The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5-–9 June 2022, Oslo, Norway

CLASSIFICATION OF VIBRATION-COMPROMISED DOFS DATA WITH LSTM NEURAL NETWORKS

VALERIA USENCO[1], KASPAR LASN[1]

[1] Department of Mechanical and Industrial Engineering
Richard Birkelands vei 2B
7491 Trondheim, Norway
email: valeriau@stud.ntnu.no, kaspar.lasn@ntnu.no

Key words: Distributed optical fiber sensors, OFDR, ANN, LSTM, Classification

Abstract. Distributed optical fiber sensors (DOFS) are gaining momentum for in-situ condition monitoring and damage detection purposes. Although DOFS are a versatile sensing method enabling high-resolution strain and temperature mapping, they are also sensitive to mechanical vibrations. Vibrations are typically created by the ambient environment (e.g acoustic background, rotating equipment) which can produce high levels of measurement noise. With physical access to DOFS installations, the principle of acoustic or mechanical vibrations can also be utilized for malicious sensor tampering.

The current lack of anomaly-detection systems suggests that practical DOFS applications would benefit from an automated analysis to detect and classify compromised measurements. Noise classification makes it possible to identify its source and potentially remove its effects from the measurement in the future. This would expand the commercial applications of DOFS systems significantly. Neural networks have been used for error detection in cyber-physical applications in numerous studies with high-accuracy results. Specifically, long short-term memory (LSTM) neural network models have become popular in recent years to classify anomalies in sequential e.g time-series data.

Our investigation conducted a series of physical experiments using magnitude-controlled mechanical disturbances on bare free-hanging DOFS. Both random low-frequency vibrations at large displacement amplitudes and a constant high-frequency acoustic source at a low amplitude were employed. Experiments revealed that strain patterns are visually different with varying types and levels of disturbances. For the numerical analysis, statistics and machine learning-based approaches were applied for DOFS vibration noise classification, and their accuracy is discussed in detail. Results from the post-processing of compromised DOFS data suggest that it is possible to develop a vibration detection or classification system based on off-the-shelf DOFS interrogation equipment coupled with LSTM numerical tools.

1 INTRODUCTION

Today, many types of sensors are routinely employed for on-demand insights into the state of a system or a structure. Among the advanced sensing equipment, distributed optical fiber sensors (DOFSs) are gaining momentum for structural health monitoring (SHM) and damage

detection related applications [1]. The main benefit of DOFSs is that they can provide continuous monitoring of strain or temperature at a high spatial resolution and over long distances. That is to say, a single DOFS can replace thousands of point sensors such as strain gauges or thermocouples. Another unique characteristic is that the optical fiber is both the communication medium and the sensor itself, making the DOFS setup more compact and less energy consuming than traditional sensors. The glass fiber of the DOFS is resilient to harsh environmental factors such as chemicals, weather and corrosion. However, a major drawback of the Rayleigh scattering-based DOFS technology is that strain measurements become noisy when exposed to mechanical vibrations [2]. Practical experience from the lab confirms that when acoustic or vibrational disturbances are present, the integrity of the DOFSs data becomes compromised.

Artificial neural networks (ANNs) are a very powerful data processing tool that can learn and predict complex patterns between inputs and outputs of a system. The availability and quality of data are central aspects for the ANN deployment. DOFSs capability of generating large amounts of data quickly makes DOFSs and ANNs a good pairing. This is confirmed by recent studies which employed DOFS data in combination with ANNs for damage detection, localization [3] and disturbance classification purposes [4]. Applying machine learning (ML) analysis on DOFS data is an emerging research topic. A review of pattern recognition and ML methods in the context of vibration sensing with DOFSs has been provided by Li et al. [5]. This study highlighted the potential of deep learning (ANNs with multiple hidden layers) to extract features automatically from the sensor system data. For example, raw amplitude data from DOFSs was used to train a convolutional neural network (CNN) model to recognize and classify intrusions for perimeter security applications with good recognition accuracy [4, 6]. Similarly, ANNs were used for pipeline integrity monitoring to detect and classify external disturbances such as mechanical digging or drilling [3, 7]. Strain data from the DOFSs was also used in combination with an ANN for structural deformation estimation [8]. Evidently, several recent studies have successfully detected and classified vibrational movements by using DOFSs. However, the sources of disturbances in these investigations were often of similar type and magnitude.

In this study, strain data is gathered from bare free-hanging DOFSs exposed to highly-varied intensities of mechanical and acoustic disturbances. The strain data from the DOFSs is used to train ANNs to classify these vibrational disturbances. DOFS measurement data is essentially similar to a time-series array, hence an appropriate ANN structure has been employed for the analyses: a long short-term memory (LSTM) ANN. This study serves as a proof-of-concept for classification of compromised DOFS data by using LSTM ANNs.

