Bridge Maintenance, Safety, Management, Digitalization and Sustainability – Jensen, Frangopol & Schmidt (eds)

Machine-learning for analyzing bridge displacement using radar data

M. Arnold
ci-tec GmbH, Karlsruhe, Germany

S. Keller
Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe, Germany

ABSTRACT: This paper introduces a novel machine-learning approach based on non-invasive ground-based radar (GBR) time series data for classifying short-span bridge vehicle crossing events. GBR is used to remotely measure the bridge displacement, which is otherwise difficult to acquire but is an essential quantity for Structural Health Monitoring (SHM). For a comprehensive Bridge SHM, monitoring traffic is beneficial to gain knowledge about the actual loading situation. This can be challenging with global responses like displacement. This study indicates that it is possible to classify crossings as single- or multi-presence from displacement signals using tree-based learners and MiniRocket. Thus, our approach serves as a proof of concept to establish remote and data-driven displacement approaches in the context of BWIM. We rely on recordings of the bridge deck taken by an unmanned aerial vehicle as reference data. Despite a small, imbalanced, and biased dataset, we achieve a balanced accuracy of 90%.

1 INTRODUCTION

Bridges are an essential part of any transportation system. Failure can lead to fatalities, but even a gradual decline in loading capacity may already significantly impact freight traffic. Therefore, it is necessary to assess their condition regularly. In Germany, bridges are inspected periodically, frequently visually. These superficial inspections may detect external damage, including cracks, but they do not provide any information about the behavior of the bridge itself, like the load-bearing capacity. On the other hand, to apply adequate measures such as load restrictions, it is vital to have insights into the weight of vehicles to which the bridge is exposed.

So-called weigh-in-motion (WIM) systems are usually deployed to monitor traffic, but they typically require directly attached sensors, inflicting damage on the bridge (Paul & Roy 2023).

Pavement-based WIMs have direct contact with vehicles, simplifying traffic characterization, but the sensors are exposed to significant stress. Nothing-on-road (NOR) bridge WIMs (BWIM) try to avoid traffic exposure by only attaching sensors to the lower side of the bridge and using the bridge itself as a scale (Yu et al. 2016). To minimize interference, remote and non-invasive systems have been investigated (Ojio et al. 2016). While BWIM reduces both the system complexity, sensor exposure, and cost, signal analysis becomes more challenging. Accordingly, traffic characterization from vehicle crossings, henceforth called events, is still an active field of research. Many standard BWIM Structural Health Monitoring (SHM) methods cannot handle multiple-presence crossings and instead require single-vehicle events (OBrien et al. 2006; Ieng 2015). Such single-events must be detected automatically for a BWIM under normal traffic conditions. Moghadam et al. (2023) introduce a method for calculating a bridge influence line from multi-presence crossings. However, they recommend using a different approach for single crossings since it is more exact, making a differentiation between single and multi events desirable.

DOI: 10.1201/9781003483755-164

Kawakatsu et al. (2023) use Deep Learning (DL) for BWIM. They train Convolution Neural Networks (CNN) to classify traffic using equally long strain time series. According to their study, the CNNs can handle multi-presence events. However, they do not study the effects of several vehicles. Ground-based radars (GBR) have recently been investigated for monitoring bridges (Gentile & Bernardini 2010; Michel & Keller 2021a, b, 2022). Since they are easy to set up and require little maintenance, inspections can be performed on the fly. More importantly, GBR measures the bridge displacement. Displacement is highly relevant in SHM, reflecting bridge stiffness and bearing capability (Zhao et al. 2015). However, it is difficult to measure. With GBR, this data is accessible directly and accurately (Michel & Keller 2021a), making it a significant contender for NOR BWIM. Since deflection is challenging to measure, displacement-based BWIM has barely been studied. For instance, Ojio et al. (2016) have determined the bridge displacement with a camera. They use a second camera, not the displacement, to monitor the traffic and extract relevant features like the axle count. However, additional hardware increases the costs. Extracting traffic information from the displacement time series directly would be cheaper. Arnold & Keller (2020) and Arnold et al. (2021) show that Machine Learning (ML) based data-driven processing of bridge displacement signals is possible. They exploit a Random Forest (RF) and a CNN to detect and extract vehicle crossings successfully. As a next step, these events need to be analyzed regarding their potential for traffic characterization. Firstly, detected events should be differentiated between single- or multiple-presence events. One big challenge is that displacement, unlike strain, does not contain a localized bridge response. Furthermore, events and their corresponding time series are of varying length due to different vehicle numbers, speeds, and lengths. Varying-length time series classification is an active but insufficiently studied topic. One option is to extract features from a given time series (Bier et al. 2022). Other approaches like Random Convolutional Kernel Transform (ROCKET) (Dempster 2020) or MiniRocket (Dempster et al. 2021) achieve state-of-the-art results in public databases (Dau et al. 2019). In this paper, we will use a feature-based approach and MiniRocket to introduce an ML-based single-/multi-classification. More specifically, the main contributions of this paper are:

