network*

Although the machine learning (ML) solutions are conquering science and technology with exponentially growing number of the affected branches, only the few examples have been demonstrated so far to boost the highly sensitive and label-free optical detectors. Nevertheless, ML methods can indeed provide novel strategies for the emerging optical sensing technologies to overcome the challenges of the real-world implementation. On this way the sensing techniques become intelligent self-learning systems capable for predictions of the probed parameters in the automated and self-adjusting manner. Here we report on the upgraded ML engine of the affordable whispering gallery mode sensor operating in the multi-resonator imaging geometry at a fixed frequency. We demonstrate for the first time that such a sensor supplemented by long short-term memory (LSTM) network processing engine enables accurate prediction of the dynamical sensor responses already within the early stages of the observed variations. On example of the experimental data of solutions mixing dynamics we show the possibility for accurate ( $> 95\%$ ) differentiation of the dynamical variations at the level of  $5.5 \times 10^{-4}$  RIU within 6 times shorter period than the whole time series. The impact of the parameters and architecture of the LSTM-based processing engine on the speed and prediction accuracy is analyzed.

## 1 Introduction

A label-free detection method based on excitation of standing waves inside an optical microcavity with circular symmetry is known as the whispering gallery mode (WGM) sensing mechanism [1–3]. Here the light energy can be effectively (with high quality factor of the resonance peaks) stored within a tiny structure thanks to the total internal reflection at the resonator's boundary [4, 5] whereas the sensing approach is based on tracking of the spectral variations caused by the interaction of the confined optical field with the surrounding medium. Depending on the morphology, material, and/or geometry of the microresonators, it is possible to monitor various external properties, e.g., pressure [6], temperature [7, 8], biomolecular composition [9–11], refractive index [12, 13], humidity [14], or gas/vapour type [15, 16]. Last years showed significant progress in the development of the optical microresonator-based sensors in terms of sensitivity enhancement [17, 18] and advancement of the integration level of the microcavities and the optical components for excitation and signal collection [19]. Nevertheless, the vast majority of the WGM sensing solutions still utilizes the original concept of spectrally-resolved data collection whose realization requires costly equipment either on the side of the light source or detector. The paradigm-changing method, that exploits the concurrent imaging of the WGM signal radiated from more than a hundred of microresonators, offers an affordable and efficient alternative, which is especially attractive for the real-world practical applications [20, 21]. The multidimensionality exhibited by the signal of such a sensor and the dispersion of properties of individual microresonators allow to apply the cost- and time-effective fixed frequency interrogation approach, albeit complicates the interpretation of the measured parameters. Here the well-known and typically used analytic models, that describe the response of the

optical field on the external influences via the resonance frequency shift [22, 23] and line broadening [24], are no longer applicable. The reason for this is that the collected signal now contains only the projection of the spectral information onto the single selected frequency. The translation of the sensor signal into the measured parameters is thus addressed by utilizing the deep-learning methods [21].

The deep-learning powered sensors and especially the optical ones establish an emerging research field, where only few rare examples have been reported so far. In addition to the above mentioned solution on imaging of the signal from multiple microresonators at the fixed frequency [21], the other machine learning-based application for WGM sensors utilizes the complete spectrally-resolved information of a single microresonator [25–28]. Further examples of optical sensors supported by the machine learning techniques include surface plasmon resonance method [29] and attenuated total reflectance approach [30].

The reported intelligent WGM sensor [21] has been proven highly efficient for parameters prediction under the equilibrium condition when a steady state of the sensed medium around the optical sensor is reached. The corresponding response and thus waiting time, though, may vary here from several milliseconds up to several hours, depending on the type and value of the measured parameter. As a characteristic example one can point out the affinity biosensors dedicated to the detection of the type and/or concentration of biomolecular components [31–33], e.g., proteins, viruses, bacteria, DNA, or RNA. In this case the sensing response is governed by the binding of the particles with the ligand preliminary immobilized on the resonator’s surface so that the description of the measured parameter within only the steady state appears not comprehensive. In fact, the complete response curve, referred to also as a sensorgram providing the shift of the resonance frequency is monitored over time, has to be considered. The measured dynamical response in terms of spectral shift of the resonance frequency can be then analyzed using the Langmuir adsorption model. The latter is defined by the association and dissociation rates of the molecules together with their concentrations [34–36].

The model-based description option is though not available for the signal collected from multiple microresonators at a fixed frequency. Instead, the dynamics analysis can be addressed by extending the previously reported [21] deep-learning engine to learn from the time series data through the use of the recurrent neural networks (RNN) and their particular type - long short-term memory network (LSTM). The number of the reported examples of the LSTM networks implementation for processing of the dynamical variations covers predominantly the problems of handwriting, speech recognition, and machine translation. At the same time, several examples of spectral data processing applications, e.g., classification of the time series spectral images [37] and mass spectrometry data [38] have been already reported. The present work reports on a novel strategy based on the long short-term memory network training for quantification of the dynamical responses within the optical sensing and is demonstrated for multi-microcavity imaging optical sensor operating at the fixed frequency. The performance of the proposed approach together with the impact of the training dataset and complexity of the deep learning engine for accuracy of the external parameters prediction has been studied. Benefits for early decision making without necessity for analysis of the whole dynamics till the steady state have been discussed.

## 2 Methods

### 2.1 Data collection

In order to match the variations of the radiated intensities with the particular microresonators on the chip, a map of the microcavities is established. For this the multistep image processing approach for edges detection has been applied as described in [39]. The radiated light observed within each microcavity area is then perceived as the microcavity signal. The intensity of the signal for each microresonator is described by the location of the fixed laser frequency in the WGM spectrum (**Figure 1**). Within a particular sensing event a set of intensities with degrees of freedom equal to the number of resonators (of the order of several hundreds) is produced.

The most frequent scenarios for the WGM-based sensing are the monitoring of the bulk refractive index and of the biochemical molecules adsorption. In both cases an overall value of the resonance frequency

shift is primarily analyzed, where, though, the dynamical variations between the steady states might be also of particular importance for complete description of the measured entity. For the molecules adsorption process the sensorgram follows a differential rate equation described by the Langmuir model, where the actual concentration impacts the shift and the kinetic rate. For the bulk refractive index sensing the period between the steady states for different RIUs is mainly described by the mixing of the sensed medium inside the flow chamber with the incoming fluids. The larger the difference between the previous and the subsequent concentrations is, the longer is the period required to reach the final resonance frequency position. Still, a more complex behavior of the dynamical variations of the spectral shift can be observed for microresonators with functional layers or with structural response enhancement [11]. In case of the fixed frequency sensing the response of each resonator is non-linearly modulated by the WGM spectrum that is unique for each microcavity (Figure 1). This results in both the non-linear character of the captured intensity variations between the steady states, as it was discussed in [21], and in the complex nature of the dynamical variations between them. As a result, depending on the physical and/or chemical nature of the process to be sensed and the time scale of the variations produced by the external process, the collected signals differ greatly.

