ISSN 1213 - 1539

Rok / Year:	Svazek / Volume:	Číslo / Number:	Jazyk / Language
2021	23	2	EN

Data-driven classification of events in the vicinity of an optical fibre

Zoltan Galaz[1], *Jiri Mekyska*[1], *Jiri Prinosil*[1], *Vit Novotny*[2]

galaz@vutbr.cz, novotny@fbm.vutbr.cz

[1] Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology

[2] Institute of Informatics, Faculty of Business and Management, Brno University of Technology

DOI: -

Abstract: This article focuses on the data-driven classification of vibration sources via a distributed optical fibre sensor system utilizing Gaussian Mixture Models and Mel-spectrograms. In total, six categories are considered: one person walking/running, group walking/running, hammering, and car riding. Two evaluation scenarios are used: individual models (per-session) and a joint model. Tenfold cross-validation is employed during the evaluation phase.

elektrorevue
ISSN 1213 - 1539

Data-driven classification of events in the vicinity of an optical fibre

Zoltan Galaz[1], Jiri Mekyska[1], Jiri Prinosil[1], Vit Novotny[2]

[1]Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology
Email: galaz@vutbr.cz

[2]Institute of Informatics, Faculty of Business and Management, Brno University of Technology
Email: novotny@fbm.vutbr.cz

Abstrakt – *This article focuses on the data-driven classification of vibration sources via a distributed optical fibre sensor system utilizing Gaussian Mixture Models and Mel-spectrograms. In total, six categories are considered: one person walking/running, group walking/running, hammering, and car riding. Two evaluation scenarios are used: individual models (per-session) and a joint model. Tenfold cross-validation is employed during the evaluation phase.*

1 Introduction

Over the last decades, optical fibres have found a wide range of applications. Especially, distributed fibre sensors are a very attractive group in the fibre optic sensor area. They behave like hundreds or thousands of sensors spread along an optical fibre. Using the distributed fibre sensors it is possible to measure quantities such as pressure, temperature, fibre loss pattern or vibrations at the scale from several metres up to tens of kilometres depending on particular solution. In this cases, the signal has to provide not only the information about the measured value but also the position of the measurement. One of the solutions to deal with this task uses reflectometry principle [12], where localization of the measurement position is based on the transmission of high power short-time pulses. These pulses go through the fibre and the portion of the light is scattered due to elastic or inelastic effects influenced by a given physical quantity (temperature, pressure, etc.). These can be measured by processing of the scattered signal that is recaptured by the fibre and is spread either back or in the direction together with the original signal or in both directions. Unlike elastic effects, which do not add any new spectral components to the signal, inelastic effects enrich the signal with new optical frequencies.

According to the time delay between instances of light pulse transmission and the instances of the response sample reception, the location of the response generation can be estimated. The time series of measurements along the fibre can give us (after the demodulation and low-pass filtering steps) an acoustic signal representing the vibrations captured by the fibre in that locations. Such a signal can be further used for acoustic events detection along the fibre [7]. The event detection technique gives us information about the presence of an object that is the source of the vibrations in the vicinity of the optical fibre at the particular location, but it can not determine what type of object it is (e. g. car, person, train, factory, etc.).

In this work, we extend the event detection technique by the classification step in order to determine the type of vibration source. The article is organized as follow: section 2 deals with data acquisition and methods description. The results are presented in section 3. And finally, section 4 concludes the article.