- a detailed description of the GBR data and the UAV dataset used in this study,
- a profound investigation of a varying-length bridge crossing classification in terms of single- versus multi-presence events using different ML approaches,
- an in-depth analysis of the potential of data augmentation for time series event classification,
- a comprehensive application of our approach to an entirely unknown dataset.

In Section 2, we will explain the setup for GBR measurements and describe the used dataset. Our methodology, including preprocessing, feature extraction and ML-models is detailed in Section 3. The results of our approach are laid out and discussed in Section 4. Finally, Section 5 summarizes the main findings of this study.

2 GBR MEASUREMENTS AND DATASET

In this section, we will briefly describe the GBR measurement setup. A more detailed explanation can be found in Michel & Keller (2021a), for example. Afterwards, we will introduce the UAV dataset for single-event classification.

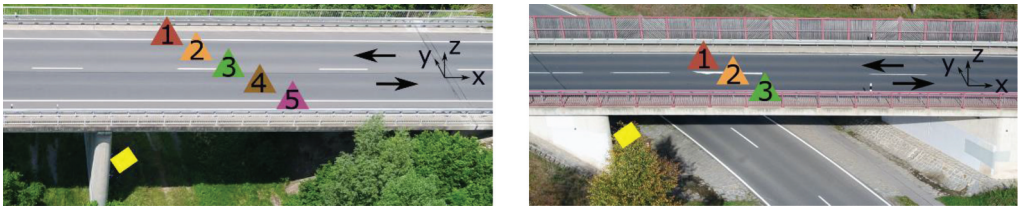

Figure 1. Overview of Bridge A (left) and Bridge B (right), including their reflectors.

2.1 *GBR measurement setup*

Measurements have been conducted at two bridges in Germany, henceforth referred to as Bridges A and B. To achieve a high signal-to-noise ratio five respectively three corner reflectors have been attached at the bottom of Bridge A and B. Both bridges and the approximate positions of the corner reflectors are shown in Figure 1. Bridge A has two fields, yet only one is monitored.

While Bridge A has undisturbed measurements, the vehicles using the street below Bridge B can cause significant disturbances in the time series data (Michel & Keller 2021a). On both infrastructures, one lane is present for each driving direction, as indicated by the black arrows. The yellow rectangle in Figure 1 represents the GBR. Its measuring principle is based on frequency modulation and interferometry (Gentile & Bernardini 2010). Modulation allows the GBR to measure multiple points along its line of sight (LOS). The maximum range resolution Δr amounts to 0.75m, according to

$$\Delta r = \frac{c}{2 \cdot B} = 0.75\ m, \tag{1}$$

with the speed of light $c = 3\times10^8$ ms^{-2} and a bandwidth B = 200 MHz. That is, every 0.75 m, the bridge displacement can be measured by summing up all reflected signals. For this purpose, all reflectors are spread along the x- and y-plane to make them distinguishable measurement points. The displacement is deducted using interferometry. With a sampling rate of 200 Hz, the phase shift $\Delta\varphi$ of each point is measured and then transformed to vertical displacement Δz using

$$\Delta z = \frac{\lambda}{4\pi}\frac{R}{\mathrm{h}} \cdot \Delta\varphi, \tag{2}$$

where h is the height difference between bridge and GRB, R is the distance of the GBR to a measurement point and λ is the GBR wavelength.