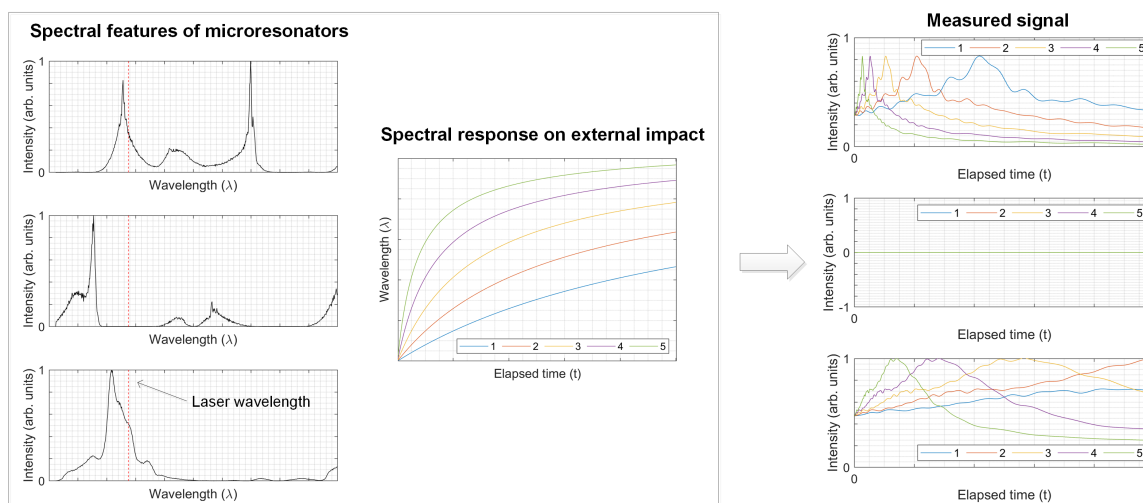


Figure 1: Graphical representation of the nature of the intensity-encoded microresonator signal measured at the fixed frequency in the multiresonator imaging scheme. The dynamical changes of the spectral shift numerically simulated according to the Langmuir model (single response shown for simplicity, in fact - individual for each microresonator) are translated into the microcavity-specific temporal variations of the radiated intensity at the selected illumination wavelength.

## 2.2 Preprocessing strategy

The preprocessing of the dynamical variations in the fixed frequency illumination scheme is proposed as following. Initially the set of the radiated intensities (number of observations -  $o_i$ ) for each particular resonator ( $r$ ) has been scaled to the  $[0;1]$  range. Then the dataset has been re-sampled with the frequency of 1 Hz,  $o_r$  (mean frame rate of the camera is 60 Hz). The sampling frequency has been selected with respect to the average duration required to observe the changes in the measured temporal variations beyond the two standard deviations of the signal for the water environment. Later  $n$  different groups representing different external impacts including  $m$  repetitions are produced. Due to the limited amount of the repetitions that can be produced within an experiment a set of the artificial data is generated for each group (output state) according to the normal distribution with the mean value and the standard deviation calculated at each time point of the experimental data. The whole amount of the repetitions for each group is set to 200. The above-mentioned steps are schematically depicted in **Figure 2**.

The demand on the computational resources is expected to rise when switching from the classification task to the regression one by increasing the number of the groups representing various dynamical variations. In order to keep the training and testing time within the reasonable limits the data preprocessing

steps could be additionally supplemented by the translation of the measured intensities variations over time into the series of the characteristic points by fitting with parametric curves. In this case, however, one should account for several substantial drawbacks: first, the temporal resolution of the sensing data is thus reduced; second, the efficiency for such conversion is expected to be low providing the captured variations of the intensities radiated by the microresonators possess rather simple dynamics with minimum inflection points over the sensing period.

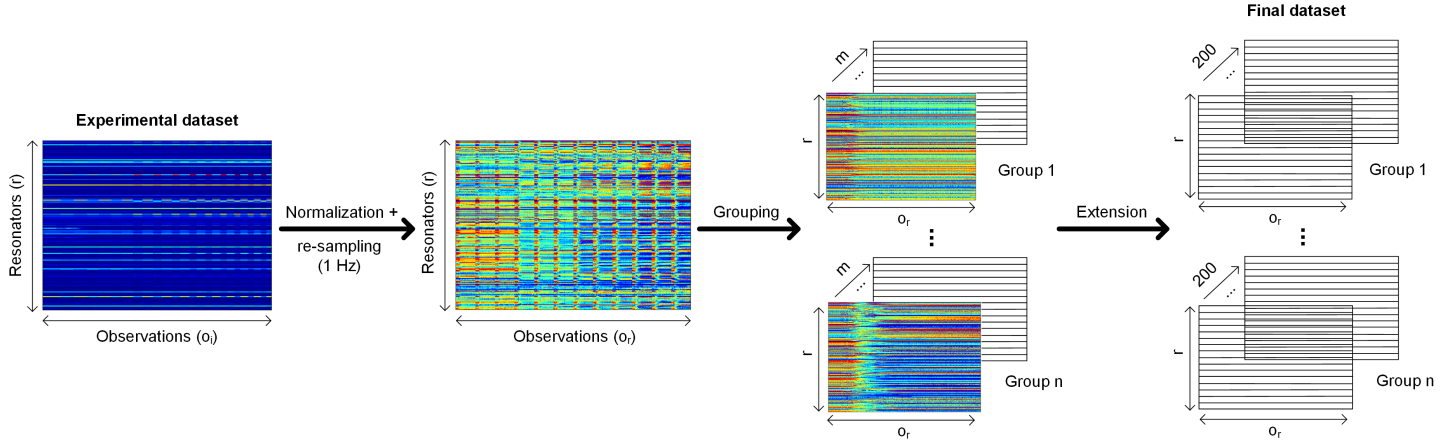


Figure 2: Schematic illustration of the experimental dataset preprocessing: initial normalization, equalization of the sampling frequency among the sensing phases, grouping of the data according to the external impact and, finally, dataset extension.

