

Learning road degradation with asphalt-integrated sensor fabric

Ralf Bruns, Jürgen Dunkel, Maximilian Greve, Christina Haxter, Joris Herrmann, Sascha Kayser

Suggested citation:

Bruns, Ralf, Jürgen Dunkel, Maximilian Greve, Christina Haxter, Joris Herrmann, and Sascha Kayser. 2025. "Learning road degradation with asphalt-integrated sensor fabric." *Innovative Infrastructure Solutions* 11. <https://doi.org/10.25968/opus-3798>.

Abstract

Efficient monitoring of the condition of road infrastructure is essential for the provision of a reliable and sustainable mobility and transportation network. The assessment of the structural condition of the asphalt base layer is particularly important in this respect. This paper presents a data-driven approach for an innovative road monitoring system for the non-destructive and continuous determination of the degree of degradation of asphalt roads. The innovation of the project lies in the application of Artificial Intelligence methods to derive the degradation state of the asphalt base layer on the basis of sensor measurements obtained by means of a novel hybrid sensor fabric integrated directly into the asphalt base layer. The proposed Machine Learning-based diagnosis relies heavily on the quality of sensor data. Therefore, we introduce a new method to evaluate the significance of sensor measurements using time series analysis techniques. The feasibility and functionality of the approach is demonstrated through extensive experiments by embedding the sensor material in real asphalt specimens, which are subject to controlled load tests.

Terms of use

CC BY 4.0



Learning road degradation with asphalt-integrated sensor fabric

Ralf Bruns¹ · Jürgen Dunkel¹ · Maximilian Greve¹ · Christina Haxter² · Joris Herrmann³ · Sascha Kayser³

Received: 14 August 2025 / Accepted: 14 November 2025
© The Author(s) 2025

Abstract

Efficient monitoring of the condition of road infrastructure is essential for the provision of a reliable and sustainable mobility and transportation network. The assessment of the structural condition of the asphalt base layer is particularly important in this respect. This paper presents a data-driven approach for an innovative road monitoring system for the non-destructive and continuous determination of the degree of degradation of asphalt roads. The innovation of the project lies in the application of Artificial Intelligence methods to derive the degradation state of the asphalt base layer on the basis of sensor measurements obtained by means of a novel hybrid sensor fabric integrated directly into the asphalt base layer. The proposed Machine Learning-based diagnosis relies heavily on the quality of sensor data. Therefore, we introduce a new method to evaluate the significance of sensor measurements using time series analysis techniques. The feasibility and functionality of the approach is demonstrated through extensive experiments by embedding the sensor material in real asphalt specimens, which are subject to controlled load tests.

Keywords Road condition monitoring · Degradation monitoring · Machine learning · Artificial intelligence · Sensor fabric · Smart infrastructure

✉ Ralf Bruns
ralf.bruns@hs-hannover.de

Jürgen Dunkel
juergen.dunkel@hs-hannover.de

Maximilian Greve
greve.maximilian@gmail.com

Christina Haxter
christina.haxter@wki.fraunhofer.de

Joris Herrmann
joris.herrmann@h2.de

Sascha Kayser
sascha.kayser@h2.de

¹ Department of Computer Science, Hannover University of Applied Sciences and Arts, Ricklinger Stadtweg 120, 30459 Hannover, Germany

² Fraunhofer WKI, Application Center for Wood Fiber Research HOFZET, Heisterbergallee 10A, 30453 Hannover, Germany

³ Department of Civil Engineering, Magdeburg-Stendal University of Applied Sciences, Breitscheidstr. 2, 39114 Magdeburg, Germany

Introduction

Road infrastructure is subject to permanent structural deterioration due to material degradation caused primarily by heavy and intensive traffic, harsh weather conditions, aging and other parameters [1, 2]. Improving the durability and sustainability of infrastructure systems is an active area of research, including studies on novel materials and advanced technologies, see for instance [3–5]. Timely detection of defects and continuous maintenance are of highest importance for the long-term provision of a high-quality road transportation network.

Asphalt roadways normally consist of three asphalt layers—the (i) surface layer, (ii) the binder (or intermediate) layer and (iii) the base layer. As the *asphalt base layer* forms the foundation of the road, the structural condition of the asphalt base layer has a decisive influence on the remaining service life of the roadway. Consequently, the assessment of the structural condition of the asphalt base layer is of particular importance.

At present, *destructive* measurement methods are used, in which the road is additionally damaged by taking drill core samples on site. A comprehensive and continuous assessment of the structural substance of roads is not possible

using conventional inspection methods due to the associated high inspection costs, time-intensive procedures, and the additional damage caused. Furthermore, inspecting and assessing the quality of the *road surface* (cracks, potholes, etc.) is not sufficient to draw reliable conclusions about the extent of structural interventions in the underlying road construction, the associated costs, and the environmental impact of maintenance measures. However, road condition assessment is an essential task for efficient and sustainable maintenance of road networks and road asset management in general [6].