2 MATERIALS AND METHODS

2.1 Optical Fibers

Single-mode optical fibers, SMB-E1550H by OFS Fitel, are hereby used as distributed strain sensors. While the light travels along the fiber, some of it becomes naturally scattered due to small-scale variations in the glass profile along the fiber length. When the variations in glass are much smaller than the wavelength of the light, this phenomenon is referred to as Rayleigh scattering. By exposing the fiber to mechanical strain, the small-scale structure of the fiber changes, which causes the Rayleigh scatter to change as well. Optical backscatter

reflectometry (OBR) is used to interpret and translate the difference in Rayleigh backscatter into a distributed strain (or temperature) measurement. This makes it possible to identify perturbations at a millimeter-scale resolution along the length of the fiber. Test equipment is typically very accurate but when optical fibers are integrated into real-life structures, strain or temperature can practically be measured at a resolution of ca. 10 $\mu\epsilon$ or 1 °C respectively.

Due to the high sensitivity of DOFSs, real-life measurements are not always entirely noise-free. Empirically, strains can easily vary by ca. ± 5 $\mu\epsilon$ [9]. Measurements can become noisy from excessive bending of the fiber, poor fusion splicing and dirty connections between the interrogation device and the fiber. Fortunately, it is easy to rectify these causes of noise by using good working procedures. A more problematic cause of noise are externally-created vibrations from mechanical sources or from sound, as these are not always possible to eliminate when the DOFSs are deployed. Therefore, classifying the source of vibrations could be highly informative and enables the extraction of new types of information from the measurement.

2.2 Measurement States

In this experimental/numerical investigation, the DOFSs were subjected to various distinct measurement states as described in Table 1. In each measurement instance, a reference trace and another trace obtained during exposure to a disturbance state were recorded by using the OBR 4600 interrogation device. The accompanying OBR software was used to process the raw trace data into strain measurements. The strain data was then used to train the LSTM ANN model to recognize different disturbance classes. As seen in Table 1, in addition to pure states *Norm*, *Vac* and *Fan1-4*, the measurement states were also assembled into a *Low* and a *High* group of composite states. The utility of *Low* and *High* composite states is explained in a later section of the paper.

2.3 Experimental Setup

Six optical fibers, approximately 0.99 m in bare (jacket-free) length, were fed through an airtight tube and a steel mesh tube in series as seen in Figure 1. The fibers were secured to two supporting polystyrene disks with hot glue. Two opposite-facing computer fans were positioned approximately 0.2 m equidistant to the center of steel mesh on either side of the setup. The air fan speed was controlled by an Arduino. A loud industrial vacuum cleaner was also positioned 1 m away from the setup. A microphone (44 kHz sampling frequency) was positioned at the lower opening of the steel mesh to record audio samples for every measurement state. The loud sound of the vacuum produced very low-amplitude but high-frequency movements of the fibers. The air fans did the opposite, inciting large-amplitude but low-frequency mechanical disturbances. The fibers were connected to the fiber optic switch (FOS) and the OBR 4600 device, which sent and received the back-scattered light from the optical fibers.

The optical fibers were subjected to one of the six pure i.e. non-composite states as defined in Table 1. Before implementing any of these states, reference measurements were taken for each fiber. Reference measurements are essentially additional measurements taken in the *Norm* state to establish the baseline trace for the fiber. Following that, the system was put through all measurement states (*Norm, Vac, Fan1-4*) and twenty repeated OBR measurements were taken for each of these classes of data by using Python UI automation. For each OBR measurement,

Table 1: Descriptions of disturbance states.

Measurement State	Description
Norm	Quiet room with minimal airflow. No deliberate disturbances activated.
Vac	Loud industrial vacuum cleaner placed 1 m away from optical fibers and turned on. Pure acoustic effect, no significant movement of air.
Fan1	Two opposite-facing air fans turned on at 1/4 maximum speed. Little visible movement in fiber ends.
Fan2	Two opposite-facing air fans turned on at 1/2 maximum speed. Moderate visible movement in fiber ends.
Fan3	Two opposite-facing air fans turned on at 3/4 maximum speed. Moderate visible movement in fiber ends.
Fan4	Two opposite-facing air fans turned on at maximum speed. Vigorous visible movement in fiber ends.
Low	A composite state consisting of {*Vac, Fan1*}.
High	A composite state consisting of {*Fan2, Fan3, Fan4*}.

distributed strains were calculated with a virtual gauge length of 10 mm and a sensor spacing of 5 mm. To exclude the tip-termination reflection effects (established in [10]), a sensing range of 800 mm was chosen, starting from the top entrance to the upper tube, which resulted in 159 data points per measurement for the given parameters. The recording of one Rayleigh backscatter trace with the OBR 4600 system took about 3-4 seconds from sending the command (i.e pushing the button) to saving the output raw data file.