During the analysis of the experimental dataset on variations of the radiated intensities there is a number of microcavities observed that demonstrate no variations in the collected signal while the parameters of the sensed medium are changed. Since the measured signal is the function of the resonance spectrum of the microcavity where the resonance peak is narrowly allocated, the illumination frequency for mentioned resonators falls into the spectrum part with no resonance. The absence of the resonance behavior in general or resonances with low quality factors that appear for corrupted microcavities are also the reason for observing insignificant or no temporal variations in the signal. In order to reduce dimensionality of the data and to extract the most significant features (microcavities), they have been rearranged according to the ReliefF algorithm [40, 41] with a modification to handle the temporally resolved data proposed in [42]. This method has been chosen as the most beneficial in terms of the algorithm efficiency for features selection and tested on various types of synthetic data [43].

The ReliefF algorithm is an iterative method to estimate the feature weights based on the variation of their value for different output states. The weights of the features are initialized with zeros. At each iteration of the algorithm a random instance ( $R$ ) is selected. For the considered type of data this is a single sample of the dynamical response. Then two different sets – one with  $k$  nearest neighbors (here  $k = 20$ ) representing the same class as the instance  $R$  and another one with  $k$  nearest neighbors from a different class – are formed. The neighbors of the same class are called hits ( $H$ ) and the ones from the different class are referred to as misses ( $M$ ). The weight of the feature is updated by subtraction of the average distance to the hits and addition of the average distance to the misses. In the original algorithm the instance has no dynamical feature and thus the Euclidean distance can act as the measure. In order to handle the time series that describes the instance, a method of the dynamical time warping (DTW) is selected for measuring the similarity between the dynamical variations of  $R$ ,  $H$ , and  $M$ . A single iteration of the weight update is repeated several times (selected to be equal to the number of the time series in the training data). Finally, the features are sorted in the descent manner according to their weights.

### 2.3 Background on RNN and LSTM

RNN is a class of the neural networks that are supplemented by the loops in the connections compared to the conventional feedforward networks [44]. These loops in the structure enable processing of the in-

put sequences by introduction of a recurrent hidden value whose state at each time point of the sequence is calculated considering its state at the previous time point. Depending on the task, the RNN architecture can be designed for different types of the input and output values.

LSTM networks are based on recurrent connections of the LSTM cells as well [45]. However, unlike the conventional RNN, the LSTM network is specially proposed to learn the long-term dependencies in the input sequences and thus allows to overcome the main problems of the RNNs: vanishing and exploding gradients when the long sequences are used for training. This endows the network with less sensitivity to the time gaps in the data sequence. In addition to the hidden value state, at each time point of the recurrent structure the LSTM block delivers a cell state. The mechanism for calculation of the outputs of the LSTM cell is based on the input ( $i$ ), output ( $o$ ), and forget ( $f$ ) gates. The input gate controls the level for the cell state to be updated with the input value, the forget gate - the level to pass the previous hidden state through the cell, and the output - the level to transfer the cell state to the next LSTM cell.

### 3 Results and discussion

In this section the performance of the LSTM network for prediction of the sensed parameters characterized by the dynamical response of the WGM signal is analyzed. The algorithms for the LSTM network construction have been implemented in Matlab 2021a using the Deep Learning Toolbox, Statistics and Machine Learning Toolbox. The local workstation with GPU Nvidia Quadro RTX 6000 has been utilized for training and testing the LSTM network. Within this research two studies on application of the LSTM network in terms of its prediction accuracy were performed. The first one was intended to verify the impact of the amount of the features (microresonators) in the dataset used for training and testing. The second study is focused on the optimization of the LSTM network architecture. The experimental dataset including the dynamical changes in the collected signal of the WGM imaging sensor in the fixed frequency illumination scheme has been produced on example of small temporal gradient of the refractive index value in the chamber. This has been realized by mixing the initial medium in the fluidic chamber (water) at the constant pumping speed with the incoming liquids of different refractive indexes.

#### 3.1 Dataset description

Time series data has been produced with five different concentrations of ethanol from 1% to 5% with a step of 1%. In terms of the refractive index change this corresponds to  $5.5 \times 10^{-4}$  to  $2.7 \times 10^{-3}$  relative to the water refractive index. The measurement for every concentration has been repeated four times with 2000 seconds in each phase. After each measurement the flow chamber has been flushed with the deionized water for 2000 seconds in order to ensure the pure water environment around the sensor. Hereby the whole duration of the experimental run lasts for 80000 seconds that taking into account the 60 Hz frame rate of the camera corresponds to 4.8 million of observations. The collected signal, which includes the intensities radiated from 181 microresonators, has been grouped into 5 data sets per 2000 seconds so that each of the sets represents a different temporal gradient of the refractive index in the flow chamber. The obtained variations of the radiated light for six representative microcavities on the imaging sensor are demonstrated in **Figure 3**.

The graphs show the time period of 2000 seconds where the slow mixing of the pure water inside the flow cell with five different ethanol solutions (1%, 2%, 3%, 4%, and 5%) happens. The whole time period can be split into three parts. The first one lasts up to approx. 250 seconds from the beginning (varies between the resonators due to different position relative to the input channel), when the solution is transported from the valve to the sensing chamber. Within this period the signal remains constant and corresponds to the pure water environment. The second is the period of mixing the present solution with the incoming one. Here the uniqueness of the response among the microresonators as well as between the different concentrations is clearly observed. Each of the resultant responses is a combination of two non-linear functions where the first one is common for all microcavities and describes the temporal gradient of the bulk refractive index inside the chamber; and the second function describes the individual