The aim of our approach is the development of an innovative *smart road monitoring system* for the non-destructive, continuous, large area, and cost-efficient determination of the degree of degradation of asphalt roads. Without affecting the ongoing traffic flow. The innovation of the project lies in the application of Artificial Intelligence (AI) methods to derive the degradation state of the asphalt base layer on the basis of sensor measurements obtained by means of a novel hybrid sensor fabric integrated directly into the asphalt base layer.

The main contributions of our smart road infrastructure monitoring system are:

- *Novel fabric-based sensor system*: Data acquisition via a new type of innovative *bio-based fabric sensor system* that is embedded directly into the *asphalt base layer*. The measurement signals recorded with this sensor fabric depict the increasing material degradation over the operational use of the road.
- *Smart machine learning-based degradation assessment*: *AI-based diagnosis* of the degradation state of a road segment on the basis of a Machine Learning (ML) model by evaluating the sensor data available for the first time from the asphalt base layer.
- *Innovative time series-based data quality analysis*: The proposed Machine Learning approach relies heavily on the quality of sensor data. Therefore, we introduce a new method to evaluate the significance of sensor measurements using *time series analysis techniques*.

The large-scale integration of the sensor system into the traffic infrastructure enables a reliable assessment of the current degradation state and the planning of effective and sustainable maintenance measures. Automatic rapid and accurate assessment of the condition of key infrastructure enables data-driven decision making and leads to more resilient infrastructure [7].

The remainder of the paper is organized as follows. In Sect. 2 we introduce the basic idea of our sensor measuring approach. Section 3 describes how road degradation can be learned by machine learning techniques. The machine

learning results are presented in Sect. 4. Section 5 discusses methods in order to assess the significance of sensor measurements. Finally, an overview of related work is given in Sect. 6 and future research directions are discussed in Sect. 7.

Embedded sensor measuring system and data acquisition

Basic concept road condition monitoring

The basic idea of our approach is to develop a *hybrid fabric* that can detect changes in the condition of the surrounding environment with the help of integrated sensor material. For this purpose, a specific sensor wire is woven into the fabric. The wire stretches under load and changes its electrical resistance in the process. This functionalized sensor fabric is then to be incorporated into the asphalt base layer of road pavements in order to detect structural damage in the layer. From the measured changes of the electrical resistance of the sensor fabric, the degradation state of the asphalt layer shall be deduced.

Machine learning methods are applied to learn the relationship between the electrical voltage measured in the tissue and the change of the structural condition of the asphalt. As a result, the trained Machine Learning models can then be used to derive the degradation state of a road on the basis of the data measured in the embedded sensor fabric.

The material selection, fabric design and fabric manufacturing process are presented in the following Sect. 2.2.

In order to demonstrate the feasibility and functionality of our road condition monitoring approach, the sensor fabric was embedded in asphalt test specimens and these test specimens were subjected to controlled stress tests under laboratory conditions. The laboratory experiments with the ongoing load tests enable the time-synchronous measurement of the change in the condition and residual strength of the asphalt specimen on the one hand and the change in the electrical resistance of the sensor material on the other hand. Section 2.3 deals with the embedding of the hybrid fabric in specially adapted test specimens and describes the laboratory test setup with the load series and results of measurements.

A more detailed discussion of the innovative sensor fabric and the laboratory experiments can be found in [8].

Design of sensor fabric

The sensor material must essentially possess two properties in order to be suitable for our application purpose:

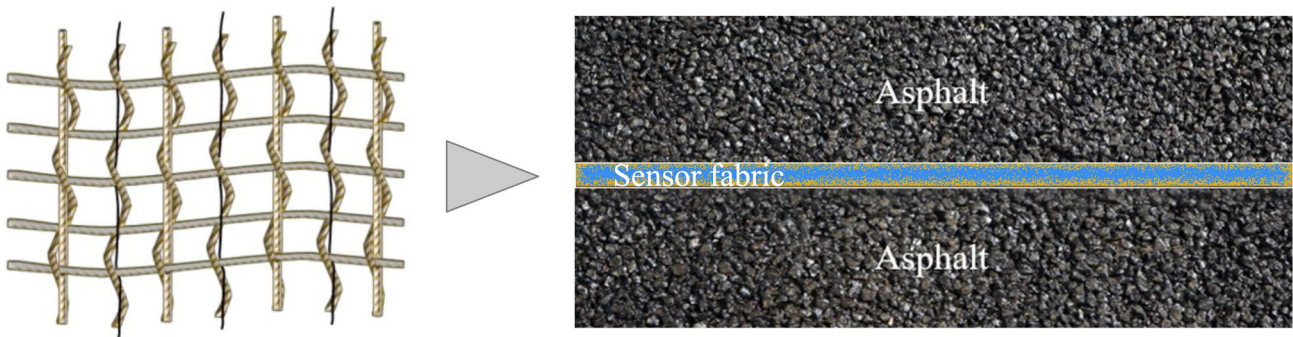
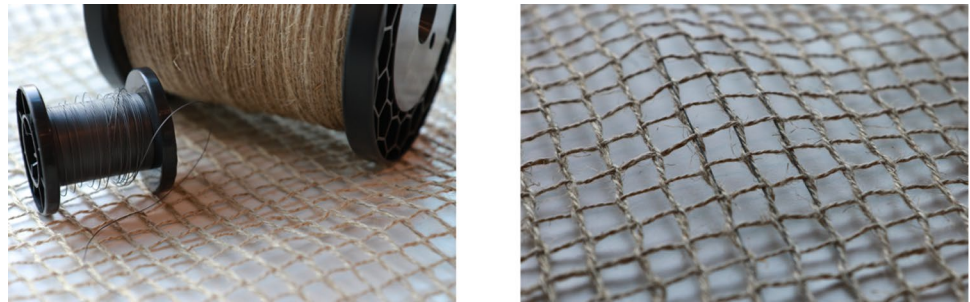


