Machine Learning Approaches for Vehicle Counting on Bridges Based on Global Ground-Based Radar Data

Matthias Arnold[1], Sina Keller[2]

[1] ci-tec GmbH, 76139 Karlsruhe, Germany - m.arnold@ci-tec.de
[2] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany - sina.keller@kit.edu

Keywords: Unequal Time-Series Data, Bridge Displacement, Bridge Monitoring, MiniRocket, Tree-Based Approaches, Feature Extraction

Abstract

This study introduces a novel data-driven approach for classifying and estimating the number of vehicles crossing a bridge solely on non-invasive ground-based radar time series data (GBR data). GBR is used to measure the bridge displacement remotely. It has recently been investigated for remote bridge weigh-in-motion (BWIM). BWIM mainly focuses on single-vehicle events. However, events with several vehicles should be exploited to increase the amount of data. Therefore, extracting the number of involved vehicles in the first step would be beneficial. Acquiring such information from global bridge responses such as displacement can be challenging. This study indicates that a data-driven machine learning approach can extract the vehicle count from GBR time series data. When classifying events according to the number of vehicles, we achieve a balanced accuracy of up to 80 % on an imbalanced dataset. We also try to estimate the number of cars and trucks separately via regression and acquire a R^2 of 0.8. Finally, we show the impact of the data augmentation methods we apply to the GBR data to tackle the skew in the dataset using the feature importance of Random Forests.

1. Introduction

Structural Health Monitoring (SHM) in bridge maintenance consists of numerous aspects. Its primary focus is monitoring the bridge's structural integrity, e.g., by detecting deterioration. Besides, it is vital to observe the load inflicted on the bridge. Heavy and significantly overweight traffic can cause permanent damage to infrastructures. With this in mind, vehicle weighting systems have been developed and applied for a long time. These systems range from simple static weighing scales to bridge weigh-in-motion (BWIM) approaches. While static weighing has a high accuracy, it interferes with the traffic flow and can easily be bypassed by vehicles. By implementing conventional acceleration or strain sensors in the road, traffic can flow uninterrupted; however, weighting is more complex. Additionally, the sensors are exposed to heavy loads, and maintenance can be challenging. BWIM systems use the bridge itself as a scale to determine the weight of a vehicle. Therefore, sensors are attached to the lower side of the bridge. The attachment inflicts damage to the bridge. Furthermore, BWIMs still often use sensors on the pavement to extract vehicle information such as axle count and speed. Nothing-on-road (NOR) solutions try to eliminate all pavement components to minimize the impact of traffic on the system, increasing durability and accessibility. Therefore, the traffic situation on the bridge has to be monitored differently. One possibility is to exploit cameras. Ojio et al. (2016) use two cameras for bridge monitoring; one camera is aimed at the road for traffic monitoring and the other for remote bridge displacement measurements. Their approach is purely contactless, avoiding any damage through sensor attachment.

Another approach for remote and non-invasive bridge monitoring applies ground-based radars (GBR) Gentile and Bernardini (2010); Michel and Keller (2021b, 2022, 2024). GBR measures the bridge deflection or displacement, which is relevant in SHM Zhao et al. (2015); Pieraccini et al. (2019); Döring et al. (2021). They can be easily set up and require little maintenance. Thus, measurement campaigns can be performed without much effort. Furthermore, they can be operated under difficult natural conditions like fog. Such foggy conditions challenge visual surveillance systems. However, relying solely on the GBR time series displacement data for bridge monitoring requires sophisticated algorithms.

Arnold and Keller (2020) and Arnold et al. (2021) exploit machine learning (ML) approaches, such as shallow learning and deep learning (DL), to detect bridge vehicle crossings in GBR displacement time series data. Arnold and Keller (2024a) investigate several approaches to extract vehicle properties such as speed and axle spacing from detected events. They only investigate single events, meaning that only one vehicle is present on the bridge during an event. It is possible to distinguish between single and multi-events during which several vehicles cross the bridge simultaneously with different ML approaches, as stated by Arnold and Keller (2024b). Although this reduces the complexity of their algorithms, they also forgo a lot of data points when neglecting multi-events. The parallel passing of two overweight trucks would, thus, be ignored despite its immense impact on the bridge. Finally, not all bridges are short or rarely frequented enough to generate sufficient single- presence data points for bridge behavior analysis. Therefore, multi-presence events should also be taken into consideration. A more detailed first-hand analysis of the traffic situation is crucial to do so. Thus, as a first step, we propose a novel approach to extract the vehicle count of bridge crossing events purely data-driven from displacement data. We use measurements of two bridges in Germany to collect ground-truth data on crossing vehicles using an unmanned aerial vehicle (UAV). Overall, the main contributions of this paper are:

- a concise introduction to GBR measurements,
- a profound description of the dataset extracted from GBR

https://doi.org/10.5194/isprs-annals-X-2-2024-1-2024

and UAV data,

- an appropriate ML framework with three models to extract vehicle counts via (1) classification and (2) regression,
- a comprehensive investigation regarding the application of data augmentation for both tasks,
- and an in-depth analysis of the feature importance of our applied models.

First, we introduce relevant work concerning traffic classification in Section 2. Then, we describe the measurement setup and the dataset in Section 3. In Section 4, our approach is detailed. Afterward, the results are stated and discussed in Section 5. Finally, Section 6 summarizes our findings.