2 Dataset and methods

2.1 Dataset

In this experiment, we used a test-bed with 1550 nm wavelength to minimize attenuation along the fibre. The pulses were generated by an acousto-optic modulator, which in addition to the modulation process with high extinction ratio (> 50 dB) shifts the spectrum of the optical signal in order of tens of MHz. The pulse width was set to 200 ns. Generated pulses are amplified by an erbium-doped fibre amplifier booster and sent to the fibre via a circulator. The back-scattered optical signal from the fibre returns via the circulator to an erbium-doped fibre pre-amplifier and is mixed with the optical local oscillator signal generated by the laser. The resulting signal is converted to the electrical form by an optical-to-electrical converter where a PIN photodiode and a transimpedance amplifier were used. Electrical signal is amplified again and then it is mixed with a sinusoidal signal with the same frequency as the radio pulse signal from the acousto-optic modulator driver. The high frequency components (above 20 MHz) of the mixer output are filtered out and the resulting signal is converted to the digital form by an analog-to-digital converter with the sampling frequency of 50 MSPS. An additional low-pass filtering is performed to obtain the acoustic signal representing the vibration source.

We collected the data in three sessions. Two of them (S1–2) were held at the military training ground Bzenec (Czech Republic) and the last one (S3) in an area of

elektrorevue
ISSN 1213 - 1539

Table 1: Number of training events in each session.

Event	S1	S2	S3
car ride	421	439	-
running of 2–3 people across a fibre	130	-	-
running of 2–3 people along a fibre	252	256	138
walking of 2–3 people along a fibre	345	635	456
hammering	206	421	392
running of one person across a fibre	79	-	-
running of one person along a fibre	453	763	319
walking of one person along a fibre	78	430	315

a transforming station owned by the E.ON company. Each environment has different conditions in terms of optical fibre placement, soil composition, etc. Thus we collected heterogeneous data which enabled us to perform more robust mathematical modelling and less vulnerable to overfitting (the fibre was laid 40 − 80 cm under the ground).

In these sessions, we simulated the following events: 1) walking of an individual along a fibre; 2) running of an individual along a fibre; 3) running of an individual across a fibre; 4) walking of a group of 2–3 people along a fibre; 5) running of a group of 2–3 people along a fibre; 6) running of a group of 2–3 people across a fibre; 7) hammering; 8) car ride. In addition, in both environments, the acquisition was running simultaneously on several parallel fibres which enabled us to collect more data. The raw data sampled with $f_s = 2$ kHz were consequently processed by a detector that (based on an empirically selected threshold) identified frames containing some (yet not classified) events. The duration of each frame was fixed to 1 s and its width was varying depending on the intensity of an event.

For the purpose of a machine learning model training we created a dataset, where we manually labelled the events in two steps: 1) we filtered out events that were detected outside a region of interest (i. e. outside an area where we performed activities listed above); 2) we manually rechecked each frame based on the waterfall visualisation and a waveform extracted from the central part of the frame. The size of the final training data set is summarised in Table 1. Finally, since we observed that walking/running along/across an optical fibre have very similar spectral properties, we have decided to fuse these activities into one. Thus in the frame of this study, we considered the following set of events that were consequently modelled by machine learning algorithms:

1. walking of an individual,
2. running of an individual,
3. walking of a group of 2–3 people,
4. running of a group of 2–3 people,
5. hammering,
6. car ride.

2.2 Feature extraction

To extract important and discriminating characteristics of the signals, we computed so called Mel-spectrograms [9, 10]. The reason for it is that most of the important information describing an event occurring in the vicinity of an optical fibre is present in the low-frequency spectrum range (empirically, we found out that most of the important information lays in the range of 0 Hz to 150 Hz). Mel-spectrogram is computed via the following steps.

First, the Short-Time Fourier Transform (STFT) [3] of the input signal $s[n]$ of length N is computed as

$$\begin{aligned} S[k,m] &= \sum_{n=0}^{N-1} s[n]w[n-mL]\mathrm{e}^{-\mathrm{j}k\frac{2\pi}{N}n}, \\ k &= 0,1,\ldots,N-1, \\ m &= 0,1,\ldots,M-1, \end{aligned} \tag{1}$$

where M denotes the number of segments obtained using a segmentation window $w[n]$ composed of L samples. In the frame of this work, we used the Hann segmentation windows with $L = 400$ samples with 50 % overlap (sampling frequency f_s of 150 Hz, and NFFT of 400).