Fig. 1 Copper-nickel wires integrated into bio-based fabric (left), sensor fabric embedded in asphalt layer (right)

Fig. 2 Spools with flax yarn and wire (left), sensor fabric with woven-in wires (right) (© Fraunhofer WKL, Manuela Lingnau)



1. Temperature-independent electrical resistance
2. Integrability into the carrier fabric directly in the weaving process

A *copper-nickel wire* (Cu55Ni44Mn1) was chosen as the sensor material, whose electrical resistance is almost independent of temperature. Moreover, the resistance of the selected wire changes in a nearly linear dependence on its elongation. In addition, the specific electrical resistance is in the right order of magnitude both for measurements on a laboratory scale and for later use on site.

The carrier fabric to be designed for the sensor system must meet the following main requirements:

- High displacement resistance to ensure installation in asphalt test specimens
- Suitable spacing between the threads/wires to avoid structural disturbances
- Minimal interference on the woven wire to minimize possible wire pre-damage during weaving procedure

One type of tissue that fulfills the above requirements is the *leno fabric*. In a leno fabric two threads are joined in warp direction: (i) the standing thread and (ii) the leno thread. The standing thread runs at right angles to the weft thread. The leno thread turns alternately to the right and left around the stationary thread after each weft insertion (see [8]).

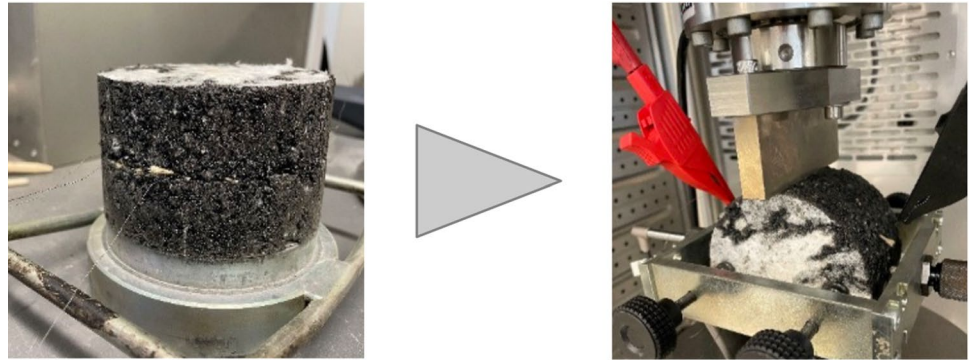
In our project, leno fabrics were designed and produced in which

1. The copper-nickel wires are used as standing threads and
2. Flax yarns (a natural fiber yarn) of different fineness are used for the leno, weft, and standing threads.

Different diameters of the copper-nickel wire with different resulting resistances were tested. Figure 1, on the left, sketches the design of the sensor fabric. Figure 2 shows two spools with flax yarn and with the wire, on the left, and visualizes the functionalized sensor fabric with the woven-in wires on the right. An important aspect in selecting the wire was its suitability for the weaving process. The selected copper-nickel wire meets all weaving requirements. The electrical resistance of the sensor fabric essentially depends on the properties of the embedded wires, in particular the alloy used and the wire thickness. The fabric serves as a carrier material for embedding in the asphalt and should have no influence on the measured voltages. However, it is important that the fabric has high displacement resistance so that neighboring wires do not touch each other and the fabric is sufficiently robust for installation. In addition, the fabric can protect the wire from damage during installation. We have therefore specifically developed a fabric made of flax fibers, which is characterized by high displacement resistance due to its construction and is also bio-based. Potential interferences between adjacent wires are the subject of further investigations.

The novel sensor system is designed as a two-dimensional fabric, which, unlike conventional point or line sensors (like distributed fiber optic sensing or piezoelectric sensors), allows for complete spatial coverage monitoring of infrastructure

Fig. 3 Sensorized asphalt specimen (left) and servo-hydraulic testing machine (right) (according to [8])



cycle	start time [msec]	end time [msec]	start voltage [mV]	end voltage [mV]	mean voltage [mV]
15	422.0	520.0	11.011584	11.012430	11.012128
16	522.0	620.0	11.014674	11.010785	11.012279
17	622.0	720.0	11.013246	11.013968	11.012760
18	722.0	820.0	11.015851	11.013479	11.014802
19	822.0	920.0	11.007137	11.000154	11.002659

Fig. 4 Data from the sensor fabric: measured voltage in successive time intervals, indicated by cycle

such as roads. This significantly increases the probability of detecting spatially distributed damage like fatigue cracks at an early stage. A key advantage is its cost-effectiveness: the system measures the change in electrical resistance of metal alloy wires, enabling the use of extremely low-cost electronics. This makes large-scale, continuous monitoring economically viable for the first time. Furthermore, the installation is seamlessly integrated into the road construction process, as the flexible fabric is simply unrolled between asphalt layers, ensuring excellent mechanical coupling and accurate measurements.