Fast Fourier transforms (FFTs) of the audio recordings taken during each measurement state are shown in Figure 2 with both linear and logarithmic intensity scales. The *Norm* state exhibits the lowest levels of noise over all frequencies. It appears that *Fan2* and *Vac* share similar frequency spectrum intensity characteristics, which is interesting because the industrial vacuum cleaner appeared much louder to the ears than the fans. However, the placement of the microphone meant that the turbulent flow of air affected the microphone directly. The audio from *Fan3* and *Fan4* shows intensities several multitudes higher than that of any other state.

In short, the OBR 4600 interrogator is used in this investigation as it is commercially intended for distributed strain measurements. The additional functionality of vibration state classification is achieved by LSTM-based post-processing of strain data, outside of the interrogator and its accompanying software package. Twenty repeated measurements of the same type, over six parallel optical fibers in similar conditions provide a large volume of independent information for the machine learning process, while also accounting for fiber-to-fiber inconsistencies within analogous measurement states.

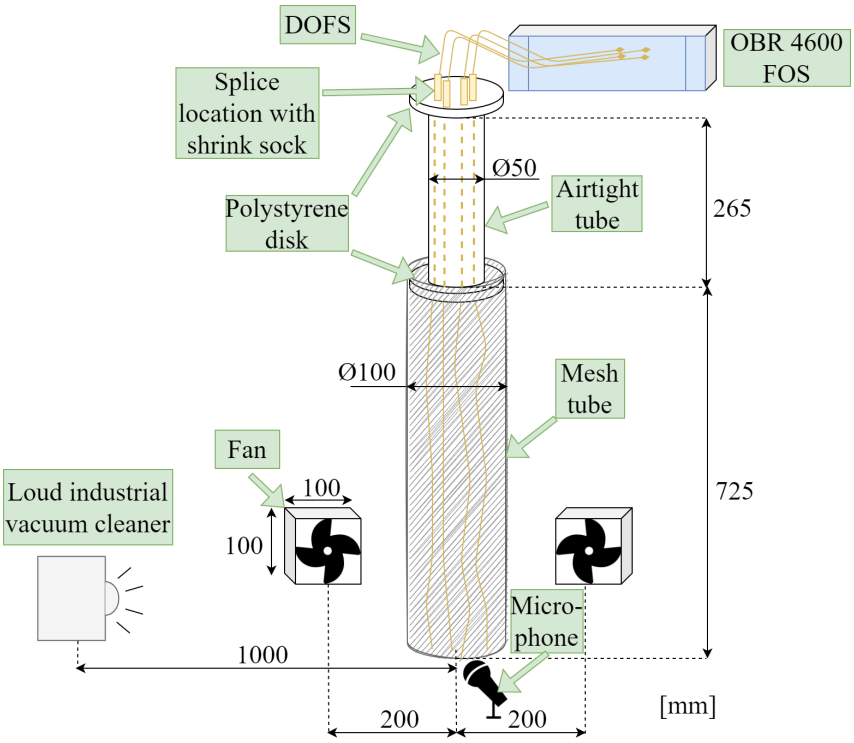
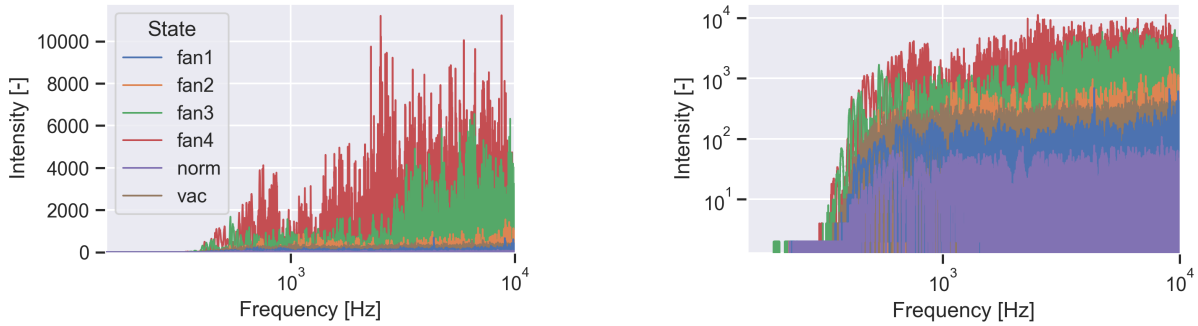


Figure 1: Experimental setup.



(a) Linear intensity

(b) Log intensity

Figure 2: Fourier transform of mic audio from each disturbance state activation.