Next, power spectrum $P[k,m]$ is computed as

$$P[k,m] = |S[k,m]|^2 \tag{2}$$

and filtered by a filer-bank $B[o,k]$ consisting of O filters.

For this purpose, we used a filter bank of 20 triangular filters (each filter had a frequency range of: minimum frequency of 0 Hz, and maximum frequency of 150 Hz) as

$$\mathrm{X} = \mathrm{BP}. \tag{3}$$

The filter bank was non-linearly distributed on the Hz scale. The non-linear distribution was accomplished by the conversion of the frequency units in Hz to mels as

$$f_m = 2595 \log_{10}\left(1 + \frac{f}{700}\right), \tag{4}$$

where f is the frequency value in Hz, and f_m is the frequency value in mels.

After the filtration, the matrix $X[o,m]$ contained O subbands $o = 1,2,\ldots,O$. With the described settings, each of the computed Mel-spectrograms had the shape of 20 x 11 (220 samples). It is important to note that the dimensionality of a feature vector is crucial as if it gets too large, the computational complexity can become unmanageable. The bigger problem however being that as the dimensionality of a feature vectors grow, more samples are needed to train a learning algorithm without the undesirable overfitting [6]. In this work, the proposed dimension of the Mel-spectrogram was selected as a trade-off between the frequency resolution of the Mel-frequency filter bank and the sample size.

elektrorevue
ISSN 1213 - 1539

2.3 Design of the classifier

When there is an unknown number of classes to be modelled by a classification model, a classical supervised learning setup [4] is not desirable. The main reason is that in a supervised N-class classification scenario, we assume there are exactly N distinct classes, and for each class, there is a reasonable number of samples that a classifier uses to learn from. However, when the number of classes is not known beforehand, and it is likely to change over time, such a classifier would have to be re-trained every time after the number of classes changes. Moreover, a classical N-class classifier is trained to predict 1–N classes (depending on the setup: i. e. one/multi-label classifier [11]), but the scenario in which none of the classes should be predicted is a special case that would have to be treated by: a) the addition of another class covering the behavior of the input feature vector in every other case than the ones covered by the original classes (such a case could be named a background or no event class); b) using the predicted class probabilities and setting a threshold that the predicted probability must match to be considered as being valid. In this work, we faced exactly this scenario. In our case, the events of interest occurring in the vicinity of an optical fibre are non-frequent and should be identified and predicted correctly. In all other cases, the classifier should be capable of recognizing that there is no event and should not detect anything (i. e. the background case is not a class to be predicted).

As one can imagine, the background scenario is almost impossible to be modelled precisely because the training samples would have to cover an enormous number of possibilities so that the classifier is provided with enough information describing the behavior of the input feature vectors given the circumstances that are specific to the particular learning task. In this work, the background signals may differ depending on the weather, season, location, and many other factors. It is quite reasonable to assume that the input signals would be significantly different in the winter, when the soil could be frozen, than in the summer, and so on. In addition, the type of the soil can vary from one location to another, not to mention that it can vary at the location of interest as well (depending on the size and other specifics of the area covered by the system). Hence, the number of training samples covering the background would have to be significantly larger than the number of samples covering the events of interest, which is also not feasible for a supervised classifier (moreover, some of the events may be so rare that there are only a handful of samples available, and such classes are far harder to model precisely) as the classes would be largely imbalanced.

For this purpose, we designed the system as an ensemble of unsupervised learning models, where each of the models is fitted to a single class (event) only. This means that when the system is used to detect the presence of N events, it comprises exactly N distinct models. This setup enables the classifier to be fine-tuned when the number of samples for some of the classes increases as only the particular models need to be re-trained. Moreover, it also provides a possibility of adding new classes without the necessity of re-training an entire classifier. When there is a new class, a new model is fitted and added to the ensemble, which makes the classifier very flexible. More specifically, in this work, we used Gaussian Mixture Models [8] as the probabilistic model that assumes all the data samples are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Each of the models can consist of 1–N dimensional Gaussian probability distributions [1] designed to model a given class (event).