Fatigue tests in laboratory setting

The hybrid functionalized fabric, introduced above in Sect. 2.2, was embedded into asphalt test specimens (see Fig. 1, on the right).

Fatigue tests of the (sensorized) specimens were conducted in a servo-hydraulic testing machine under laboratory conditions. The road load was simulated by a drop hammer with cylindrical guide rod. Displacement transducers measured the horizontal transverse strains. The degree of degradation of the asphalt specimen can be calculated from these measurements.

Figure 3 shows the experimental setup with the sensorized asphalt specimen on the left and the servo-hydraulic testing machine with the connected measuring system on the right.

The load tests cause structural damage in the asphalt and, thus, also lead to an elongation of the sensor wires, which results in a change in the electrical resistance. During the mechanical loading of the sensorized asphalt test specimens, the electrical resistance values are continuously measured and

recorded and the corresponding degree of structural damage is calculated synchronously. These test series make it possible to correlate and interpret the resistance and degradation data obtained.

For more details about the laboratory setting and the fatigue functions please refer to [8].

Data acquisition by laboratory test series

A series of fatigue tests were performed in the laboratory environment introduced above. Sensor fabrics with interwoven wires of four different thicknesses were evaluated.

Samples of the data measured in the laboratory tests are shown in the following two figures. Figure 4 shows the recorded voltage measurements from the sensor fabric. The first column 'cycle' indicates the number of time steps in the experiment, in particular how many times the drop hammer has fallen to the specimen. The following two columns refer to the time in milliseconds of the corresponding voltage measurement interval. The time values are related to the time that has elapsed since the start of the experiment. The other columns contain the voltage values at the beginning and end of the measurement interval, as well as the medium value (all in *mV*).

Figure 5 shows the results from the fatigue tests carried out with the hydraulic testing machine. Based on the measured transverse strains, the fatigue of the asphalt specimen was calculated. The columns in Fig. 5 indicate (i) the time step of the measurement, using the 'cycle' as shown in Fig. 4, (ii) the measured strain and (iii) the calculated fatigue, both in percentage.

cycle	strain [%]	fatigue [%]
15	0,0653420959393891	0,0471505359444252
16	0,0674499054858214	0,0502939050073869
17	0,0674499054858216	0,0534372740703486
18	0,0674499054858214	0,0565806431333103
19	0,0674499054858213	0,059724012196272

Fig. 5 Degradation states for each cycle: calculated from strain measurement

The degree of fatigue damage is analyzed in the laboratory using the *Dynamic Indirect Tensile Fatigue Test*.¹ In this test, a cylindrical asphalt specimen is subjected to repeated vertical compression. This compression indirectly generates horizontal tensile strain, which simulates the real-world stress at the bottom of an asphalt layer under traffic. By recording how many load cycles are required to cause failure at a specific strain level, the material’s characteristic fatigue function is determined. This function describes the fundamental resistance of the asphalt to the type of cracking that the sensor system is designed to monitor in the field. The procedure for calculating the fatigue state is as follows:

1. *Initial strain*: First, the initial elastic horizontal strain ($\epsilon_{el,anf}$) is calculated for each test.
2. *Number of load cycles to failure (N_{Macro})* is defined as the number of load cycles² at which macrocracking occurs in the specimen. It is determined using the concept of the *energy ratio ER(N)* (using the stiffness modulus at load cycle N). N_{Macro} corresponds to the load change N at which the *ER(N)* curve reaches its maximum.
3. *Fatigue function*: The relationship between the initial elastic horizontal strain ($\epsilon_{el,anf}$) and the corresponding number of load cycles up to the macrocrack (N_{Macro}) is described by a power function: $N_{Macro} = k \times \epsilon_{el,anf}^n$. The parameters k and n are material-specific values that are determined by regression from the test results.

For further details regarding the fatigue calculation formulas see [8].

Around 4000 data points are received for each experiment carried out with a specific wire.

Machine learning of road degradation

The main goal of our approach is to learn the road degradation state from voltage values measured in the sensor fabric. More precisely, we want to learn the hypothesis function $h()$:

$$h(U) = D \tag{1}$$

with voltage U measured in the sensor fabric and degradation state D of the road segment. In our Machine Learning approach U serves the feature and D can be considered as the target of the hypothesis function.