2.2 *Dataset*

To acquire ground truth data for the event classification, an unmanned aerial vehicle (UAV) has been deployed to film the bridge topside. During postprocessing, those videos have been synchronized with the radar data and used to build a database of vehicle crossings. Table 1 gives a general overview of the dataset for both bridges. As mentioned before, disturbances occur at Bridge B. We will only regard undisturbed events in this work. Overall, 1111 vehicles have been recorded for Bridge A and 730 for Bridge B. Since an event can contain several vehicles, the number of events is smaller, with 964 events for Bridge A and 500 for Bridge B, respectively. Besides the small dataset size, the imbalance of the dataset poses another challenge. This is true in two ways: Firstly, multi events are scarce compared to single events. Secondly, trucks appear less often than cars. Since trucks are inherently slower than cars and cars tend to cue behind a slow truck, trucks are underrepresented in single-events. In the case of Bridge A, for example, 241 trucks are single events, whereas 608 single events contain cars.

Table 1. Overview of the UAV-recorded events used for the event classification.

Bridge	Vehicles	Trucks	Cars	Events	Single Events	Multi Events
A	1110	341	769	962	849	112
B	491	114	377	343	271	72

While single events can differ significantly, multi-presence events come in an even greater variety: For instance, two heavy trucks might cross the bridge simultaneously in opposite directions. This scenario would be very relevant for SHM due to the significant stress induced on the bridge. Still, it is also challenging to visually differentiate from a single-event using only time series displacement data. To better understand, Figure 2 shows one single and one multi event for each bridge, although these examples do not cover the broad spectrum of traffic situations in the dataset. The corresponding time series data for all reflectors is drawn below each UAV

image. For each vehicle, its time on the monitored field is highlighted. Since there are overlapping areas in the lower plots, these are classified as multi events. Bridge A multi event shows a truck finishing its overtaking maneuver of another truck on the bridge field. The time series already resembles a single event of a very slow vehicle, e.g., when compared to the Bridge B single event. Thus, parallel crossings are barely distinguishable from single events. Attributing the maximum displacement or the maximum load of such an event falsely to one vehicle would lead to overestimating the bridge load and possibly to unnecessary reinforcements. A small car follows after the two trucks, but the vibration heavily masks the bridge bending the car causes. Finally, the Bridge B multi event shows a more straightforward example of two consecutive cars. The curious signal shape of RB3 in the Bridge B single event can be attributed to the fact that more than one displacement component is measured by the GBR (Michel & Keller 2021a).

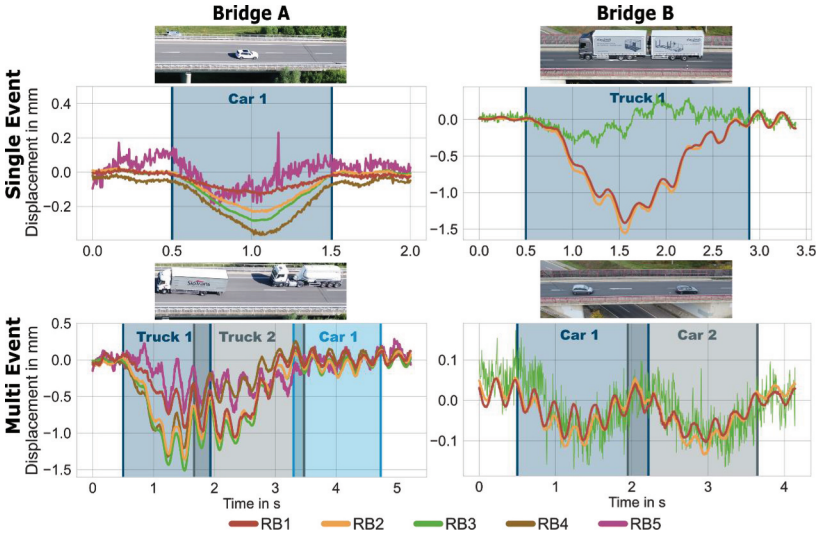

Figure 2. Example for a single (upper) and a multi event (lower) for Bridge A and Bridge B.