From the high-level perspective, the workflow of the event classifier while computing a prediction for an unknown sample can be described as follows:

1. When an unknown sample is to be classified, it is first put into the feature extraction module that computes the feature vector based on the pre-defined settings.

2. Next, the feature vector is inserted into an ensemble of N Gaussian mixtures to compute the posterior probability of the sample belonging to each of the mixtures.

3. The computed posterior probabilities are weighted by the probability threshold (each mixture has its own threshold, and if the given probability is bellow the threshold, it is converted to zero).

4. The thresholded probabilities are joined into a vector of output probabilities, and the largest probability is used to determine the predicted class (if none of the thresholded probabilities is larger than zero, a uncertain class is returned indicating that it is likely that there is no event occurring, i. e. it is the background only).

High-level overview of the event classifier can be found in Figure 1. The figure demonstrates the high-level functionality only, and the detailed schema of the classifier (ensemble of Gaussian mixtures) can be seen in Figure 2.

2.4 Experimental evaluation

To evaluate the classification performance of the trained models for each of the sessions (S1–S3) individually (i. e. each session was evaluated individually), we employed the stratified 10-fold cross-validation [2]. To evaluate the classification performance of the model for a combination of the three sessions (joint model), we joined the data acquired from these three sessions, shuffled them, and applied the stratified train-test split with the following ratios: training data 85 % and testing data 15 %. To quantify the classification performance of the trained models we computed the following classification metrics [5]: a) precision (PRE); b) recall (also known as sensitivity; SEN), and c)

elektrorevue
ISSN 1213 - 1539

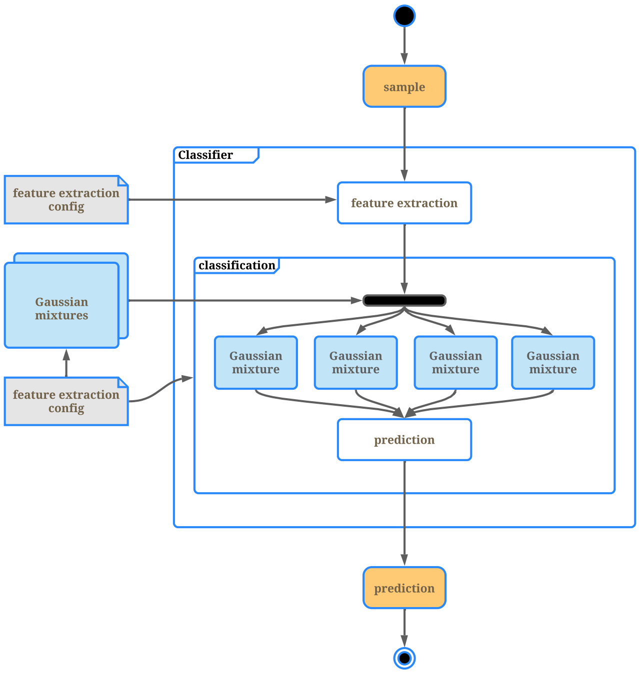

Figure 1: Overview of the system's architecture.

specificity (SPE). These metrics are computed as

$$\text{PRE} = \frac{TP}{TP+FP}, \quad (5)$$

$$\text{SEN} = \frac{TP}{TP+FN}, \quad (6)$$

$$\text{SPE} = \frac{TN}{TN+FP} \quad (7)$$

where $N = (TP+FP)(TP+FN)(TN+FP)(TN+FN)$, TP (true positive) and FP (false positive) represent the number of correctly identified events of interest and the number of incorrectly identified events of interest, respectively. Similarly, TN (true negative) and FN (false negative) represent the number of correctly identified events other that the event of interest and the number of incorrectly identified events other that the event of interest (event of interest being the true label of an event).