2. Related Work

In this section, we give an overview of relevant studies. First, we will deal with time series classification, focusing on unequal length time series data. Subsequently, we cover approaches for data-driven traffic identification with a focus on vehicle count determination.

Since most ML approaches are developed with images as input in mind, only a few are compatible with input of varying lengths. However, as Arnold and Keller (2024b) have shown, unequal time series are present in bridge monitoring due to varying speeds and vehicle lengths. Furthermore, the length correlates with class affiliation so that padding might skew the results. Therefore, they apply MiniRocket, among others, to handle variable-length datasets. MiniRocket was designed by Dempster et al. (2021) and is currently among the state-of-the-art methods on public time series datasets, such as the UCR dataset (Dau et al., 2019). The developers also provide source code, which can handle variable-length input (Löning et al., 2019). Another approach is to extract features manually using expert knowledge (Arnold and Keller, 2024b) or automatically via methods such as the auto-regressive integrated moving average model (ARIMA) (Wang and Tang, 2020). Dynamic time warping (DTW) is often used as a baseline with a 1-nearest neighbor classifier (Ruiz et al., 2021). DTW aligns two-time series along their temporal axis, minimizing the Euclidean distance between both. It is particularly suitable for sequences that vary in speed, e.g., speech or movement recognition (Tan et al., 2019). A disadvantage of DTW is its extensive computation time, especially for relatively long time series. DL can be applied for feature extraction when many data points are available. Hertel et al. (2016) use convolutional neural networks (CNN) combined with masking and padding. They also implement a global pooling layer, which can handle the filtered sequences of different sizes for acoustic scene classification.

Many approaches for vehicle counting systems also apply CNNs, but they use images from surveillance systems as input data (Fachrie and others, 2020; Gomaa et al., 2022). However, visual techniques require an unobstructed view and might fail, e.g., in the event of snowfall or fog. Therefore, other methods are researched. Taghvaeeyan and Rajamani (2014) exploit four magnet sensors at the roadside for vehicle counting and classification. Passing vehicles induce peaks in the measured time series data and with a simple threshold 186 of 188 vehicles could be detected. Arnold and Keller (2024b) use MiniRocket and manually crafted features for Shallow Learners to classify single and multi-events. MiniRocket achieves a balanced accuracy (BA) of 87 % on a single bridge dataset. They also

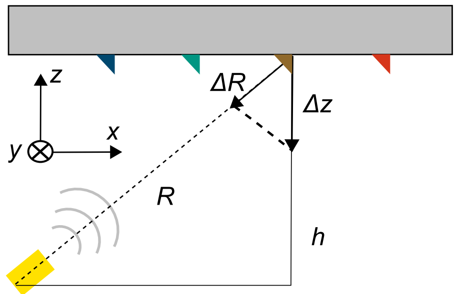

Figure 1. Schematic illustration of a GBR measurement. The GBR (yellow box) measures the LOS distance for each reflector (colored rectangles). From consecutive measurements of the phase difference $\Delta\phi$, the displacement difference ΔR can be calculated. With a known height difference h between bridge and GBR ΔR can be projected to the vertical displacement Δz.

find that data augmentation can improve the BA to 90 %. Finally, they apply their models to a completely unknown dataset, where MiniRocket achieves a BA of 93 %. While distinguishing between single and multi-presence events, they make no statement on the vehicle count.

3. GBR Measurements and Dataset

In this section, we will shortly explain the working principle of GBRs and show some real-world measurements. For example, a more detailed explanation can be found in Michel and Keller (2021a). Then, we will describe the data basis for this study.

GBR uses frequency modulation to measure the displacement in the line of sight (LOS). They utilize a bandwidth of B = 200 MHz resulting in a range resolution of

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

with the speed of light c = $3 \times 10^8\,\mathrm{m\,s^{-1}}$. Thus, the displacement is measured for every 0.75 m. For this, GBR exploits interferometry. Figure 1 gives a schematic illustration of this measurement principle. The final vertical displacement Δz can be calculated according to

$$\Delta z = \frac{R}{h} \cdot \frac{\lambda}{4\pi} \cdot \Delta\phi. \tag{2}$$

Δz is sampled with 200 Hz for each measurement point along LOS. The driving direction corresponds to the x-axis. Often, reflectors are spread along the y-axis to acquire information about the lateral vehicle position.

Measurements have been conducted at two short-span bridges in Germany. Bridge A has two loosely coupled fields, from which one has been monitored, whereas Bridge B only has one. One lane for each driving direction is present for both bridges. As a road runs beneath Bridge B, disturbances occur caused by passing vehicles. Both have been equipped with corner reflectors to achieve low-noise measurements. The monitored field of Bridge A has five reflectors attached; Bridge B has three reflectors. During the GBR measurements, a UAV was deployed to record the bridge deck as ground-truth data. Figure 2 shows example events for each bridge. All crossing vehicles are shown above the corresponding time series data. Only the displacements for two reflectors are shown since we only use these