2.4 LSTM Data Analysis Method

Machine learning (ML) is a broad term for data processing methods that are established by learning from data. Artificial neural networks (ANNs) are a subset of ML, which use parameters in non-linear functions to map inputs to outputs. Deep learning refers to ANNs with multiple hidden layers. For sequential data, it is common to use a recurrent neural network (RNN) structure. Unlike conventional ANNs, which consider outputs to be independent of the input

order, RNNs learn the sequential dependencies within data arrays by combining outputs from previous data points in the sequence with the data point it is currently processing. An LSTM network is an improved RNN, which mitigates the vanishing/exploding gradients problem that RNNs suffer from and allows LSTM networks to learn longer patterns than usual RNNs. This is accomplished by the addition of a memory unit, which stores outputs from data points further back in the array.

The Keras API was used with Python 3.9 to develop the ANN. An LSTM network consisting of two LSTM blocks of 32 and 16 units and a dropout layer was trained on 60% of all measured data. The training data set was augmented by flipping (i.e. multiplying the series by -1). The rest of the data set was split evenly between the validation (20%) and test (20%) data sets. L2-regularization with a regularization factor of 0.01 for the kernel, recurrent and bias weights was used in the 16 unit layer. The same structure was used both for the network with six pure states and the network with composite states. The learning rate was set to 0.0001 and a batch size of 16 was used. Early stopping was invoked when validation loss increased for more than three consecutive epochs.

3 RESULTS

3.1 Statistical Analysis

The strain measurements from distributed optical fibers in states *Norm, Vac* and *Fan1-4* revealed clearly how data becomes increasingly noisier with higher disturbance intensities. The strain curves also share some similarities between the different states. To exemplify the variation in strain curves, four arbitrary strain measurements for each disturbance state of fiber No. 3 (F3) are shown in Figure 3. The vertical dashed line marks the transition between the airtight tube (to the left) and the open mesh (to the right). Although no direct force was applied on the free-hanging fibers at any time, noise from the acoustic sound or from the turbulent flow of air manifested itself in the measurements as high strain values, increasingly so for higher disturbance intensities. This is reiterated in the probability density function (PDF) graph in Figure 4: for lower disturbance states (*Norm, Vac, Fan1*), the majority of strain values are very small, whilst for higher disturbance states (*Fan2, Fan3, Fan4*), the strain values tend to be higher, as shown by lower, horizontally-spread out center peaks on the PDF graph.

On the whole, the FFT curves of the vibration sources in Figure 2 and the strain PDF curves in Figure 4 are in agreement with each other: the states with high and sharp PDF curves are produced by low acoustic intensities, whilst the states with shallow PDF curves correlate with high acoustic intensities. One exception is *Fan2*, which gave a shallow PDF curve but had low acoustic intensities, as measured by the microphone.

Gramian angular difference field (GADF) matrices are an intuitive way of visualizing time series data [11]. To generate a GADF matrix, each point in the time series is parameterized into a polar coordinate representation. Then, the phase angle difference between each point and every other point in the time series is plotted in a matrix form. In this way, the GADF matrix makes it possible to observe general trends a time series data. Figure 5 shows the mean GADFs of all measurements. The matrices were generated by calculating the GADF matrix for each measurement and taking the average value of each matrix cell at the same index. On average, the GADF matrices become progressively noisier for higher disturbance intensities from left to right. Some disturbance states produce similar GADF matrix textures, for example, *Fan2-Fan4*.