Figure 2 gives the impression that it could be sufficient to monitor the length of an event or its maximum displacement for one or two reflectors to classify it successfully. To illustrate that the distributions for both classes overlap significantly regarding such features, Figure 3 shows the duration of an event and the maximum displacement for reflector 3 of Bridge A and their corresponding Kernel Density Estimations (KDE). Although multi events tend to take longer, two consecutive cars are often faster than a single truck. And events of two trucks crossing the bridge in parallel on opposite lanes might last as long as a single truck event. On the other hand, most multi events have a higher deflection since slow and heavy trucks are involved, yet a clear decision boundary is not apparent. Although these two features alone are not enough to make a correct classification, additional features in combination with ML might be more successful.

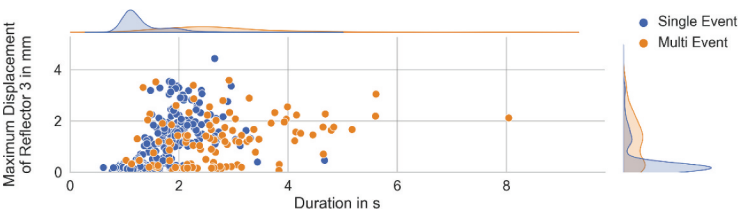

Figure 3. Maximum displacement and duration with normalized KDE.

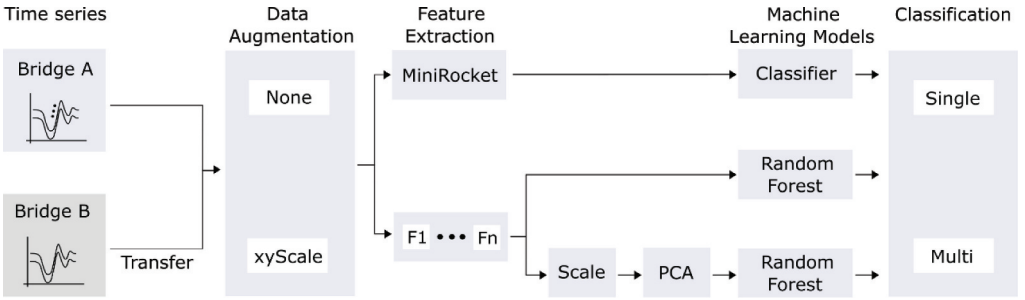

Figure 4. Schema of the methodological classification approach.

3 METHODOLOGY

In this chapter, we will shortly explain the methodology of the single- and multi-presence classification. Our proposed methodology consists of three steps: the data augmentation, the feature extraction and the ML models to classify the events. Figure 4 provides the schema of all steps.

3.1 *Preprocessing*

As evident from Section 2.1, bridges can have a varying amount of measurement points. For this study, we assume that at least two points are available at each bridge. Therefore, we use the two reflectors 2 and 4 for Bridge A and 1 and 3 for Bridge B, respectively. For comparison, we also train our models with all five reflectors of Bridge A. We split Bridge A data into training, validation, and test sets using a 70:15:15 ratio, with each subset mirroring the overall distribution of single and multi events. Finally, to investigate transferability, we did not train our models on the dataset of Bridge B.

In accordance with Ruiz et al. (2021), we try to minimize the amount of preprocessing, i.e., we did not normalize the time series. We ensure, that all time series start at a displacement of 0 mm by removing the offset. As described in Section 2.2, single events tend to have a smaller maximum deflection. To mitigate this bias, we apply data augmentation in the form of y-scaling. We choose five values in the range of -0.1 mm to -4.0 mm (see Figure 3) superimposed with random values, in which the maximum displacement of each event would be scaled. To tackle the bias in duration, we oversampled all single-events by factors 2 and 3 and downsampled all multi events accordingly. We combine both scalings (xyScale) and compare it to no augmentation (None).