https://doi.org/10.5194/isprs-annals-X-2-2024-1-2024

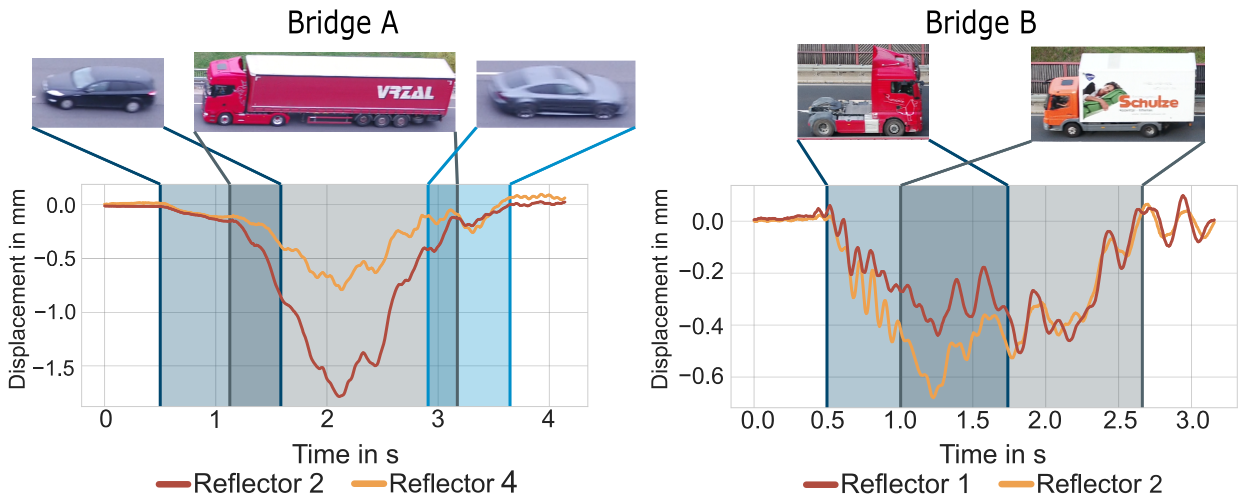

Figure 2. For each bridge, a multi-event is shown. In the time series plot, the respective time of each vehicle on the monitored field is highlighted and attributed to the corresponding vehicle images.

for our datasets. This enables transferability, assuming at least two measurement points per bridge are available. The reflectors are selected such that the driving side of a vehicle is apparent from the relative maximum displacement, and the overall signal curvature for both bridges is similar.

The Bridge A event in Figure 2 depicts two cars and a truck. The first car overtakes the truck. The second car drives in the direction opposite to them. Visually, this event could pass as a single event. Arnold and Keller (2024b) indicate that it is possible to classify this event as a multi-event using ML. However, it would be beneficial to know the number of vehicles included in an event to extract vehicle properties from multi-event with approaches introduced in, e.g., Arnold and Keller (2024a). Even more so if the vehicle count would be split into the number of trucks and cars. As cars usually have two axles, this information can be helpful to determine the truck properties more accurately. The example of Bridge B visualizes two trucks driving in opposite directions (see Figure 2). Although it can be recognized as a multi-event, the number of vehicles is not also apparent. The displacement time series data also contains the bridges’ oscillation in their respective first natural frequency of approximately $3.66\,\mathrm{Hz}$ for Bridge A and $3.75\,\mathrm{Hz}$ for Bridge B (Arnold et al., 2021).

Figure 2 shows that events can come in a significant variance regarding vehicle type, count, order, and driving direction. A superficial analysis of the used dataset is stated in Table 1. As mentioned, disturbances caused by passing vehicles can occur at Bridge B. Michel and Keller (2021a) show that they can be detected; therefore, we only consider undisturbed events. With 1118 events, the dataset is relatively small overall. Single events are more frequent than multi-events, and cars form the majority of registered vehicles overall. The highest number of vehicles during one event amounts to 5 for Bridge A and 7 for Bridge B, respectively. However, events with more than five vehicles rarely occur; therefore, we cluster these events in a 4+ class.

Multi-events often consist of several vehicles in a series. Cars would, e.g., queue behind a slower truck. Figure 3 illustrates the skew in the dataset. It depicts the distribution of event duration grouped into single and multi-events for each bridge. Most of the time, single events are shorter than multiple events, especially for Bridge B. Thus, the duration can indicate the number of vehicles on the bridge. Yet, both distributions overlap, so additional features need to be found. Furthermore, Figure 3 highlights the dataset imbalance, as there are many more single-vehicle events than multi-events.

4. Methodology

This section will describe our methodological approach as depicted in Figure 4. First, we will describe preprocessing and data augmentation. Afterwards, the feature extraction step is explained. Finally, our ML models concerning our classification and regression tasks are introduced. For reasons of comparison, we orient us at the methodology of Arnold and Keller (2024b).

4.1 Preprocessing and Data Augmentation

We reduce the time series data preprocessing to a minimum. The only step we apply is to remove the offset for each sequence. This is necessary as a long-term drift occurs during GBR measurement campaigns (Arnold and Keller, 2020). We also do not remove bridge oscillations as they often correlate with the presence of heavy vehicles.

We test all our approaches on two event types. First, we use all available events as long as an event is in a not disturbed mode. In a second step, we build on the results of Arnold and Keller (2024b). Assuming we can distinguish between single and multi-events, as they indicate, it would be enough only to consider multi-events for our tasks. Even though using only multi-events instead of all events drastically reduces the size of the dataset, it also reduces its imbalance.