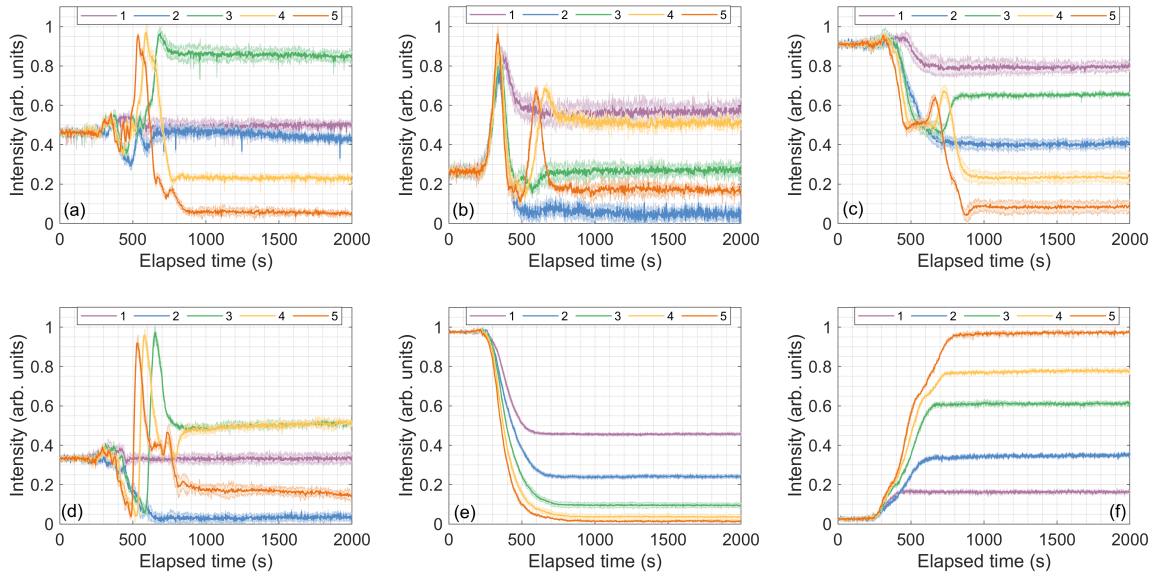


Figure 3: Overview of the measured temporal variations of the radiated light intensities for six ((a)-(f): #1-6) representative microresonators of the sensor (the data has been scaled to  $[0;1]$ ). The curves on the graphs belong to the different bulk refractive index states (numbers correspond to the percentage of the pure ethanol in the water). Each of the measurement has been repeated four times (lines follow the mean values and the shaded areas show 95% confidence intervals).

WGM spectrum of the microcavities. This can be clearly observed for the microresonators #1-4 in Figure 3 where the temporal variations of the radiated intensity exhibit the signs of the resonance peaks of the microcavities. Here the amount of these peaks and their linewidth depend on the filling medium concentration, where higher values of concentration lead to appearance of more spectral features and to faster variation rate. In contrast to microcavities #1-4, resonators #5-6 show less complex behavior in the variations that is explained by almost linear part on the resonance curve where the refractive index changes appear: falling (#5) and rising (#6) part of the resonance peak. The third period starting from approx. 800 seconds represents a steady state for the chosen ethanol solution.

The repeatability between the identical sensing events (captured with the time delay up to several hours) varies between 0.01 (#6) and 0.1 (#2) arb. units for the steady state. Dispersion of the repeatability value between different resonators arises from the different non-linear response of each microresonator collected at the fixed wavelength of the illumination source. Here the same repeatability in terms of spectral shift is translated into different variations of the intensity signal depending on the shape of the spectral curve and location of the illumination wavelength on this curve. At the same time, considering several hundreds of microresonators simultaneously the overall repeatability of the sensor approaches the lowest value among them. This is tuned by the distribution of the weights among the features during the process of the NN training. The stability of the signal at the steady state (third period) for the vast majority of the microcavities on the sample demonstrates the relative drift below 1%. The particular cases when the relative drift increases such as observed for microcavities #1 and #4 are expected to be caused by the incomplete enveloping of the resonator with liquid.

As it was previously reported [21], the RIU states can be accurately predicted by the deep-learning engine that is pre-trained on the data from the third period. However, for the represented dataset at least 550 seconds are required to wait during the experiment until steady state is reached. So the deep-learning engine is intended to learn from the time series data from the second period in order to enable the prediction of the external impact without measuring the signal till the equilibrium.

### 3.2 LSTM network architecture

The measured experimental data represents the sequence input to single output classification problem, where the temporal variations of the radiated intensities correspond to one particular concentration of the incoming agent. The LSTM network structure has been constructed out of the following blocks (**Figure 4**):



sequence input layer, bidirectional LSTM (BiLSTM) layer, dropout layer with 0.2 probability, fully connected layer for 5 output classes, a softmax layer to compute the probability of each dynamics to belong to a certain class, and a classification output layer. Compared to the conventional LSTM layer, the bidirectional one processes the sequence data in the forward and backward directions in two separate hidden layers. Dropout layer is intended to avoid the overfitting via dropping the random neuron from the network during the training with the given probability. Adam optimizer including L2 regularization factor has been selected for the network training with the gradient threshold of 1 and initial learning rate of 0.01. The amount of the training epochs for LSTM network is controlled by both the maximum number of epochs (here equal to 100) and by the validation patience parameter (here: 5). The latter is introduced to avoid the overfitting as well, and describes the number of times when the loss on the validation set is larger or equal to the previously smallest loss before the training stops. If the loss parameter is continuously decreasing during the training process, it stops when the maximum number of the epochs is reached. The performance of the LSTM network for classification of the sensed mixture concentrations is finally evaluated by the correct classification rate.

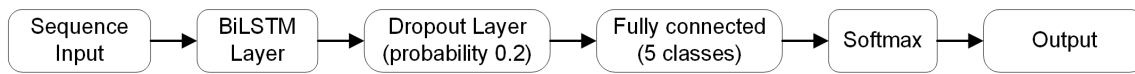


Figure 4: Architecture of the LSTM network used for classification of the dynamical changes sequences of the microresonator imaging sensor.

The preprocessed experimental dataset with the reorganized position of the features according to their weights descending (ReliefF) has been split into 10 subsets with different number of features [10:10:100]. Later the sets have been further divided into the subsets of different sequence duration (from 100 to 400 s with a step of 10 s) with the starting point at 150 seconds (see Figure 3). The above-mentioned operations are intended to perform the studies on the impact of the number of the most significant features (microresonators) and of the duration of the sequence response of the multiple microresonator imaging sensor on the prediction accuracy. For each hyperparameter tuning operation the input data has been split into training (70%), validation (15%), and test (15%) parts where the temporal sequences have been randomly selected from the whole dataset. For generalization every LSTM network training procedure has been repeated 100 times.