Data preprocessing

To apply Machine Learning methods, feature engineering must first be performed. Specifically, we followed these steps:

- First, the voltage data from the table in Fig. 4, representing the feature values, must be integrated with the fatigue data from Fig. 5, which serves as the target variable. Fortunately, the data can be easily linked through the common ‘cycle’ column, which is present in both tables, see Fig 6.
- Absolute values of the measured voltages are not useful, as the voltage values measured at the start of the tests may deviate depending on the wire thickness and small deviations in the wire length. To ensure comparability across different experiments, we used the percentage increase in voltage relative to the starting point of the measurement, rather than the absolute values.³
- To facilitate the interpretation of the test results, we have mapped the percentages of the calculated fatigue to classes, i.e. the intervals of fatigue. The fatigue class 0 corresponds to the initial phase of the road’s life cycle, characterized by fatigue values below 10%. In contrast, fatigue class 9 represents the final phase of the road’s life cycle, with fatigue levels exceeding 90%. The use of fatigue classes enables the application of standard evaluation metrics, such as precision and recall. For more details, refer to Sect. 4.
- It should be noted that the data processing procedure inherently includes a basic form of outlier elimination. First, the voltage increase is calculated based on the mean of all voltage values measured during a load cycle, which reduces the influence of single outliers within individual cycles. Since occasional strong downward outliers were observed in some experiments (see Fig. 7), all data points with voltage values below the starting point are automatically removed from the dataset. Nonetheless, more advanced outlier detection and elimination mechanisms should be implemented in the future.

¹ In accordance with technical specification TP Asphalt-StB, Part 24.

² i.e. hammer drops.

³ The percentage increase in voltage at a given time instant t is calculated relative to the initial (start) voltage and is defined as Percentage increase in voltage(t) = $\frac{\bar{U}(t) - U_{start}}{U_{start}} \times 100$, where $\bar{U}(t)$ denotes the mean voltage at time t and U_{start} is the voltage at the starting point.

Fig. 6 Voltage values mapped to road degradation classes

mean voltage [mV]	voltage increase	fatigue[%]	fatigue class
5.083777	0.030254	66.597736	5
5.083193	0.018768	66.601364	5
5.083388	0.022599	66.604992	5
5.084436	0.043225	66.608620	5
5.085062	0.055547	66.978668	6
5.083715	0.029034	66.982296	6
5.082771	0.010468	66.985924	6
5.082515	0.005429	66.989552	6
5.083119	0.017318	66.993180	6
5.083023	0.015418	66.996807	6

In particular, sudden and pronounced drops in voltage values should be identified and corrected to improve data reliability.

These data preprocessing steps yield to the integrated Machine Learning dataset shown in Fig. 6.

The two newly calculated columns, ‘voltage increase’ and ‘fatigue class,’ serve distinct roles: ‘voltage increase’ acts as the feature for the hypothesis function to learn, while ‘fatigue class’ is the corresponding target. The feature data may appear somewhat limited, but it is important to note that the sensor system only provides voltage measurements that can be used to predict road conditions.

Machine learning pipeline

The preprocessing of the data described above leads to a *multi-class classification problem* according to the following Eq. 2.

$$h(\Delta U) = C \tag{2}$$

The hypothesis function $h()$, which must be learned, maps the voltage increase ΔU to a set of N degradation states $C = \{C_0, C_1, \dots, C_{N-1}\}$, where each C_i represents a distinct degradation category. More precisely, the degradation states refer to a sequence of degradations starting with a new road in state C_0 until it has to be renewed in state C_{N-1} .

According to the classical Machine Learning process, we proceed the following steps for finding an appropriate hypothesis function:

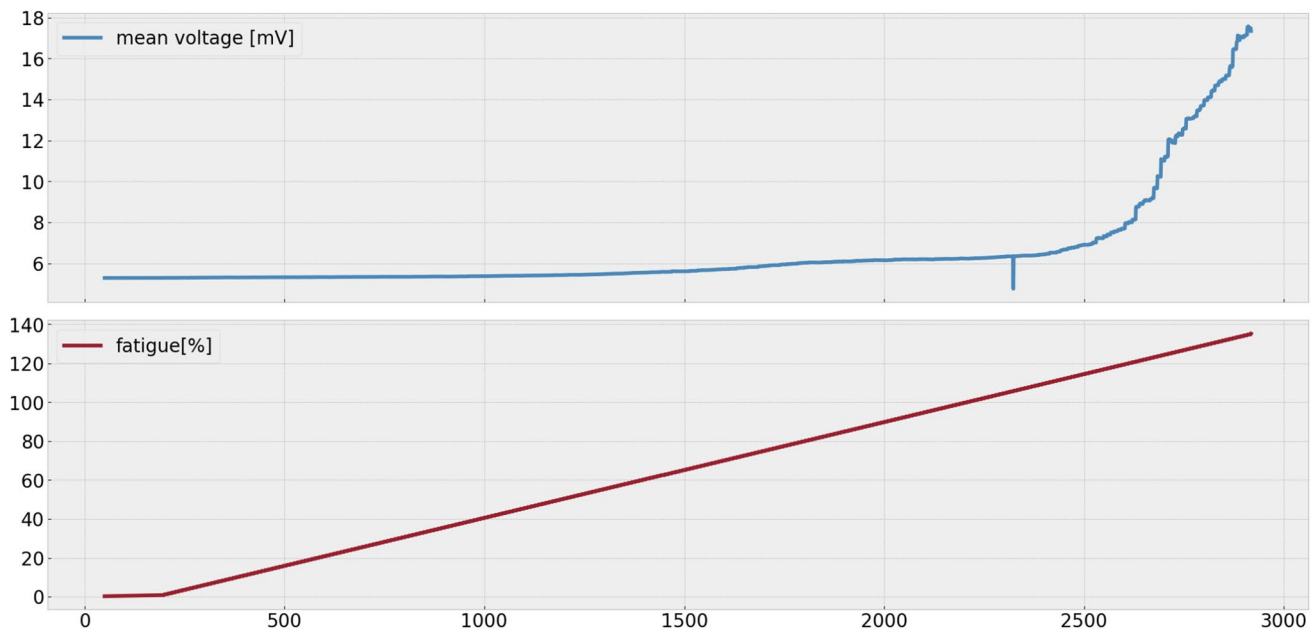


Fig. 7 Voltage measured in the fabric (top) and fatigue state (bottom) in a very good experiment: the linear increase of the fatigue corresponds with an exponential voltage increase

1. *Splitting data into training and test set:* The dataset is randomly split into training set and test set using standard *k-fold cross-validation*, with 20% of the data reserved for testing and evaluating the prediction model. In the experiments presented below, we consistently used $k = 5$ folds throughout.
2. *Model selection and training:* The next step is to select an appropriate classification model for training on the dataset. Among the models considered were the *Logistic Regression* and the *Random Forest Classifier*. A series of evaluation experiments was conducted for each model, using a range of datasets to ensure a comprehensive assessment of the performance. In particular, we performed a grid search to optimize each model’s hyperparameters, systematically testing various parameter combinations.
3. *Model evaluation:* Given our multi-class classification problem, we employed the following classical evaluation metrics to quantify the performance of the learned hypothesis function.
 - *Recall* measures how well the model identifies the relevant instances for each class. Specifically, recall is the proportion of correctly predicted instances of a class compared to the actual instances of that class in the dataset:

$$recall(C_i) = TP(C_i)/(TP(C_i) + FN(C_i)), \tag{3}$$

where True Positives $TP(C_i)$ is the number of instances correctly predicted as belonging to class C_i and False Negatives $FN(C_i)$ is the number of instances belonging to C_i but are incorrectly predicted as another class. A recall of 1.0 means that all instances of a class are predicted correctly.

- *Precision* quantifies how many of the instances predicted as belonging to a certain class are actually correct. It is calculated by:

$$precision(C_i) = TP(C_i)/(TP(C_i) + FP(C_i)), \tag{4}$$

where False Positives $FP(C_i)$ is the number of instances incorrectly predicted as belonging to class C_i . A precision of 1.0 means that every instance the model predicted as positive for a class is correct, there are no false positives.

- *Precision**: Because in our road monitoring system, the sequence of fatigue classes refers to a consecutive sequence of deterioration states, we can distinguish how ‘good’ wrong predictions are. If our

model predicts an adjacent fatigue class, the predicted state is relatively close to the correct road state and could be accepted as a valuable result. According to road maintenance domain experts, precise classification of the fatigue state is not critical during operational monitoring. When using a 10-level fatigue scale, a deviation of one class (i.e., an error of approximately 10%) in either direction is considered acceptable. In most cases, asphalt pavement renewal is a long-term, planned activity. As such, immediate intervention is rarely required, and it is generally sufficient to accurately capture the overall trend of deterioration rather than determine the exact fatigue state.

To reflect this domain-specific tolerance, we define the modified *precision**(C_i) as follows:

$$precision^*(C_i) = TP^*(C_i)/(TP^*(C_i) + FP^*(C_i)) \tag{5}$$

where $TP^*(C_i)$ is the number of predictions of the correct class C_i and its adjacent classes C_{i-1} and C_{i+1} with $i \in \{0, 1, \dots, N - 1\}$, as well. Accordingly, $FP^*(C_i)$ is the number of instances incorrectly predicted as belonging to class C_i , when they are not belonging to class C_i or its adjacent classes.

- The *F1 score* provides a more balanced evaluation of a model’s performance, especially in cases where there is an uneven class distribution or when both precision and recall are important. It is the harmonic mean of precision and recall and can be calculated as:

$$F1(C_i) = 2 \times \frac{Precision(C_i) \times Recall(C_i)}{Precision(C_i) + Recall(C_i)}. \tag{6}$$

In road monitoring, precision is of the utmost importance: the predictions should be as accurate as possible so that no incorrectly detected fatigue condition results in costly repairs. We use *precision* (and *precision**) to denote the average precision over all classes.

Machine learning evaluation results

To test the suitability of the intelligent sensor fabric, different experiments were carried out with copper-nickel wires of 4 different diameters: 0.2 mm, 0.3 mm, 0.4 mm, 0.6 mm. For each wire thickness, 7-9 asphalt specimens were produced and a total of 32 fatigue tests with specimens were conducted. Each experiment produced a dataset, which

served to learn the relationship between the measured voltage and the associated fatigue class as formally described in Eq. 2.

Model learning was performed with scikit-learn,⁴ the well-known open source Python Machine Learning library, which provides various classification algorithms.