For data augmentation, we use a combination of x- and y-scaling (`xyScale`) and no augmentation at all (`None`). Apart from increasing our dataset, we try to tackle its skew as described in Section 3 with data augmentation. x-Scaling corresponds to down- and oversampling along the temporal axis. For events with one vehicle, we oversample each event by factors two and three, as single events are often shorter. Conversely, we downsample events with more than one vehicle by the same amount. y-Scaling is done afterward for all sequences by rescaling them to a range of $0.1\,\mathrm{mm}$ to $4.0\,\mathrm{mm}$ (Arnold and Keller,

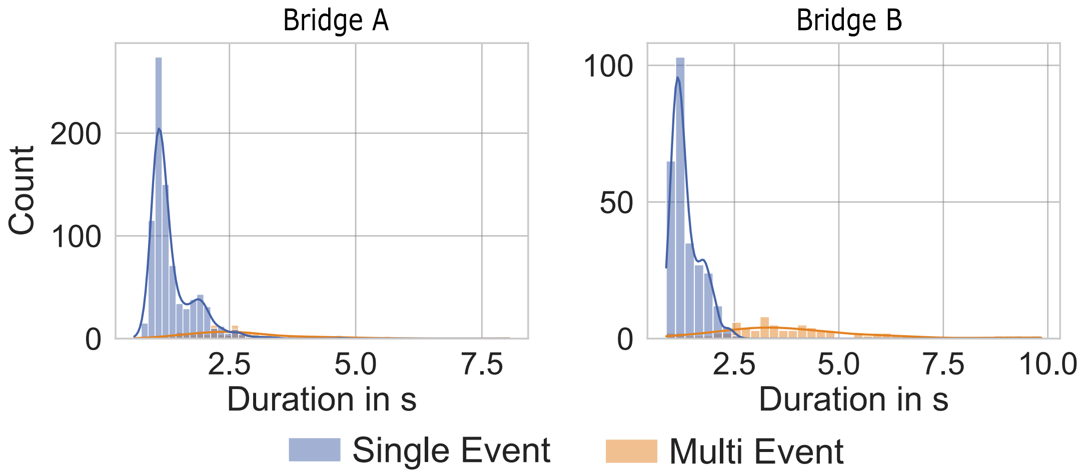

Figure 3. Distribution of event duration for both bridges grouped by single and multi-events. A more detailed breakdown of multi-event into vehicle count classes has been omitted for presentation reasons.

Table 1. Overview of used dataset.

Structure	Number of cars	Number of trucks	`1 vehicle`	`2 vehicles`	`3 vehicles`	`4+ vehicles`
Bridge A	112	769	849	84	24	5
Bridge B	337	114	271	35	22	15

2024b). From each sequence, five new sequences are generated through y-scaling.

4.2 Feature Extraction

Three different approaches for feature extraction are evaluated. First, we use manually developed features listed in Table 2. They are extracted for each reflector time series. Thus, with 14 features and two used reflectors, we generate 28 features per event. We also apply scaling and principal component analysis (PCA) as a second approach to reduce the dimensionality. We transform our 28 features onto eight components since they explain over 95 % of the variance for the Bridge A dataset without data augmentation (Arnold and Keller, 2024b).

Finally, we use MiniRocket for automatic feature extraction. MiniRocket applies 9996 convolutional kernels to the time series and, from the results of the convolution, then calculates the proportion of positive values (PPV) with

$$\mathrm{PPV}(Z) = \frac{1}{n} \sum_{i=0}^{n-1} [z_i > 0]. \tag{3}$$

More details can be found in Dempster et al. (2021). We use the implementation of Löning et al. (2019) since it can handle variable-length input.

4.3 Machine Learning Models

We investigate two different tasks in this study: **Task (1)**, the classification of events according to the number of vehicles within, and **Task (2)**, the prediction of cars and trucks within an event. The first task is treated as a multi-class classification task with the classes `1 vehicle`, `2 vehicles`, `3 vehicles`, and `4+ vehicles`. We implement the second task as a multi-output regression by predicting a value for both car and truck. Since regression can produce floating numbers, but vehicles only occur in natural numbers, we round the prediction to the nearest integer.

The manually crafted features are inputted to a Random Forest (RF). Concerning MiniRocket and **Task 1**, we follow the suggestion of Dempster et al. (2021) to use a logistic regression model for `xyScale` augmentation as there are more than 10 000 samples in the training set and a ridge model otherwise, meaning for `None` augmentation. However, in the case of the regression task, we will always utilize a ridge regression model.

We apply grid-search with 5-fold cross-validation for hyperparameter optimization during training. We split our dataset in a 80 : 20 manner for training and testing. Due to class imbalance in **Task (1)**, we apply class weights and use “balanced_accuracy” as the score for a grid search. The mean absolute error (MAE) is the scoring metric for the regression task. We do not optimize the hyperparameters for our MiniRocket approach. Instead, we follow the suggestions of Dempster et al. (2021) and use default parameters otherwise.

5. Results and Discussion

This study aims to investigate the potential of ML to count vehicles in GBR bridge crossing events. In this section, we will state and discuss the results of all approaches. The classification task models are evaluated based on balanced accuracy (BA), overall accuracy (OA), precision (P), and recall (RC). Regression performance is expressed by the coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE). The results for both tasks have been combined in Table 3.