3 Results

The results for the two previously mentioned evaluation scenarios (individual models, joint model) use the following class notation: a) individual walk; b) individual run; c) group walk; d) group run; e) hammer; and f) car. To present the classification performance of each model, we used the confusion matrix (figure) and table with the previously described classification metrics. In the confusion matrix, we added the "uncertain class" to show the number of samples that the classifier was not capable of classifying with enough certainty. These samples were therefore not classified at all, and would not be recognized by the model (background case).

Table 2: Classification performance of the trained models.

Event	PRE [%]	SEN [%]	SPE [%]	N
S1				
individual walk	4.00	12.00	85.00	78
individual run	31.00	7.00	95.00	453
group walk	100.00	3.00	100.00	345
group run	0.00	0.00	100.00	252
hammer	0.00	0.00	100.00	206
car	0.00	0.00	100.00	421
S2				
individual walk	24.00	25.00	86.00	430
individual run	36.00	16.00	90.00	763
group walk	48.00	21.00	94.00	635
group run	0.00	0.00	100.00	256
hammer	100.00	13.00	100.00	421
car	0.00	0.00	100.00	439
S3				
individual walk	46.00	7.00	98.00	315
individual run	80.00	1.00	100.00	319
group walk	23.00	13.00	84.00	456
group run	0.00	0.00	100.00	138
hammer	100.00	14.00	100.00	392
Joint model				
individual walk	10.00	20.00	83.00	76
individual run	53.00	20.00	95.00	202
group walk	56.00	20.00	95.00	211
group run	0.00	0.00	100.00	60
hammer	100.00	14.00	100.00	173
car	0.00	0.00	100.00	141

[1] Event – event of interest; PRE – precision (in percentages); SEN – sensitivity (in percentages); SPE – specificity (in percentages); N – number of samples (support).

The classification performance of the models trained for the data acquired in sessions S1–S3 as well as the joint model are summarized in Table 2. Moreover, the confusion matrix of the joint model is shown in Figure 3.

As can be seen, in every scenario, the trained models achieve good specificity. However, their sensitivity as well as precision differs quiet dramatically. In the case of the individual models, it can be seen that the hammer event was identified with 100 % precision in two out of the three sessions (sensitivity ranged from 13–14 %). Next, it can also be seen that the car event was not recognized at all (it is very likely that there was not clear pattern in the spectral properties of this class and much more data would be needed to model it precisely). As opposed to this event, the events based on the walking and running were modeled with reasonable precision besides the one based on the running of the group of people. This event was not recognized in any of the three sessions, which shows that it is too similar with other events (e. g. group walk event). It can also be seen that other events such as individual walk and individual run were modeled with different precision and