### 3.3 Features selection

For the first test intended to determine the optimal number of the features (microresonators) in the dataset, the amount of the hidden LSTM units has been fixed at 5. The evolution of the correct prediction rate at different amounts of significant features (10, 20, 30, 50, 70, 100) in the input data is represented in **Figure 5**. Each plot reveals the calculated classification errors together with their mean values over 100 training repetitions for different durations of the input sequence. In addition to them, the exponential functions ( $y = ae^{1-bx} + c$ ) that are fit to the mean error values over the time are represented. Except for the results for 10 and 20 most significant features, the selected function accurately follows the mean accuracy values for the represented durations of the sequences and approaches 100% accuracy for the longer ones. For the case of 10 and 20 features one can observe the reduction of the mean prediction accuracy when comparing longer input sequences with the shorter ones (e.g., 160 s with 140 s and 180 s with 160 s). Due to the moderate number of features together with rather short duration of the input sequences, these results are expected to be related with the dominant impact of the short-term local variations of the collected intensities on the LSTM network parameters. Reduction of the prediction accuracy when the number of the features increases for the same sequence duration confirms this interpretation. Here the impact of the long-term dependencies in the dataset on the LSTM network output rises. The prediction accuracy for longer durations of the training data follows the expected behavior and the 95% rate level shifts from 260 s to 240 s when 20 most weighted features are considered. The subsequent increase in the amount of the significant features up to 30 enables to slightly shorten the sequence duration to ensure the same accuracy. Here the prediction accuracy is continuously increasing

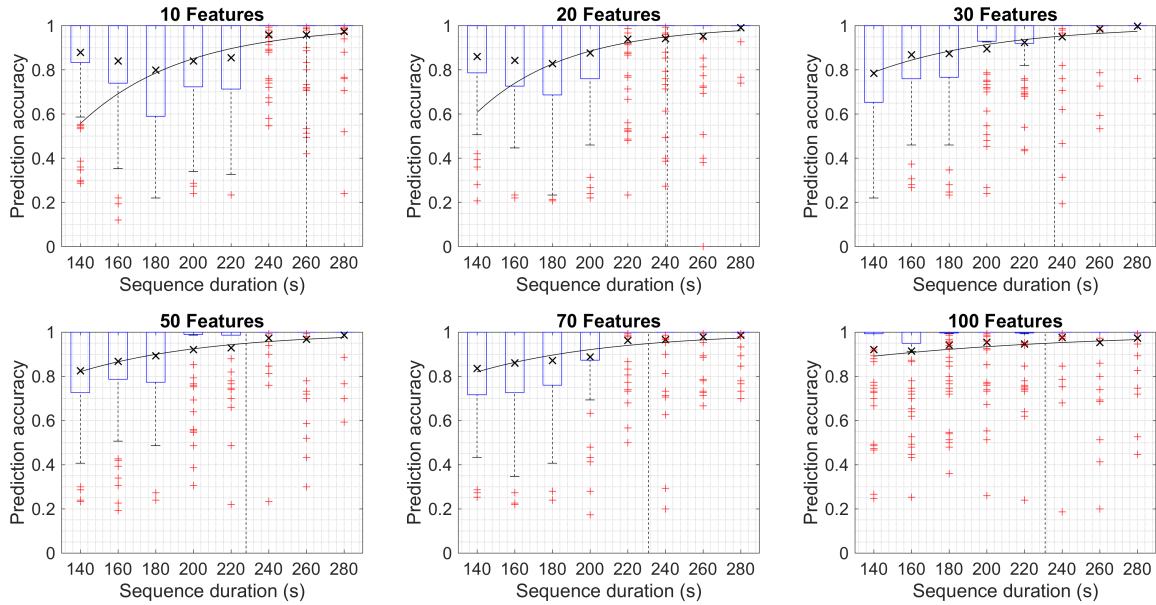


Figure 5: Statistics on the prediction accuracy of the proposed LSTM-based engine for different amounts of significant features (10, 20, 30, 50, 70, 100) in the input data (black crosses mark mean values, the bottom and top edges of the blue box indicate the 25th and 75th percentiles).

when longer sequences have been used for training. For the dataset with 50 features the 95% level is observed at  $\approx 230$  s and variability among the repetitions is reduced. When considering more features up to 70 and 100, the 95% accuracy level slightly shifts towards longer durations. Among them, the results for 100 significant features stands particularly out, where the improvement of the mean prediction error for shorter input sequence durations (up to 200 s) is observed. After a certain number of significant features, the portion of those with low variability and thus with less distinguishable dynamical changes between the output states grows up. When those features become majority, the more weight during the training of the neural network is redistributed to the short-term memory representing the last local variations in the radiated microcavity signal. In fact, the result for 100 features represents exactly the case when the LSTM network starts to follow predominately the local variations. Reduced prediction accuracy for 70 features compared to the case with 50 features agrees with this explanation as well. Here the dataset has been supplemented with the microresonators with low variability compared to the case of 50 features. However, the amount of such resonators remains minor compared to those with clear distinction in the dynamics.

The above results reveal that the optimum amount of features in the training set exists. This allows to obtain the high prediction rate for the minimal duration of the measured temporal variations in the WGM data obtained at the fixed frequency. The generalized strategy for definition of the required number of significant features for training of the LSTM network can be formulated as the necessity to pick all the microcavities that demonstrate the clearly distinguishable dynamical changes for different external conditions. Hereby, the number of features in the reported experimental set can be reduced from 181 to the first 50 significant ones.

### 3.4 LSTM layer complexity optimization

The second study is indented to determine the optimal number of the hidden LSTM units in order to ensure the low error rate classification of the sequences within the minimal time. The results on classification of the sequences dataset with 50 significant features of the varying duration (from 140 s up to 280 s) with LSTM networks of different complexity of their architecture are demonstrated in **Figure 6**. The obtained results demonstrate the clear correlation between the number of the hidden units and the prediction accuracy. It has been determined that an increase in the number of the hidden units in the LSTM network from 5 to 10 is followed by the shift of the 95% mean detection rate level from 230 s of