Regarding **Task 1**, the classification of events according to their number of contained vehicles, RF with no data augmentation outperforms the other models with a BA of 80.4 %. Its confusion matrix can be seen in Figure 5. MiniRocket comes second with a BA 70.2 %. Regarding all events, OA, P, and RC are very high for all models independent of the data augmentation method. This is mainly due to the imbalance of the dataset.

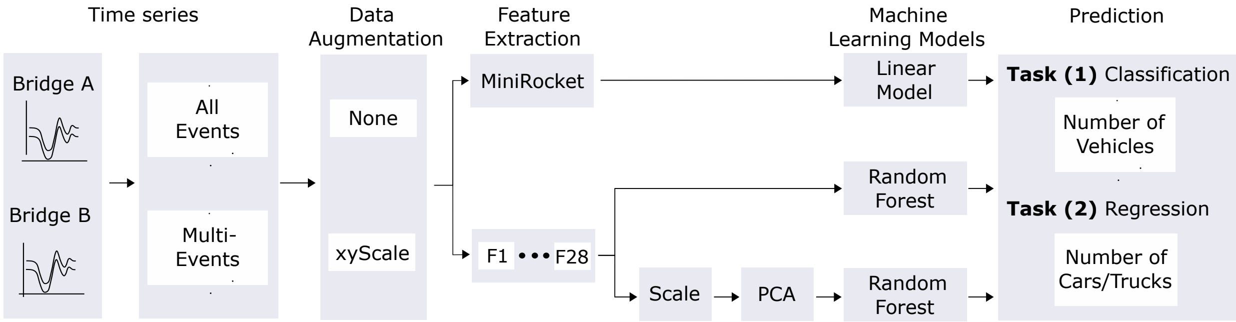

Figure 4. Schema of our methodological approach.

Table 2. These 14 features have been extracted from the GBR time series data $\underline{x}$. The Python packages numpy and scipy have been used for calculation. All non-default values are stated. They correspond to the features used in Arnold and Keller (2024b).

Feature No.	Name of Feature	Basis of calculation
1	Maximum	$\max(\underline{x})$
2	Minimum	$\min(\underline{x})$
3	Mean	$\min(\underline{x})$
4	Standard Dev.	$\text{mean}(\underline{x}^2)$
5	Skewness	$\text{mean}(\underline{x}^3)$
6	Kurtosis	$\text{mean}(\underline{x}^4)$
7	Median	$\text{median}(\underline{x})$
8	Length	$\text{len}(\underline{x})$
9	Quantile25	$\text{quantile}(\underline{x}, 0.25)$
10	Quantile75	$\text{quantile}(\underline{x}, 0.75)$
11	NbrPeaks	$\text{len}(\text{find\_peaks}(\underline{x}, \text{distance} = 4, \text{width} = 5, \text{rel\_height} = 0.5))$
12	xMinPosRatio	$\text{argmin}(\underline{x})/\text{len}(\underline{x})$
13	Power	$\text{sum}(\underline{x}^2)/\text{len}(\underline{x})$
14	MAD	$\text{median\_abs\_deviation}(\underline{x})$

Table 3. Test results for all **Task 1: Classification** and **Task 2: Regression**. The best results concerning BA respectively R^2 for each event type are highlighted.

Event Type	Data Augmentation	Model	Task 1				Task 2		
			OA in %	P in %	RC in %	BA in %	R^2	MSE	MAE
All events	None	RF	95.1	94.7	95.0	**80.4**	0.74	0.14	0.10
		PCA RF	92.7	92.0	92.7	66.4	0.75	0.14	0.11
		MiniRocket	93.5	93.0	93.4	70.2	**0.80**	0.11	0.08
	xyScale	RF	92.3	92.1	92.3	63.8	0.35	0.25	0.22
		PCA RF	88.5	90.2	88.5	67.9	0.50	0.20	0.18
		MiniRocket	90.8	90.1	90.8	62.3	0.48	0.21	0.19
Multi-events	None	RF	81.1	81.1	81.1	67.1	0.64	0.65	0.30
		PCA RF	70.2	70.9	70.2	50.0	**0.66**	0.54	0.30
		MiniRocket	72.9	78.4	72.9	58.3	0.53	0.69	0.36
	xyScale	RF	70.3	70.0	70.3	66.2	0.31	0.68	0.46
		PCA RF	70.2	68.7	70.2	56.9	0.21	0.73	0.54
		MiniRocket	78.4	78.9	78.4	**72.7**	0.25	0.77	0.55

Models tend to predict single events because the dataset is heavily skewed towards this class. While both RF and MiniRocket decrease in BA for `xyScale` compared to `None`, PCA RF performs similarly in both cases. This indicates that scaling adds no new information that linear transformation methods can extract, but also, the PCA makes this pipeline more robust. Overall, the PCA RF BA improves slightly when data augmentation is applied. Conversely, RF decreases in BA for `xyScale`.

Using only multi-events in **Task 1** leads to considerably worse results, especially OA, P, and RC. Since no single events are present in the dataset, it is much more balanced than all events. Therefore, the models cannot simply predict one vehicle. MiniRocket outperforms all other models for this event type and `xyScale` with a BA of 72.7 %. It even improves its BA compared to all events. Figure 6 shows its confusion matrix.