elektrorevue
ISSN 1213 - 1539

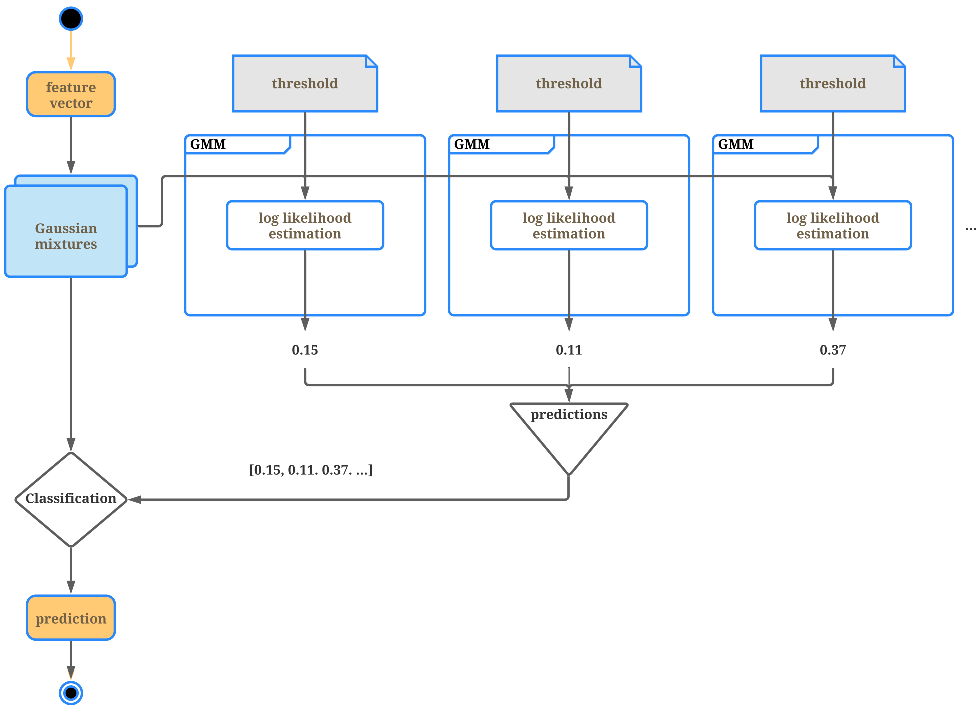

Figure 2: Design of the system's classifier.

sensitivity in each session. This demonstrates how big the effect of the season, location and other external factors is.

Regarding the joint model, it is obvious that joining the data (increasing the sample size) helped to model the classes. However, the events car and group walk were still not recognized. With respect to other events, the classification performance of the associated models were improved, which shows that there is still room for improvement, and more data samples are necessary to train the models yielding better results. But still, when checking the confusion matrix, it can be seen that most of the samples fell into the "uncertain class" showing that the models were not trained enough to provide confident predictions, which further highlights an importance and necessity of much larger and much more heterogeneous dataset.

Even though that at first sight it may seem that the achieved classification performance is not desirable, the following facts need to be noted:

1. Each classification task is bound to the criterion of success. In this case, the maximum sensitivity is not the key performance indicator as the system is not designed to discriminate the events as precisely as possible, nor is it required to capture every single event. Since the system is supposed to create alerts in the case any of the events of interest happens, it is desirable to aim for the maximum precision and specificity (it is expected to create minimum false alerts, and the confidence of the event when alerted should be as high as possible).

2. The classification performance of the model depends on the data, and the learning task itself. When the system is to classify a class, it is assumed that this class differs from the other classes significantly so that the learning algorithm can differentiate the classes based on the discovered patterns in the provided feature vectors. With this information, it is quiet evident that other classes should be introduced as the classes used in this experiment did not differ enough to be distinguished with higher confidence.

3. We summarize only the preliminary efforts and despite unsatisfactory results, the potential of this method is pretty clear. In future work, we will have to focus on acquiring much more data. With more data, we will be able to check the properties of the signals for given classes, and fine-tune the feature extraction as well as model training steps (addition of another features, training other models, etc.).

4 Conclusion

This paper describes a new approach of data-driven classification of acoustic signals. The signals correspond to events (vibration source) captured by the distributed sensor fibre system based on reflectometry principle. The six different event types are considered from classification task (individual person walking, individual person running, group of persons walking, groups of persons running,