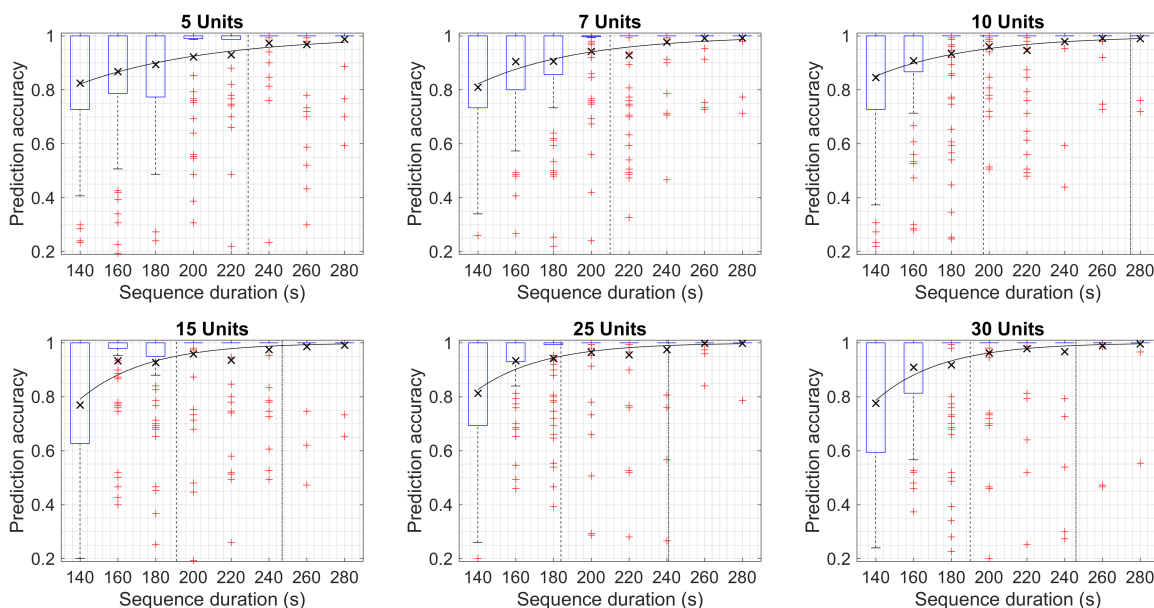


Figure 6: Statistics on the prediction accuracy for the input series data of different duration with 50 significant features provided by the LSTM networks of the varying architecture complexity (5, 7, 10, 15, 25, and 30 hidden units). The black cross points indicate the mean values of the prediction accuracy over the training repetitions, the solid lines are the fitting curves over the mean values. Dashed vertical line indicates the duration of the time series to reach the accuracy level of 95% and the dash-dotted vertical line - 99%.

the required sequence duration down to approx. 200 s. Among these architectures, the 99% detection rate in case of 10 hidden units is observed for the input sequence duration of approx. 275 s. The results for more complex network architectures with 15, 25, and 30 hidden units exhibit the minor variability in the required sequence duration to reach the 95% and 99% accuracy. The first level of the correct detection rate is reached for the dynamical variations duration of  $\sim 190$  s and the second - at  $\sim 250$  s. Hereby 15 hidden units in the architecture of the LSTM network are determined as sufficient to meet the best prediction accuracy at the minimum duration of the training sequence.

When analyzing the error rates for individual output states, the difference among them in the required time to reach the desired accuracy rate is examined. Here the shortest sequence length to reach the 95% level equal to  $\sim 150$  s has been observed for the first output state (1% ethanol) and the longest of  $\sim 250$  s - for the forth (4% ethanol). This is predominantly caused by the nature of the experiment (Figure 3) where lower gradient between the incoming and the present liquids in the sensing chamber results in earlier steady state.

Since the experimental sequence part before 150 s has been excluded from the LSTM training process, this value has to be added to the actual sequences duration. At the same time, as discussed in Sec. 3.1 and demonstrated in Figure 3, the first 250 s are required to transport the agent to the sensing chamber and have to be excluded. Thus the measured time series dynamical variations can be accurately classified within the first 90 s after the start of the reaction in the chamber with the 95% correctness. The prediction rate of 99% is possible already after 150 s. Hereby, the accurate prediction of the external impact that demonstrates a set of dynamical variations of the intensities radiated by the microresonators is possible approx. 6 times faster than the timescale (550 s) required to observe the steady state in the microresonator sensor signal. The averaged training time for this configuration of the input sequence and LSTM network is measured to be 12 s and the testing time as 0.08 s.

## 4 Conclusion

This paper demonstrates the first example of application of the recurrent neural networks to build an intelligent whispering gallery mode sensor capable for quantification of the external impacts on the time

series dynamics of the sensor response. This has been realized by combining the long short-term memory network together with the affordable instrumental approach based on the multiple microcavity imaging sensor illuminated by the diode laser at the fixed frequency. The reported method for parameters quantification of the external impact that is characterized by the dynamical changes in the measured signal is a generalized intelligent WGM sensing solution with respect to the feedforward network-based solution [21]. In terms of the LSTM-based approach the latter is an extreme case where a set of the microresonator signals under the equilibrium condition has been utilized for prediction.

Under optimization of the number of the microresonator signals to consider for the LSTM network training together with the complexity of its architecture we demonstrated the discrimination of the set of complex non-linear highly specific time series signals of the intensities radiated by the microcavities within the substantially reduced timescale than required to reach the steady state. The exact portion of the time series data required for precise estimation of the external parameters is expected to depend on the dynamics of the measured response, sensor spectral features, and the difference in the parameters to be predicted. At the same time, unlike for the common study of spectral variations of the microresonators, here the presence of the additional subpeaks in the initial WGM spectrum is beneficial since it provides more variability (as being modulated by the non-linear function) between the measured temporal variations. On the reported in this paper example of five different dynamical variations provided by the temporal gradient of the refractive indexes with relative difference of  $5.5 \times 10^{-4}$  RIU due to slow mixing of the solutions inside the flow chamber, a decision on the concentration of the incoming analyte with  $> 95\%$  accuracy is possible 6 times faster than the steady state is observed.

The reported WGM-based sensing device supplemented by the LSTM network processing engine is a highly promising solution within the emerging field of the machine-learning inspired sensing devices and fulfills the main critical requirements such as affordability, reduced instrument complexity, and fast response time. The discussed mechanism for accurate prediction of the external impact utilizing the part of the captured set of the temporal variations in the multiple microresonator imaging sensor is not limited to the demonstrated in this paper experimental data of slowly changing bulk refractive index. The proposed LSTM network-based approach can be directly implemented also for other applications, e.g., for prediction of the protein concentrations. The latter is possible due to the similar nature of the collected signal for the adsorption of the molecules on the resonator surfaces in the fixed frequency illumination scheme [20].