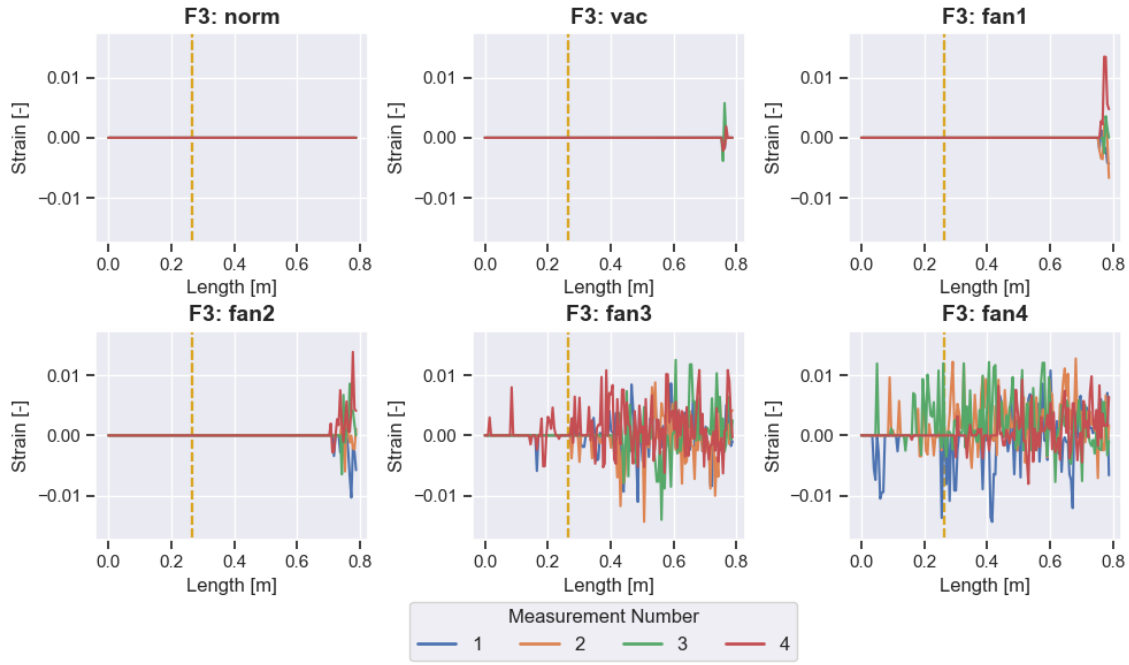
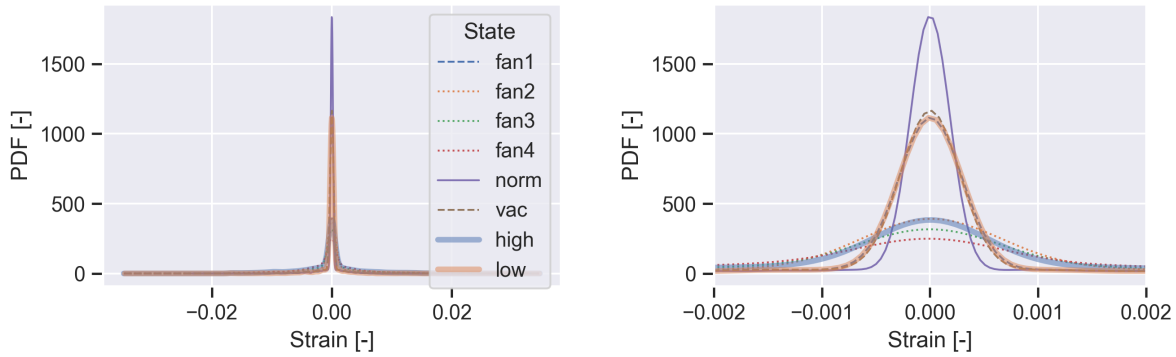


Figure 3: Representative examples of DOFS strain curves for six different disturbance states. Four measurements were arbitrarily selected from twenty taken for fiber No. 3 (F3).

Small stripes of constant color represent a time series that fluctuates often and randomly, similar to Gaussian noise. Fiber No. 5 (F5) was considered to be anomalous as its data produced similar noisy GADF matrices in all disturbance states. Therefore, data from F5 was not used for the following ANN data analysis.

The GADF visualization matrices through all optical fibers (the vertical direction in Figure 5) appear more consistent for the extreme states of *Norm*, *Fan3*, and *Fan4*. For each of these three states, the GADF textures of all fibers, excluding the outlier F5, are visually similar. In contrast, the intermediate states *Vac*, *Fan1*, and *Fan2* display a higher variation in GADF visualizations for different fibers. This is troublesome because all fibers should in principle be affected by similar external perturbations. As the noise variability from fiber to fiber is higher, these intermediate states are more difficult to classify already with the human eye, and similar difficulties can be expected for the ANN algorithm.

The transition from the enclosed tube to the open mesh tube occurs at approximately 1/3 length from the beginning of the data-series array (dashed lines in Figure 3). In Figure 5 this transition is, surprisingly, easiest to identify from the GADF matrix textures of the "disturbance-free" *Norm* state. The same transition line is also recognizable for the mellow *Fan1* state. Whereas for the acoustic noise of *Vac* and especially for the intense airflow of *Fan2-Fan4* states the transition line between two tubes becomes unrecognizable.



(a) Wide range of strains **(b)** Narrow range of strains

Figure 4: Distributions of strain values over twenty measurements for all fibers in different disturbance states.