3.2 *Feature extraction*

Both approaches are based on features extracted from the time series. The features of the RF classifiers are listed in Table 2, together with the functions to extract them. We have extracted every feature for each used reflector time series. With 14 features, this yields 28 features for two reflectors and 70 for five reflectors, respectively. We also train an RF after scaling the features and applying principal component analysis (PCA) to reduce the dimensionality. We use eight components since they explain over 95% of the variance for the None dataset.

Unlike manual feature selection, MiniRocket uses random convolutional kernels to extract 9,996 features itself. The properties and specifications of those kernels can be followed by Dempster et al. (2021). We use the sktime implementation (Löning et al. 2019), which can handle variable-length input. The final extracted feature of each convolution is the proportion of positive values (PPV). Theoretically, the number of kernels can be regarded as a hyperparameter. However, Dempster et al. (2021) recommend using 10000 kernels. Due to using the PPV and the bias from the convolution results, normalization is unnecessary.

Table 2. These 14 features have been extracted from the GBR time series data $\underline{x}$. For calculation, the Python-packages numpy and scipy have been used. All non-default values are stated.

Feature No.	Name of Feature	Basis of Calculation
1	Maximum	max($\underline{x}$)
2	Minimum	min($\underline{x}$)
3	Mean	mean($\underline{x}$)
4	Standard Dev.	min($\underline{x}$)
5	Skewness	skew($\underline{x}$)
6	Kurtosis	kurt($\underline{x}$)
7	Median	median($\underline{x}$)
8	Length	ken($\underline{x}$)
9	Quantile25	quantil($\underline{x}$, 0.25)
10	Quantile75	quantil($\underline{x}$, 0.75)
11	NbrPeaks	len(find_peaks($\underline{x}$, distance=4, width=5, rel_height=0.5))
12	xMinPosRatio	argmin($\underline{x}$)/len($\underline{x}$)
13	Power	sum($\underline{x}^2$)/len($\underline{x}$)
14	MAD	median_abs_deviation($\underline{x}$)

3.3 *Machine learning models*

Using the validation set, we apply grid search to the models in the second to last column in Figure 4 to find the optimal set of hyperparameters. Since the dataset is imbalanced, we use "balanced_accuracy" as the score for the grid search. Furthermore, we apply class weights during training. The features extracted from MiniRocket are passed to a classifier.

For small datasets with less than 10000 observations, like in the None case, a ridge classifier is recommended by Dempster et al. (2021). Otherwise, logistic regression, which we use for xyScale, is suggested. Since MiniRocket is independent of y-scaling, we might introduce overfitting to the classifier with our data augmentation approach.

4 RESULTS AND DISCUSSION

The study's objective is to investigate the potential of ML approaches to classify GBR variable-length time series vehicle crossings. Table 3 shows the results of the single vs. multi event classification concerning Overall Accuracy (OA), Precision (P), Recall (RC) and Balanced Accuracy (BA).

Concerning Bridge A with two reflectors, all models have an OA, P, and RC above 90%. However, these values are heavily skewed since single-events are much more present. Therefore, we included BA, the average of recall from each class. All BAs are close to each other, between 82% to 87%, with MiniRocket having the highest score. These results are deceptive, as there are correlations between duration and maximum displacement and event type, as described in Section 2.2. Figure 5 highlights this aspect. The decision boundary for MiniRocket in the left plot almost exclusively depends on the duration, with every event longer than 2.5 s classified as a multi event.

The boundary becomes more complex after applying xyScale data augmentation. Fewer long single-events are now misclassified as multi events, as seen in Figure 5 on the right. Conversely, MiniRocket has more difficulties with medium-length single-events. The effect of xyScale also shows in the BA of all three models. RF and PCA RF achieve a worse BA since they have to find more complex relationships. MiniRocket improves with data augmentation, achieving the best results for two reflectors on one bridge, with a BA of 90%. Considering the small size of the dataset and without providing any prior domain knowledge, these are very auspicious results. Figure 5 also shows that short multi events are still misclassified as single-events. Heavier augmentation in time might solve this issue. When using five reflectors, we did not expect any additional information in those time series, but on the contrary, to achieve worse results due to the higher complexity. Accordingly, RF and MiniRocket achieve equal or worse results. Interestingly enough, PCA RF has better results than with only two reflectors. One explanation could be that the driving side is more distinct with more data, which can help detect multiple events.