In general, the collected signal in the fixed frequency illumination scheme includes the whole palette of the WGM spectrum variations (spectral shift, linewidth broadening, and mode splitting) that may arise in the optical microresonator-based sensing. Thus the proposed method for the dynamical changes prediction is applicable to the whole variety of the possible sensing tasks where the training data can be collected in automated manner, where, depending on the dispersion and sampling rate of the sensed parameter in the experimental dataset used for the LSTM network training, the dynamical changes prediction problem can transform from the classification to the regression one. Albeit with the reduced efficiency due to the small variability of the dynamical changes among the microresonators, the LSTM-based approach could be employed also with the conventional frequency sweeping scheme. In addition, the method is expandable for other than WGM types of sensors which are capable to demonstrate the reproducible dynamical response under the same conditions. Among them the most promising might be classification of the biomolecules concentrations for the affinity-based sensors in different configurations, e.g., electrochemical, surface plasmon resonance, attenuated total reflectance, etc.

## 5 Experimental section

*Multiple microresonator sensor:* Spherical microcavities (Cospheric LLC) with varying diameter in the range from 90 to 120  $\mu\text{m}$  out of the soda-lime glass are allocated on the 150  $\mu\text{m}$  thick 20×20 mm cover glass substrate with  $\approx 500$  nm water-matched adhesive layer (MyPolymers MY-133-MC). The latter ensures both the stable coupling conditions for excitation of the whispering gallery modes and the minimal energy dissipation for the water-based environment. No special arrangement procedures are required

to allocate the microcavities. The detailed description of the optical microresonator-based sensor fabrication has been previously reported in [7, 20, 21]. In order to enable the WGM detection, each of the fabricated chips is enclosed in the two-component structure out of polyether ether ketone with the polydimethylsiloxane (Dow Corning Sylgard 184) layers as gaskets. When assembled together the structure forms a flow chamber whose dimensions have been numerically optimized in the CFD Module of the Comsol Multiphysics to ensure the homogeneous distribution of the fluid velocities within the whole sensor. The viewing window of 10×10 mm covered with the glass plate enables the imaging of the microresonators. The loaded quality factors of the individual microspheres on a single chip vary in the range of  $10^3 - 10^6$  and the sensitivity for the bulk refractive index ranges from 4 nm/RIU to 20 nm/RIU [7, 20, 21]. These values are common for the glass microresonators of the selected dimensions operating in the spectral range around 785 nm and agree well with the numerical estimations for the fundamental mode. Variability of both the quality factors and sensitivity arises from the dispersion of the microresonators properties, difference in the excitation efficiency, and the order of the excited resonance mode among the microresonators on the sample.

**Instrument:** The BK-7 optical prism with the anti-reflection coating on the legs has been selected for the robust excitation of the multiple microresonators in a parallel manner. The fiber-coupled fixed frequency diode laser (Thorlabs LP785-SAV50) with central wavelength of 785 nm and 10 MHz linewidth has been supplemented with a polarization controller (Thorlabs FPC030) whereupon the collimated beam is guided to the prism (**Figure 7**).

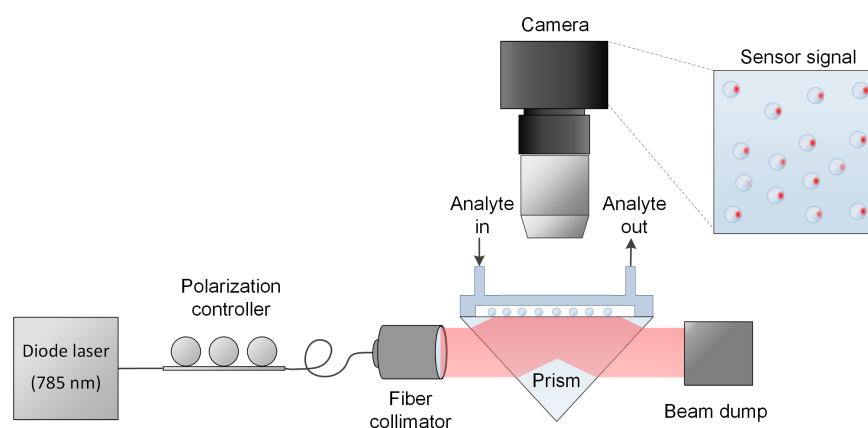


Figure 7: Instrument configuration for the radiated intensity collection in the multiple microresonator imaging scheme at the fixed illumination frequency.

The incidence angle and polarization state of the laser light have been selected to excite the first order TE mode (numerically estimated for the average microresonator size of 105  $\mu\text{m}$ ). The pressure-based controller (Fluigent LINEUP FLOW EZ) supplemented with the flow rate sensor (Fluigent FLOW UNIT) and the selection valve (Fluigent M-SWITCH) have been utilized for sequential pumping of different fluids through the sensor with constant pumping speed of 100  $\mu\text{l}/\text{min}$ . The variations of the radiated energies due to the changes of the properties of the resonators and/or external medium are collected by the objective and captured with the camera.

## Acknowledgements

The authors thank the German Federal Ministry of Education and Research (BMBF) for financial support of this project (FKZ: 03VPO8220, IntelLOSS).

## Conflict of Interest

The authors declare no conflict of interest.