Model selection

To select the most suitable model we conducted experiments using several Machine Learning classification methods: *Logistic Regression*, *Support Vector Classifier*, *Decision Trees*, *Random Forest*, and *Gradient Boosting*. Following the Machine Learning pipeline described in Sect. 3.2, we carried out the following steps for each of our 32 experiments:

- The data was split into a 80% training set and a 20% test set.
- On the training set, we performed 5-fold cross-validation to obtain average performance metrics across folds. Each fold served as a validation set once, while the remaining four folds were used for training. Additionally, we employed grid search to systematically explore hyperparameter combinations and identify the optimal settings for each model.
- The test set was completely held out during training and cross-validation. It was used exclusively for the final, unbiased evaluation of the model after hyperparameter tuning. The final model, trained on the full training set with the best-found hyperparameters, was evaluated on this test set.

Table 1 lists the final results for the evaluated models using their optimized hyperparameters. The columns show the mean and the median precision values achieved across all experiments.

Among the evaluated models, the *Random Forest Classifier* achieved the best overall performance. In addition, most of the other models, particularly the *Support Vector*

Table 1 Comparison of model classifiers based on precision scores across all experiments

Classifier	Mean precision	Median precision
Random forest	0.47	0.41
Support vector	0.44	0.41
Decision tree	0.45	0.35
Gradient boosting	0.42	0.30
Logistic regression	0.34	0.20

⁴ <https://scikit-learn.org>.

Our models are also using further libraries such as pandas and NumPy for data wrangling.

Classifier performed only slightly worse, indicating that classification quality was generally comparable across most models.

However, it can be observed, that a simple model such as *Logistic Regression* provides significantly lower classification performance due to its inability to model non-linear relationships in the data. Considering that our dataset is likely to contain complex, non-linear interactions and decision boundaries, this linearity constraint limits the model's ability to capture relevant patterns. From an ablation perspective, this result indicates that overly simplistic models, such as Logistic Regression, are inadequate for capturing the complexity inherent in our data.

Based on these findings, all results presented in the following sections were obtained using the *Random Forest Classifier*.⁵

Evaluation results

We start with an experiment that works almost perfect to predict the road fatigue. Figure 7 shows the progression of data points over about 3000 load cycles, measured during a specific experiment using a wire with a 0.3 mm diameter.

As expected, the figure initially shows that fatigue increases linearly.⁶ Only at the very beginning of the stress test can no noticeable effect of the drop hammer be seen. The voltage curve looks a bit different: at first, it also increases linearly, but in the end an exponential increase can be observed. Furthermore, there is a downward outlier at around 2400 load cycles. In this particular experiment, it should be possible to derive the relation between voltage and fatigue without significant issues.

This impression is confirmed by the results of a Random Forest classification, which are shown in Fig. 8. For each fatigue class, the achieved results for precision, recall, and F1 score are given. Furthermore, the support column indicates how many data points in the test set belong to this class.

The corresponding confusion matrix shows that only in a few cases a neighboring fatigue class was predicted instead of the correct one. Each row represents, for a given fatigue class, the number of correct and incorrect classifications. In the first row, we see that all 79 data points for class 0 were correctly classified. For class 1, out of 63 predictions, 55 were correctly classified, 1 were misclassified as class 0, and 7 were misclassified as class 2.

⁵ The following hyperparameters are used: 'estimators = 5', 'bootstrap = true', 'min samples split = 4'.

⁶ It should be noted that the calculation method can also result in fatigue values of more than 100%. In such a case, the road section can still be used.

Fig. 8 Evaluation results of Random Forest classification for a very good experiment: very good precision values across all classes, and a perfect precision* value. There is only a very small number of misclassifications to adjacent fatigue classes, indicated in the confusion matrix by the non-diagonal elements

	precision	recall	f1-score	support	<u>confusion matrix:</u>
0	0.99	1.00	0.99	79	[[79 0 0 0 0 0 0 0 0 0]
1	0.93	0.87	0.90	63	[1 55 7 0 0 0 0 0 0 0]
2	0.89	0.93	0.91	60	[0 4 56 0 0 0 0 0 0 0]
3	1.00	1.00	1.00	62	[0 0 0 62 0 0 0 0 0 0]
4	1.00	1.00	1.00	40	[0 0 0 0 40 0 0 0 0 0]
5	1.00	1.00	1.00	53	[0 0 0 0 0 53 0 0 0 0]
6	0.98	0.94	0.96	51	[0 0 0 0 0 0 48 3 0 0]
7	0.94	0.98	0.96	47	[0 0 0 0 0 0 1 46 0 0]
8	1.00	1.00	1.00	51	[0 0 0 0 0 0 0 0 51 0]
9	1.00	1.00	1.00	68	[0 0 0 0 0 0 0 0 0 68]