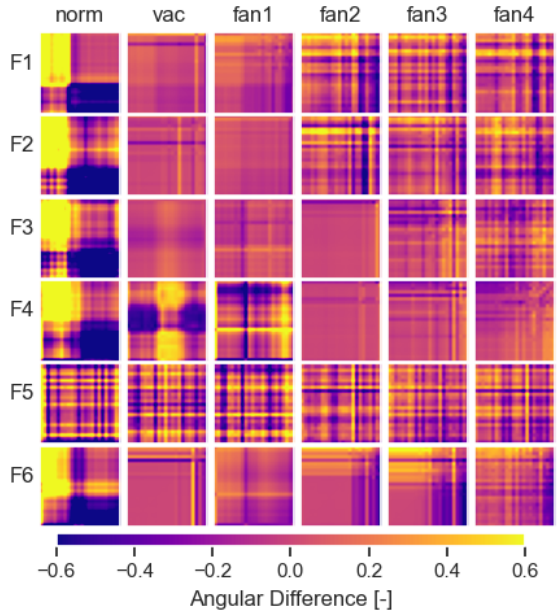
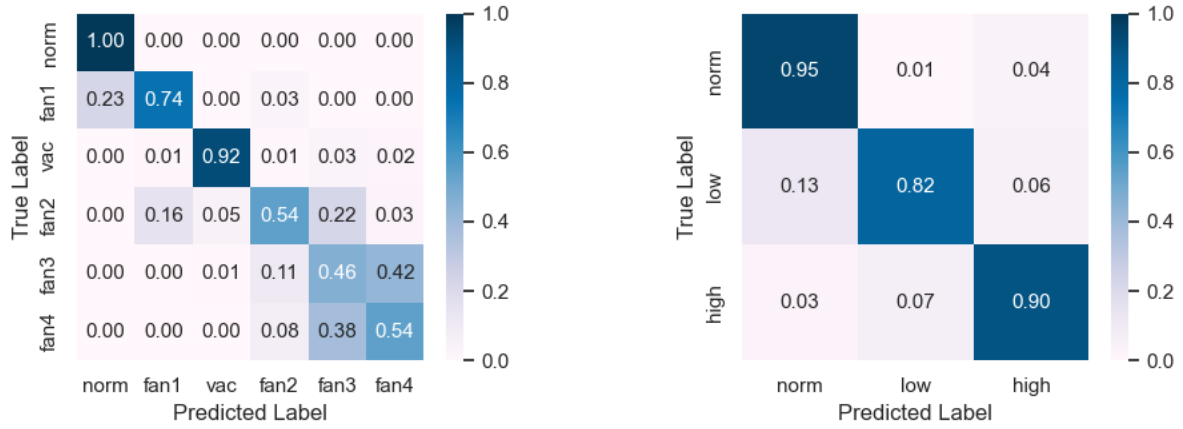


Figure 5: Mean of GADF matrices for all optical fibers (F1-F6).

3.2 LSTM Model Predictions

For the machine learning analysis, the LSTM ANN was applied on the optical fiber strain data. Figure 6a shows the predictions made by the LSTM model using six pure states (defined as Model M1) in a confusion matrix. The states are ordered according to the intensity levels from the FFT (Figure 2). Figure 6a shows that Model M1 sometimes confused *Fan1* and *Norm*, which can be attributed to the states' similar frequency spectrum characteristics as evident in

Figure 2. The model also struggled to differentiate between *Fan2-Fan4*, which exhibit clear similarities in Figure 5. However, the strong diagonal trend in the confusion matrix shows that when the model made an incorrect prediction, it was often one of the adjacent states (94% of Model M1's predictions were correct within the neighborhood of one state). Grouping the disturbance states into three larger, i.e. composite groups (Table 1) gave even better results, as shown in the confusion matrix for Model M2 in Figure 6b. A summary of the prediction accuracy of each data set for both models is provided in Table 2. The training, validation and test accuracy for both models were within 2-6%, suggesting that the models generalized well.



(a) M1: average test accuracy of 70%.

(b) M2: average test accuracy of 89%.

Figure 6: Confusion matrices for initial M1 and grouped M2 LSTM models.

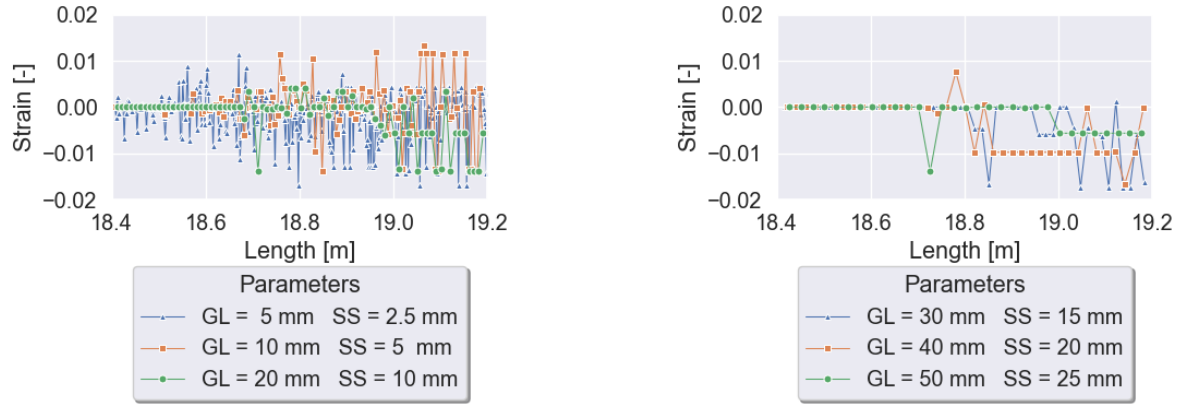
Table 2: Summary of accuracy of models M1 and M2.