PCA RF always has the worst BA except for all events `xyScale`. Arnold and Keller (2024b) state that when only training on one bridge and predicting on another one, PCA RF achieved the best results for single- vs multi-presence classification. However, with two bridges in the training dataset, PCA RF seems to have issues extracting helpful features. One reason could be that in the Bridge B dataset, events with several vehicles take considerably longer than single events (see Figure 3), making it difficult to generalize. Data-driven models, in general, might learn to rely mainly on the signal length (Feature 8 in Table 2).

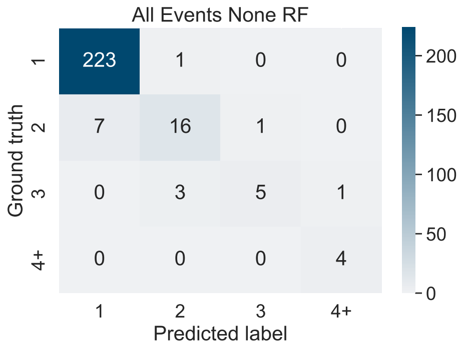

Figure 5. Confusion matrix for RF with all events and without data augmentation.

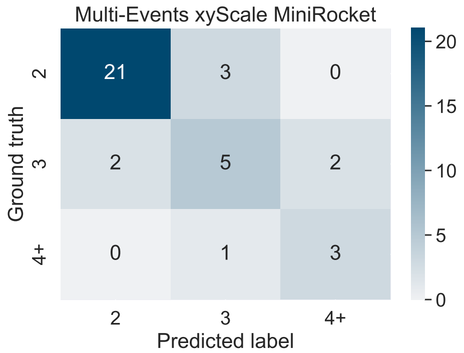

Figure 6. Confusion matrix for multi-events MiniRocket with data augmentation.

For a more detailed analysis of our extracted features and the impact of data augmentation, we show the feature importances for **Task 1** of selected features in Figure 7. RF trained without data augmentation rate the length (see Length_1 and Length_2) of a signal highly. A single length feature would have sufficed since the length is the same for both reflector time series of one event. However, since a PCA does not considerably improve an RF's performance, we have refrained from doing so. Both length features have a high importance. The overall length rating can be understood as the combination of those two. This coincides with our observations from Section 3 regarding the duration skew of the dataset. ML approaches trained using `xyScale` place less value on the length. This suggests that our intention of balancing the duration skew through scaling was successful. Overall, their importance is spread more homogeneously over several features. They value, e.g., skewness and kurtosis more highly. Also, "xMinPoRatio" has a high feature importance. It describes the position of the global minimum within an event relative to the signal length. For example, a small "xMinPoRatio" means that the highest deflection happens early in an event, as it might be caused by a truck followed by several cars. Interestingly, the feature importance of "Minimum" is less than 0.02 for all models. Intuitively, a high minimum is caused by heavy vehicles, which are often involved in multi-presence events due to their speed. This feature seems to be of less importance, though, when determining vehicle count.

The results of **Task 2** are also stated in Table 3. This task aims to estimate the number of cars and trucks in an event separately using multi-output regression. We have rounded the regression results to the nearest natural number before evaluation. This might skew our results. However, it is closer to reality. MiniRocket achieves the best overall results with a R^2 of 0.8 using all events without data augmentation. In this setup, RF and PCA RF have comparable results. Also, MSE and MAE are close to each other, indicating that the number of prediction outliers is small. PCA RF achieves the best results after removing single events from the dataset with a R^2 of 0.66. RF performs comparably. However, MiniRocket falls considerably behind with a R^2 of 0.53. It is barely better than predicting the mean of the dataset, which is the case for an R^2 of 0.5. The decrease in performance of all models between all events and multi-events indicates that the imbalance of the dataset regarding the overwhelming majority of one-vehicle events leads to good results for all events. Also, both MSE and MAE increase due to this aspect.

Regarding the effect of data augmentation, it shows that the performance drops heavily when using the `xyScale` dataset. For all events and only multi-events, the R^2 for all models is smaller or equal to 0.5. MiniRocket only achieves a R^2 of 0.25 for multi-events with data augmentation. This suggests that the length of a signal is an essential factor for vehicle count regression. These observations coincide with those from **Task 1**. However, the impact seems stronger for **Task 2**.

Figure 8 shows the feature importances of selected features for the regression task. Similar to the classification, our data augmentation reduces the relevance of the signal length. However, there is a more significant difference in the importance of the length between all events and multi-events with data augmentation than for **Task 1**. With data augmentation, the same features are more critical, such as "skewness", "kurtosis" and "xMinPosRatio" showing concise behavior. Overall, the ML approaches use considerably more features compared to task 1. The similarity in the feature importance and the broader spread of relevant features indicates that regression is a more complex challenge despite both tasks being related.

6. Conclusion

In this study, we discuss ML approaches for data-driven determination of vehicle count from GBR bridge displacement time series data. Two bridges in Germany have been monitored over several days. Together with UAV data for ground truth, a database has been built. With this database, we investigate the potential ML approaches to (1) classify events according to the number of involved vehicles and (2) extract the exact number of cars and trucks in an event via regression. To this end, we use two measurement points per bridge. To simplify the classification task, we group all events with four or more vehicles in one class. We investigate the effect of applying data augmentation and using only multi-events in the dataset.