elektrorevue
ISSN 1213 - 1539

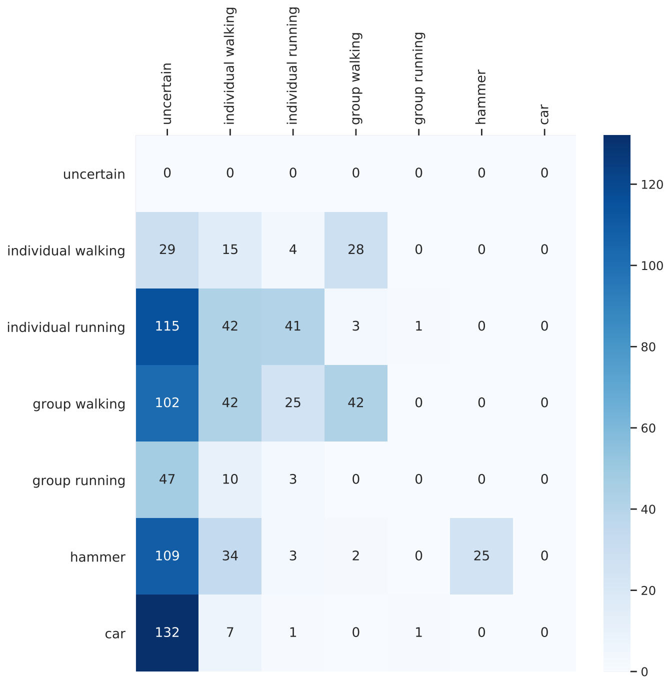

Figure 3: Confusion matrix for the joint model.

hammering, and car riding). The classification consists of Mel-spectrogram feature extraction and Gaussian Mixture Model for class-probability estimation. Two evaluation scenarios were performed upon three independent data session. In the firsts scenario individual model is generated and evaluated for each session. In the second scenario joint model is generated and evaluated for combination all session data. According to the results, it is obvious that combing all data together provide better model outperforming individual models of all sessions. Nevertheless some event types (car riding and group of persons walking) a has poor performance in the both scenarios.

This study has some limitations, first limitation being relatively small sample size limiting the statistical significance of the inference made on the population of events occurring in the vicinity of an optical fibre. The second limitation of this study is the fact that a single parametrization method was used only. The reason behind this choice however is the exploratory nature of the study and the fact it aims at deeper investigation of a single method being integrated into the system rather than investigating an entire spectrum of signal processing methods that may be suitable for the given task. Despite unsatisfactory results, the potential of this method has been proven. In future work, we will have to focus on acquiring much more data as well as to explore the possibility of using data augmentation to increase the number of data samples quiet radically. With more data, we will be able to check the properties of the signals for given classes, and fine-tune the feature extraction and the machine learning model.

Acknowledgment

This research was funded by Ministry of Interior of the Czech Republic under grant no. VI20172020078.

References

[1] Peter Ahrendt. "The multivariate gaussian probability distribution". In: *Technical University of Denmark, Tech. Rep* (2005).

[2] Michael W Browne. "Cross-validation methods". In: *Journal of mathematical psychology* 44.1 (2000), pp. 108–132.

[3] Daniel Griffin and Jae Lim. "Signal estimation from modified short-time Fourier transform". In: *IEEE Transactions on acoustics, speech, and signal processing* 32.2 (1984), pp. 236–243.

[4] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. "Overview of supervised learning". In: *The elements of statistical learning*. Springer, 2009, pp. 9–41.

[5] Mohammad Hossin and MN Sulaiman. "A review on evaluation metrics for data classification evaluations". In: *International Journal of Data Mining & Knowledge Management Process* 5.2 (2015), p. 1.

[6] Mario Köppen. "The curse of dimensionality". In: *5th Online World Conference on Soft Computing in Industrial Applications (WSC5)*. Vol. 1. 2000, pp. 4–8.

[7] Vit Novotny et al. "Critical Infrastructure Monitoring System". In: *2021 IEEE 17th International Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE. 2021, pp. 165–170.

[8] Carl Edward Rasmussen et al. "The infinite Gaussian mixture model." In: *NIPS*. Vol. 12. 1999, pp. 554–560.

[9] Yuma Sakashita and Masaki Aono. "Acoustic scene classification by ensemble of spectrograms based on adaptive temporal divisions". In: *Detection and Classification of Acoustic Scenes and Events (DCASE) Challenge* (2018).

[10] Jonathan Shen et al. "Natural tts synthesis by conditioning wavenet on mel spectrogram predictions". In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2018, pp. 4779–4783.

[11] Grigorios Tsoumakas and Ioannis Katakis. "Multilabel classification: An overview". In: *International Journal of Data Warehousing and Mining (IJDWM)* 3.3 (2007), pp. 1–13.

[12] Eric Udd and William B Spillman Jr. *Fiber optic sensors: an introduction for engineers and scientists*. John Wiley & Sons, 2011.