	Data set		
Model	Training	Validation	Test
Initial (M1)	76%	70%	70%
Composite (M2)	90%	88%	89%

3.3 The Effect of the Virtual Strain Gauge Size

Traditional physical strain gauges come in different sizes depending on the structure being measured. In the post-processing stage of OBR strain measurements, virtual gauge lengths have to be chosen depending on the application. Until now, the optical fiber strains were calculated using GL = 10 mm gauge lengths at SS = 5 mm sensor spacing. However, the effect of these parameters on the LSTM ANN performance is unclear. A parametric study was conducted using six different virtual gauge size configurations with gauge lengths between 5 mm and 50 mm, to investigate how the sensing resolution affects the LSTM network prediction accuracy. To maintain the quasi-continuity of the strain data, the parameters were constrained

by the condition: SS = $\frac{1}{2}$GL. Figure 7 illustrates, by a generic example, how changing the post-processing parameters affects the calculated strains. The overall trend is that the strain curves become less noisy with increased gauge length. This is because the strain value is calculated using more points from the raw OBR data, which has a smoothing effect.



(a) Small virtual gauges

(b) Large virtual gauges

Figure 7: Comparison of the same strain measurement data with various post-processing parameters, GL and SS, using an example curve from the *Fan4* state.

Models M1 and M2 were retrained using the recalculated strain data for each added gauge size configuration. Figure 8 shows how the prediction accuracy was relatively stable for both models for gauge lengths above 10 mm. However, training the networks on strain data generated using the smallest gauge length of 5 mm resulted in a significant decrease in the performance of both models, especially for Model M2. This is likely due to the data containing excessive noise (Figure 7a) from the new source (OBR internal post-processing) which makes the vibrational noise (i.e. another independent source) more difficult to predict.

3.4 The Effect of Data Set Size

Another numerical investigation was conducted to understand how the quantity of training data files affects the test accuracy. The total number of files used for training was 360: 60% of 20 strain files × 6 states × 5 fibers. A pseudorandom reduced selection of files was additionally chosen with each disturbance class represented evenly. The findings in Figure 9 show a clear positive correlation i.e. more input data enables more accurate predictions. It is important to note that the climbing trends have not started to plateau, suggesting that the prediction accuracy of the classification models could be improved further with more data. Due to stochastic variations in exactly which data files were chosen for each data segment, the results for small data batches can break the trend. For example, the first point for Model M1 is abnormally high presumably because the randomly selected strain curves were among those that are the easiest to classify. The same random seed was used throughout this study, however, the models may perform slightly differently with a different random seed.

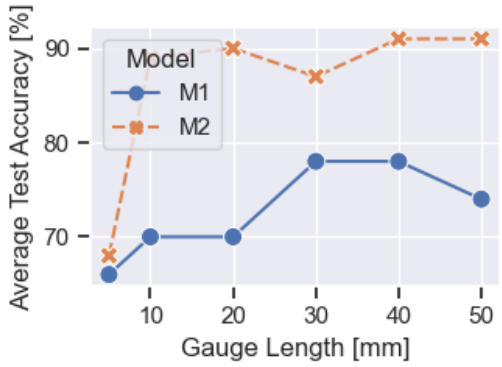


Figure 8: The effect of gauge length on the LSTM prediction accuracy (SS = $\frac{1}{2}$GL).

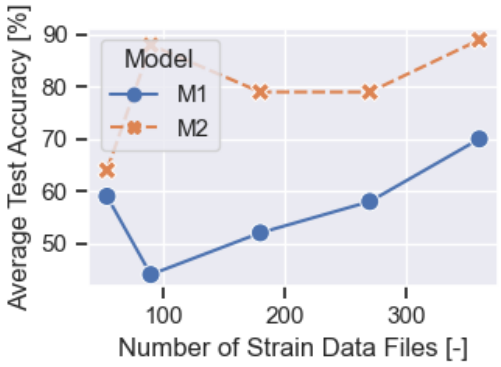


Figure 9: The effect of training data quantity on the LSTM prediction accuracy.