Finally, the tests on Bridge B show that a transfer to an unknown dataset can be successful to a certain degree. With None augmentation, all models suffer again under the correlation depicted in Figure 5. The BA is even 90% and above. This score is partially due to the more extensive dataset, where usual events, e.g., cars following a truck, are more prominent, whereas complex events, such as two trucks simultaneously, are rare. The transfer seems challenging for RF and MiniRocket trained with xyScale, since even OA and RC are around 50%. Only PCA RF has values of over 80% and even a BA of 74%. As with five reflectors for Bridge A, the PCA appears to extract robust features. Considering that both bridges have unequal numbers of fields and are of different size, this is a promising result regarding transferability.

Table 3. Test results for all classificators.

Bridge	Number of Reflectors	Data Augmentation	Model	OA	P	RC	BA
A	2	None	RF	0.91	0.97	0.93	0.85
			PCA RF	0.94	0.95	0.98	0.82
			MiniRocket	0.95	0.97	0.98	0.87
		xyScale	RF	0.90	0.94	0.93	0.76
			PCA RF	0.90	0.95	0.94	0.79
			MiniRocket	0.98	0.98	0.97	0.90
	5	xyScale	RF	0.94	0.94	0.99	0.76
			PCA RF	0.95	0.97	0.98	0.87
			MiniRocket	0.94	0.96	0.97	0.84
B		None	RF	0.96	0.97	0.98	0.94
			PCA RF	0.92	0.97	0.93	0.90
			MiniRocket	0.95	0.97	0.96	0.93
		xyScale	RF	0.59	0.92	0.57	0.69
			PCA RF	0.84	0.89	0.92	0.74
			MiniRocket	0.59	0.97	0.50	0.72

5 CONCLUSION

This paper introduced a data-driven ML approach to classifying multi-variate GBR time series from bridge vehicle crossings. The GBR data originates from two bridges in Germany, which a UAV also monitors as ground truth. One bridge has been used for training, validation, and testing, whereas the second bridge only served as a test set to evaluate transferability. We also investigated the effect of varying measurement points as input. The classification goal is to differentiate between single- and multi-presence events. Particularly challenging is the variable length of the time series and the correlation of the length with a class. For this purpose, we have examined the potential of data augmentation. Methodically, we implement and apply three different ML approaches. We use an RF with 14 hand-crafted features per reflector as input, an RF that uses the output of a PCA as input. Finally, we compare their performance to MiniRocket, a state-of-the-art ML model in time series classification. It shows that MiniRocket outperforms the other models when applying data augmentation to tackle biases. With a BA of 90%, it achieves satisfying results.

Concerning transferability to an utterly unknown bridge, MiniRocket needs to catch up in BA compared to PCA RF and has only a BA of 74%. Keeping the second bridge out of the training cycle allowed us to test the transferability and reduced the training size, making it more challenging for the models. In the future, we will monitor further bridges from which the models can more easily extrapolate to new structures. Also, pre-training in combination with DL could be exploited to widen our dataset.

In sum, we showed that a data-driven classification of GBR data is possible. These promising results can link an event detection, as described by Arnold et al. (2021), and a GBR-based BWIM, which focuses, in the first step, on single-events. With this, a non-invasive and remote BWIM and SHM would be possible.

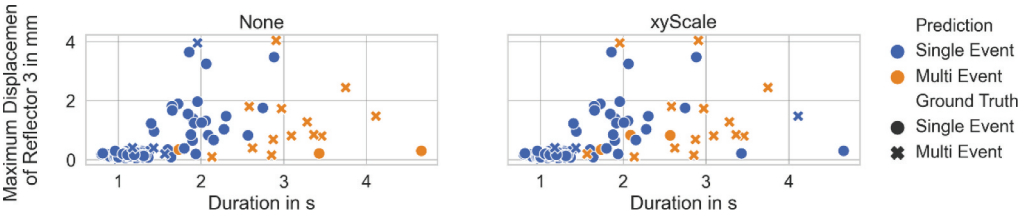

Figure 5. Test set classification results for MiniRocket with None and xyScale data augmentation.