## References

- [1] M. R. Foreman, J. D. Swaim, F. Vollmer, *Advances in optics and photonics* **2015**, 7, 2 168.
- [2] Y.-N. Zhang, T. Zhou, B. Han, A. Zhang, Y. Zhao, *Nanoscale* **2018**, 10, 29 13832.
- [3] X. Jiang, A. J. Qavi, S. H. Huang, L. Yang, *Matter* **2020**, 3, 2 371.
- [4] V. B. Braginsky, M. L. Gorodetsky, V. S. Ilchenko, *Physics Letters A* **1989**, 137, 7-8 393.
- [5] K. J. Vahala, *Nature* **2003**, 424, 6950 839.
- [6] A. Bianchetti, A. Federico, S. Vincent, S. Subramanian, F. Vollmer, *Optics Communications* **2017**, 394, 6950 152.
- [7] A. V. Saetchnikov, E. A. Tcherniavskaia, V. A. Saetchnikov, A. Ostendorf, *IEEE Journal on Miniaturization for Air and Space Systems* **2021**, 2, 2 84.
- [8] J. Liao, L. Yang, *Light: Science & Applications* **2021**, 10, 1 32.
- [9] J. Su, A. F. Goldberg, B. M. Stoltz, *Light: Science & Applications* **2016**, 5, 1 e16001.
- [10] S. F. Wondimu, M. Hippler, C. Hussal, A. Hofmann, S. Krämmmer, J. Lahann, H. Kalt, W. Freude, C. Koos, *Optics express* **2018**, 26, 3 3161.
- [11] A. V. Saetchnikov, E. A. Tcherniavskaia, V. A. Saetchnikov, A. Ostendorf, *Journal of Lightwave Technology* **2020**, 38, 8 2530.
- [12] I. M. White, H. Zhu, J. D. Suter, N. M. Hanumegowda, H. Oveys, M. Zourob, X. Fan, *IEEE Sensors Journal* **2007**, 7, 1 28.
- [13] H. Li, X. Fan, *Applied Physics Letters* **2010**, 97, 1 011105.
- [14] M. Eryürek, Z. Tasdemir, Y. Karadag, S. Anand, N. Kilinc, B. E. Alaca, A. Kiraz, *Sensors and Actuators B: Chemical* **2017**, 242, 10 1115.
- [15] A. K. Mallik, G. Farrell, M. Ramakrishnan, V. Kavungal, D. Liu, Q. Wu, Y. Semenova, *Optics express* **2018**, 26, 24 31829.
- [16] C. Lemieux-Leduc, R. Guertin, M.-A. Bianki, Y.-A. Peter, *Optics express* **2021**, 29, 6 8685.
- [17] V. R. Dantham, S. Holler, C. Barbre, D. Keng, V. Kolchenko, S. Arnold, *Nano letters* **2013**, 13, 7 3347.
- [18] Y. Zhi, X.-C. Yu, Q. Gong, L. Yang, Y.-F. Xiao, *Advanced materials (Deerfield Beach, Fla.)* **2017**, 29, 12.
- [19] Z. Yao, K. Wu, B. X. Tan, J. Wang, Y. Li, Y. Zhang, A. W. Poon, *IEEE Journal of Selected Topics in Quantum Electronics* **2018**, 24, 6 5900324.
- [20] A. V. Saetchnikov, E. A. Tcherniavskaia, V. V. Skakun, V. A. Saetchnikov, A. Ostendorf, *IEEE Sensors Journal* **2019**, 19, 17 7644.
- [21] A. Saetchnikov, E. Tcherniavskaia, V. Saetchnikov, A. Ostendorf, *Opto-Electronic Advances* **2020**, 3, 11 200048.
- [22] S. Arnold, M. Khoshshima, I. Teraoka, S. Holler, F. Vollmer, *Optics Letters* **2003**, 28, 4 272.
- [23] F. Vollmer, S. Arnold, D. Keng, *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105, 52 20701.
- [24] M. D. Baaske, M. R. Foreman, F. Vollmer, *Nature nanotechnology* **2014**, 9, 11 933.

- [25] V. A. Saetchnikov, E. A. Tcherniavskaia, A. V. Saetchnikov, G. Schweiger, A. Ostendorf, *Nonlinear Phenomena in Complex Systems* **2011**, 14, 3 253.
- [26] E. A. Tcherniavskaia, V. A. Saetchnikov, *Journal of Applied Spectroscopy* **2011**, 78, 3 457.
- [27] D. Hu, C.-L. Zou, H. Ren, J. Lu, Z. Le, Y. Qin, S. Guo, C. Dong, W. Hu, *Sensors (Basel, Switzerland)* **2020**, 20, 3.
- [28] J. Lu, R. Niu, S. Wan, C.-H. Dong, Z. Le, Y. Qin, Y. Hu, W. Hu, C.-L. Zou, H. Ren, *IEEE Sensors Journal* **2021**, 21, 7 9046.
- [29] J. C. M. Gomes, L. C. Souza, L. C. Oliveira, *Biosensors & bioelectronics* **2021**, 172 112760.
- [30] V. V. Kornienko, I. A. Nechepurenko, P. N. Tananaev, E. D. Chubchev, A. S. Baburin, V. V. Echeistov, A. V. Zverev, I. I. Novoselov, I. A. Kruglov, I. A. Rodionov, A. V. Baryshev, A. V. Dorofeenko, *IEEE Sensors Journal* **2020**, 20, 13 6954.
- [31] G. C. Righini, S. Soria, *Sensors (Basel, Switzerland)* **2016**, 16, 6.
- [32] Y. Wang, S. Zeng, G. Humbert, H.-P. Ho, *Laser & Photonics Reviews* **2020**, 14, 12 2000135.
- [33] N. Toropov, G. Cabello, M. P. Serrano, R. R. Gutha, M. Rafti, F. Vollmer, *Light: Science & Applications* **2021**, 10, 1 42.
- [34] I. Langmuir, *Journal of the American Chemical Society* **1918**, 40, 9 1361.
- [35] A. L. Washburn, L. C. Gunn, R. C. Bailey, *Analytical chemistry* **2009**, 81, 22 9499.
- [36] Y. Kim, H. Lee, *Analytical chemistry* **2020**, 92, 9 6189.
- [37] J.-L. Xu, S. Hugelier, H. Zhu, A. A. Gowen, *Analytica chimica acta* **2021**, 1143 9.
- [38] J. Liu, J. Zhang, Y. Luo, S. Yang, J. Wang, Q. Fu, *IEEE Access* **2019**, 7 10734.
- [39] A. V. Saetchnikov, E. A. Tcherniavskaia, V. A. Saetchnikov, A. Ostendorf, In C. Gorecki, A. K. Asundi, W. Osten, editors, *Optical Micro- and Nanometrology VII*. SPIE, ISBN 9781510618824, **2018** 33.
- [40] K. Kira, L. A. Rendell, In D. Sleeman, P. Edwards, editors, *Machine Learning Proceedings 1992*, 249–256. Morgan Kaufmann, San Francisco (CA), ISBN 978-1-55860-247-2, **1992**.
- [41] I. Kononenko, E. Šimec, M. Robnik-Šikonja, *Applied Intelligence* **1997**, 7, 1 39.
- [42] M. D. Radovic, M. Jordanski, N. Filipović, Z. Obradovic, **2016** .
- [43] V. Bolón-Canedo, N. Sánchez-Marño, A. Alonso-Betanzos, *Knowledge and Information Systems* **2013**, 34, 3 483.
- [44] R. J. Williams, D. Zipser, *Neural Computation* **1989**, 1, 2 270.
- [45] S. Hochreiter, J. Schmidhuber, *Neural Computation* **1997**, 9, 8 1735.