Methodologically, we implement three different ML approaches, which can handle variable-length time series data. First, we manually craft features extracted from the time series data and pass it to an RF. Second, we scale those features and reduce their dimensionality via PCA before passing the resulting eight components to an RF. Finally, we exploit MiniRocket, which extracts features by applying convolutional kernels to the time series data. It shows that RF achieves the best results for the classification with a BA of $80.4\,\%$. However, this is mainly due to a data set heavily imbalanced towards events with only one vehicle. When removing these events, MiniRocket outperforms all other models when applying data augmentation. It can

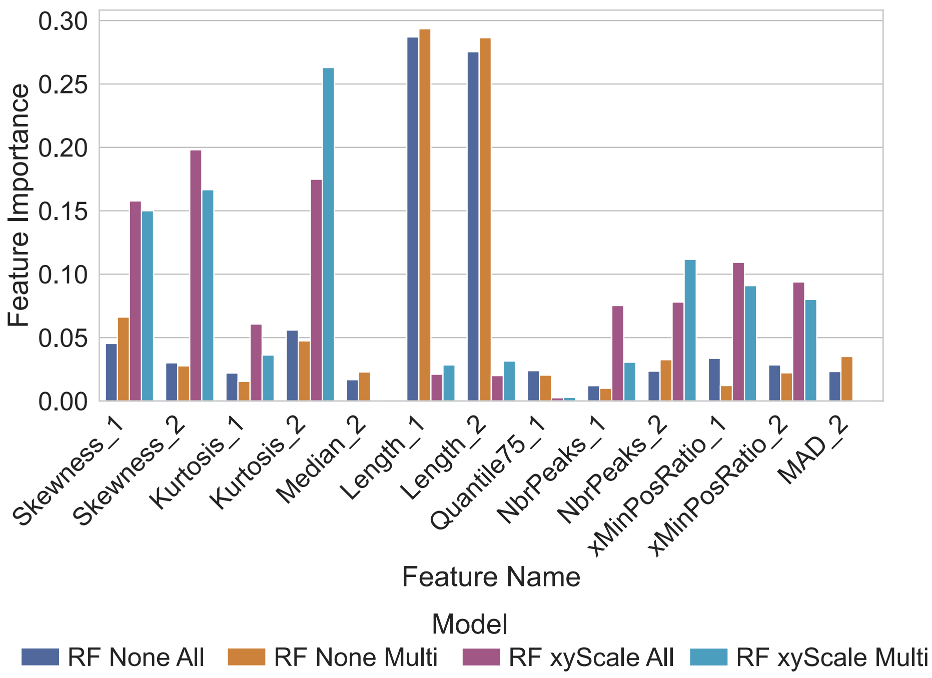

Figure 7. Feature importances of selected features for **Task 1**. We only show features if at least one model rates its importance over 0.02. This is done for the sake of clarity. The suffix number of each feature name represents the index of the time series from which it was extracted.

classify multiple events according to their vehicle count with a BA of 72.2 %. From the feature importance of the RF, we can also deduce that the event duration heavily impacts the results, leading to good results for all events. Data augmentation can be helpful regarding this aspect, as the feature importance of the signal duration reduces when the `xyScale` dataset is used.

Similar observations can be drawn from our regression approach. Using all events and applying no data augmentation, MiniRocket achieves R^2 of 0.80. PCA RF has the best R^2 of 0.66 with only multi-events. Data augmentation worsens the results for all events and only multi-events. The RF's feature importance shows similar behavior to the classification task.

We showed that a data-driven classification and regression approach for vehicle count determination is feasible. These promising results can lead to more sophisticated methods for GBR-based BWIM as current state-of-the-art focuses on single events (Arnold and Keller, 2024a).

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 Sciences*, V-1-2021, 31–38. https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/V-1-2021/31/2021/.

Arnold, M., Keller, S., 2020. Detection and Classification of Bridge Crossing Events With Gound-Based Interferometric Radar Data and Machine Learning Approaches. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020, 109–116. https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/V-1-2020/109/2020/.

Arnold, M., Keller, S., 2024a. Machine Learning and Signal Processing for Bridge Traffic Classification with Radar Displacement Time-Series Data. *Infrastructures*, 9(3), 37. https://www.mdpi.com/2412-3811/9/3/37. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Arnold, M., Keller, S., 2024b. Machine-learning for analyzing bridge displacement using radar data. IABMAS Copenhagen. Submitted and accepted.

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., 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*, ACM, Virtual Event Singapore, 248–257.

Döring, A., Waibel, P., Matthes, J., Scherer, O., Keller, H. B., Keller, S., Müller, J., Schneider, O., 2021. Ratio-based features for data-driven bridge monitoring and damage detection. *Bridge Maintenance, Safety, Management, Life-Cycle Sustainability and Innovations*, CRC Press. Num Pages: 10.

Fachrie, M., others, 2020. A simple vehicle counting system using deep learning with YOLOv3 model. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(3), 462–468.

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. http://www.tandfonline.com/doi/abs/10.1080/15732470903068557.