REFERENCES

Arnold, M., Hoyer, M., & Keller, S. 2021. Convolutional Neural Networks for Detecting Bridge Crossing Events With Ground-Based Interferometric Radar Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science*s 1-2021, 31–38.

Arnold, M. & Keller, S. 2020. Detection and Classification of Bridge Crossing Events With Ground Based Interferometric Radar Data and Machine Learning Approaches. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial In formation Sciences* 12020, 109–116.

Bier, A., Jastrzebska, A., & Olszewski, P. 2022. Variable-Length Multivariate Time Series Classification Using ROCKET: A Case Study of Incident Detection. *IEEE Access* 10, 95701–95715.

Dau, H. A., Bagnall, A., Kamgar, K., Yeh, C.-C. M., Zhu, Y., Gharghabi, S., Ratanamahatana, C. A., & Keogh, E. 2019. The UCR Time Series Archive. *arXiv:1810.07758* [cs, stat].

Dempster, A. 2020. ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery* (34), 42.

Dempster, A., Schmidt, D. F., & Webb, G. I. 2021. MINIROCKET: A Very Fast (Almost) Deterministic Transform for Time Series Classification. *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 248–257)

Gentile, C. & Bernardini, G. 2010. An interferometric radar for non-contact measurement of deflections on civil engineering structures: laboratory and full-scale tests. *Structure and Infrastructure Engineering* 6(5), 521–534.

Ieng, S.-S. 2015. Bridge Influence Line Estimation for Bridge Weigh-in-Motion System. *Journal of Computing in Civil Engineering* 29(1), 06014006.

Kawakatsu, T., Aihara, K., Takasu, A., Nagayama, T., & Adachi, J. 2023. Data-Driven Bridge Weigh-In-Motion. *IEEE Sensors Journal*, 1–1. Conference Name: IEEE Sensors Journal.

Löning, M., Bagnall, A., Ganesh, S., & Kazakov, V. 2019. sktime: A Unified Interface for Machine Learning with Time Series.

Michel, C. & Keller, S. 2021a. Advancing Ground-Based Radar Processing for Bridge Infrastructure Monitoring. *Sensors* 21(6), 2172. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Michel, C.& Keller, S. 2021b. Introducing a non-invasive monitoring approach for bridge infrastructure with ground-based interferometric radar. In EUSAR 2021; *13th European Conference on Synthetic Aperture Radar*, pp. 1–5.

Michel, C. & Keller, S. 2022. Determining and Investigating the Variability of Bridges' Natural Frequencies with Ground-Based Radar. *Applied Sciences* 12(11), 5354.

Moghadam, A., Al Hamaydeh, M., & Sarlo, R. 2023. Dual-purpose procedure for bridge health Monitoring and weigh-in-motion used for multiple-vehicle events. *Automation in Construction* 148.

OBrien, E. J., Quilligan, M. J., & Karoumi, R. 2006. Calculating an influence line from direct measurements. *Proceedings of the Institution of Civil Engineers - Bridge Engineering* 159(1), 31–34.

Ojio, T., Carey, C. H., OBrien, E. J., Doherty, C., & Taylor, S. E. 2016. Contactless Bridge Weigh-in-Motion. *Journal of Bridge Engineering* 21(7), 04016032.

Paul, D. & Roy, K. 2023. Application of bridge weigh-in-motion system in bridge health monitoring: a state-of-the-art review. *Structural Health Monitoring*, 147592172311544.

Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., & Bagnall, A. 2021. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery* 35(2), 401–449.

Yu, Y., Cai, C., & Deng, L. 2016. State-of-the-art review on bridge weigh-in-motion technology. Advances in *Structural Engineering* 19(9), 1514–1530. Publisher: SAGE Publications Ltd STM.

Zhao, X., Liu, H., Yu, Y., Xu, X., Hu, W., Li, M., & Ou, J. 2015. Bridge Displacement Monitoring Method Based on Laser Projection-Sensing Technology. *Sensors* 15(4), 8444–8463.