4 CONCLUSIONS

It was experimentally confirmed that both large-amplitude random vibrations created by the air fans and low-amplitude acoustic vibrations produce additional noise in the nearby Rayleigh scattering-based DOFS strain measurement system. The noise manifests itself as spikes of strain even though no external force is applied on the fiber and the strain is expected to be zero. When the intensity of vibrations is increased, a larger proportion of strain data take on abnormally large values.

For automated classification of noise, LSTM ANN models were trained and applied on the experimental DOFS strain data. The accuracy of Model M1, differentiating between six diverse states of perturbations, was 70% with good generalization. Furthermore, incorrect predictions were almost always in the adjacent disturbance states in terms of noise intensity. Noise classification was improved by grouping data into three larger composite classes, giving an accuracy of 89%.

A parametric numerical study, where strain data was calculated with different virtual gauge lengths, revealed that the performance of both Model M1 and M2 is stable for gauge lengths above 10 mm. This confirms that using LSTM ANNs is a robust method of disturbance classification for DOFS data. Investigating the effect of training data set size on LSTM prediction accuracy showed a positive correlation, confirming that it is important to have a large quantity of data. In the current case, the prediction accuracy could likely be further improved with more measurement data.

In conclusion, this experimental/numerical analysis indicates that ANN-based models such as LSTM can distinguish between different classes of vibrational disturbances on the DOFS measurements. The method proved to be relatively indifferent to the main post-processing parameters but significantly dependent on the training data quantity.

ACKNOWLEDGMENTS

This project was partially funded by the Fuel Cells and Hydrogen 2 Joint Undertaking under grant agreement No. 826262, https://thor-fch2.eu/.

REFERENCES

[1] Ismail Alj, Marc Quiertant, Aghiad Khadour, Quentin Grando, Benjamin Terrade, Jean-Claude Renaud, and Karim Benzarti. Experimental and numerical investigation on the strain response of distributed optical fiber sensors bonded to concrete: Influence of the adhesive stiffness on crack monitoring performance. *Sensors*, 20(18):5144, 2020.

[2] Stephen T. Kreger, Justin W. Klein, Nur Aida Abdul Rahim, and Joseph J. Bos. Distributed rayleigh scatter dynamic strain sensing above the scan rate with optical frequency domain reflectometry. In *Fiber optic sensors and applications xii*, volume 9480, page 948006. International Society for Optics and Photonics, 2015.

[3] Yu Bai, Jichuan Xing, Fei Xie, Sujie Liu, and Jinxin Li. Detection and identification of external intrusion signals from 33 km optical fiber sensing system based on deep learning. *Optical Fiber Technology*, 53:102060, December 2019.

[4] Chengang Lyu, Ziqiang Huo, Xin Cheng, Jianying Jiang, Alimina Alimasi, and Hongchen Liu. Distributed Optical Fiber Sensing Intrusion Pattern Recognition Based on GAF and CNN. *Journal of Lightwave Technology*, 38(15):4174–4182, August 2020. Conference Name: Journal of Lightwave Technology.

[5] Junchan Li, Yu Wang, Pengfei Wang, Qing Bai, Yan Gao, Hongjuan Zhang, and Baoquan Jin. Pattern Recognition for Distributed Optical Fiber Vibration Sensing: A Review. *IEEE Sensors Journal*, 21(10):11983–11998, May 2021. Conference Name: IEEE Sensors Journal.

[6] Seedahmed S. Mahmoud and Jim Katsifolis. Robust event classification for a fiber optic perimeter intrusion detection system using level crossing features and artificial neural networks. In *Fiber Optic Sensors and Applications VII*, volume 7677, pages 60–71. SPIE, 2010.

[7] Huijuan Wu, Ya Qian, Wei Zhang, and Chenghao Tang. Feature extraction and identification in distributed optical-fiber vibration sensing system for oil pipeline safety monitoring. *Photonic Sensors*, 7(4):305–310, 2017.

[8] Gong-Yu Hou, Zi-Xiang Li, Kai-Di Wang, and Jin-Xin Hu. Structural deformation sensing based on distributed optical fiber monitoring technology and neural network. *KSCE Journal of Civil Engineering*, 25(11):4304–4313, 2021.

[9] Erik Saeter, Kaspar Lasn, Fabien Nony, and Andreas T. Echtermeyer. Embedded optical fibres for monitoring pressurization and impact of filament wound cylinders. *Composite Structures*, 210:608–617, February 2019.

[10] Usenco Valeria. Analysis of DOFS Measurements Compromised by Mechanical and Acoustic Vibrations Using ANN. Master's thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, 2022.

[11] Zhiguang Wang and Tim Oates. Imaging Time-Series to Improve Classification and Imputation. *arXiv:1506.00327*, May 2015. arXiv: 1506.00327.