Gomaa, A., Minematsu, T., Abdelwahab, M. M., Abo-Zahhad, M., Taniguchi, R.-i., 2022. Faster CNN-based vehicle

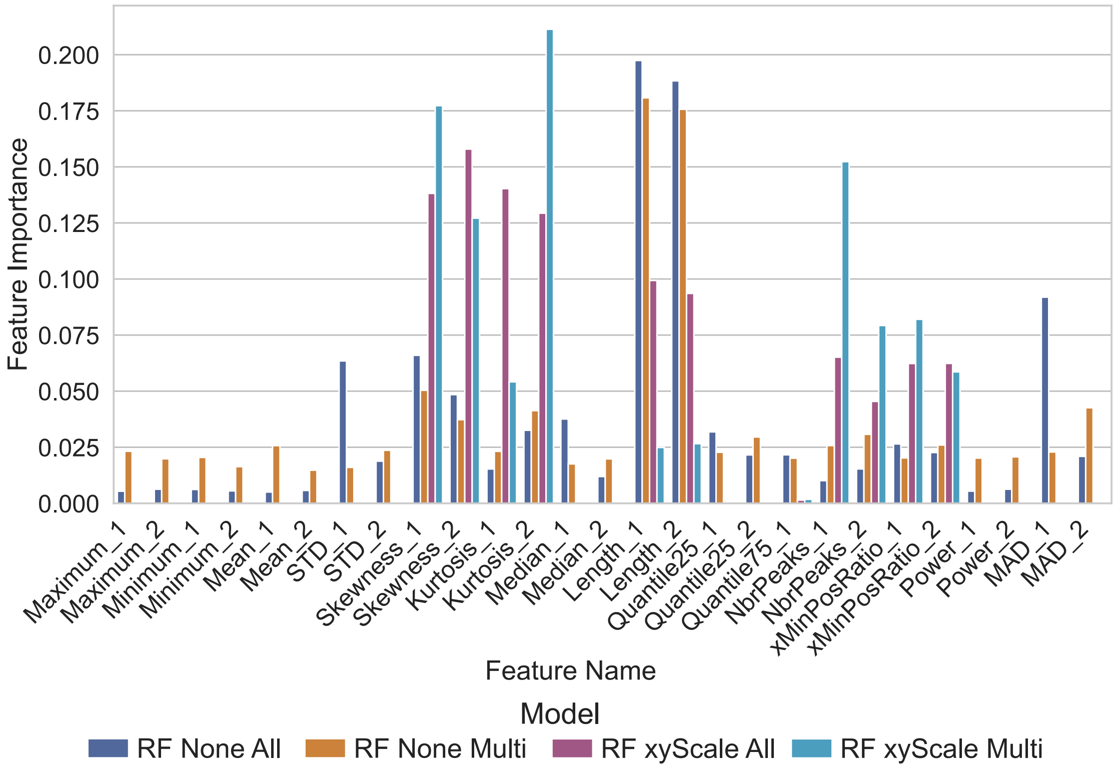

Figure 8. Feature importances of selected features for **Task 2**. We only show features if at least one model gives its importance over 0.02 This is done for the sake of clarity. The suffix number of each feature name represents the index of the time series from which it was extracted.

detection and counting strategy for fixed camera scenes. *Multimedia Tools and Applications*, 81(18), 25443–25471. https://link.springer.com/10.1007/s11042-022-12370-9.

Hertel, L., Phan, H., Mertins, A., 2016. Classifying Variable-Length Audio Files with All-Convolutional Networks and Masked Global Pooling. *CoRR*. http://arxiv.org/abs/1607.02857.

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. https://www.mdpi.com/1424-8220/21/6/2172. 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. *EUSAR 2021; 13th European Conference on Synthetic Aperture Radar*, 1–5.

Michel, C., Keller, S., 2022. Determining and Investigating the Variability of Bridges' Natural Frequencies with Ground-Based Radar. 17.

Michel, C., Keller, S., 2024. Assessing Important Uncertainty Influences of Ground-Based Radar for Bridge Monitoring. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5. https://ieeexplore.ieee.org/document/10360164/.

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. https://ascelibrary.org/doi/10.1061/

Pieraccini, M., Miccinesi, L., Abdorazzagh Nejad, A., Naderi Nejad Fard, A., 2019. Experimental Dynamic Impact Factor Assessment of Railway Bridges through a Radar Interferometer. *Remote Sensing*, 11(19), 2207. https://www.mdpi.com/2072-4292/11/19/2207. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

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. http://link.springer.com/10.1007/s10618-020-00727-3.

Taghvaeeyan, S., Rajamani, R., 2014. Portable Roadside Sensors for Vehicle Counting, Classification, and Speed Measurement. *IEEE Transactions on Intelligent Transportation Systems*, 15(1), 73–83. Conference Name: IEEE Transactions on Intelligent Transportation Systems.

Tan, C. W., Petitjean, F., Keogh, E., Webb, G. I., 2019. Time series classification for varying length series. arXiv:1910.04341 [cs, stat].

Wang, J., Tang, S., 2020. Time series classification based on arima and adaboost. *MATEC Web of Conferences*, 309, 03024. https://www.matec-conferences.org/10.1051/matecconf/202030903024.

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. https://www.mdpi.com/1424-8220/15/4/8444. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.