precision: 0.97 precision*: 1.0

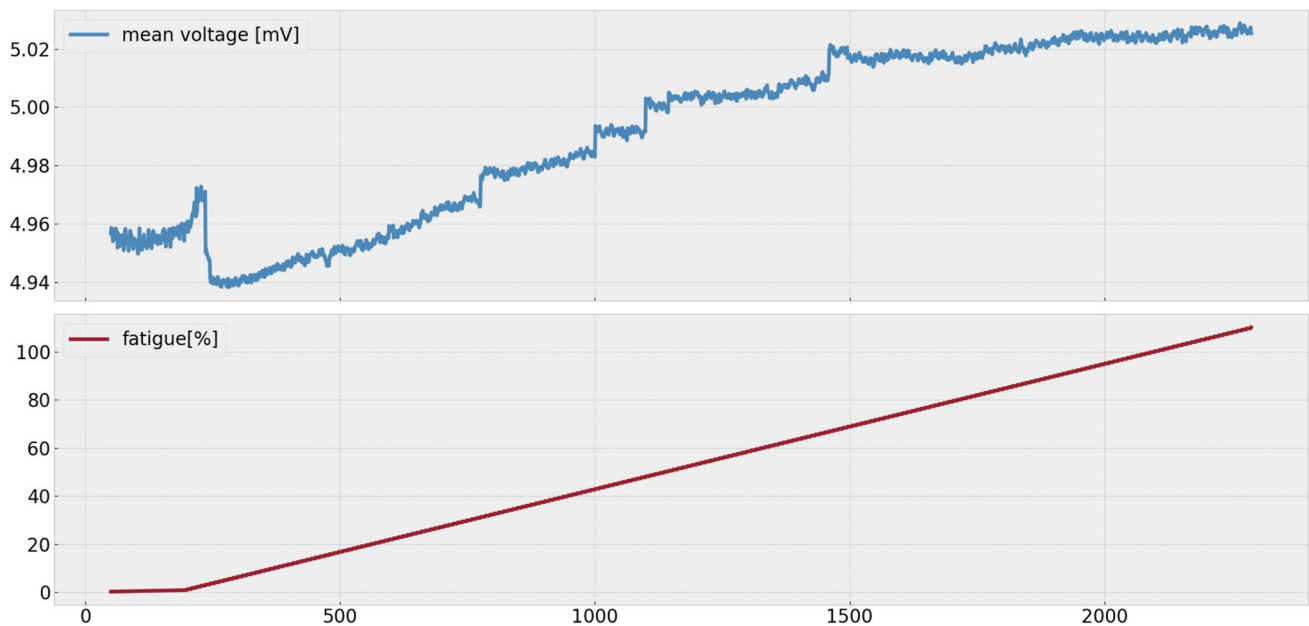


Fig. 9 Voltage measured in the fabric (top) and fatigue state (bottom) in a good experiment: the linear increase of the fatigue corresponds with a continuous voltage increase, albeit with a drop at the beginning

Considering a test set of 574 classifications from 10 classes, almost perfect precision and recall values of 0.97 each were achieved. Because each misclassification predicts an adjacent class, we achieve a perfect modified $precision^* = 1.0$.

The experiment demonstrates that the developed sensor fabric is an effective means of determining the deterioration condition of a road section, with a high degree of accuracy.

But the quality of the data provided by the sensor fabric depends heavily on the installation circumstances. Loose wires that are not tightly embedded in the asphalt base layer or a small stone in the asphalt that exerts pressure on the wire cause poorer data quality. An example gives the results shown in Fig. 9. The voltage rise curve looks rather similar as that before, but at the beginning it is slightly decreasing. In addition, the course of the curve has a certain amount of noise.

Obviously, the classification results are correspondingly worse. As shown in Fig. 10, only precision and recall values of about 0.66 can be achieved.

This means that only 66% of the predicted classes are correct. However, the learned model still provides valuable results, as the confusion matrix reveals that it generally predicts adjacent fatigue classes. This indicates that the predictions are relatively close to the correct state. This is proven by a modified $precision^* = 0.97$.

Unfortunately, the sensor fabric may sometimes fail to provide any meaningful insight into the fatigue condition of a road section. Figures 11, 12 illustrate an experiment where the voltage oscillates, making it impossible to infer the fatigue state from the voltage values.

As a result, the classifier can only achieve a low precision about 0.36. A look at the confusion matrix also shows that it is practically impossible to predict road conditions on the

Fig. 10 Evaluation results of Random Forest classification for a good experiment: reasonable precision values across all classes, and a very good precision* value. Only a moderate number of misclassifications indicated in the confusion matrix by the non-diagonal elements

	precision	recall	f1-score	support	<u>confusion matrix:</u>
0	0.71	0.72	0.72	72	[[52 10 10 0 0 0 0 0 0 0]
1	0.76	0.68	0.72	47	[13 32 2 0 0 0 0 0 0 0]
2	0.72	0.76	0.74	50	[8 0 38 4 0 0 0 0 0 0]
3	0.72	0.77	0.74	30	[0 0 3 23 4 0 0 0 0 0]
4	0.75	0.60	0.67	55	[0 0 0 5 33 17 0 0 0 0]
5	0.61	0.79	0.69	34	[0 0 0 0 7 27 0 0 0 0]
6	0.44	0.56	0.49	36	[0 0 0 0 0 0 20 16 0 0]
7	0.47	0.33	0.39	52	[0 0 0 0 0 0 25 17 10 0]
8	0.57	0.78	0.66	36	[0 0 0 0 0 0 0 2 28 6]
9	0.80	0.67	0.73	36	[0 0 0 0 0 0 0 1 11 24]]

precision: 0.66 precision*: 0.97

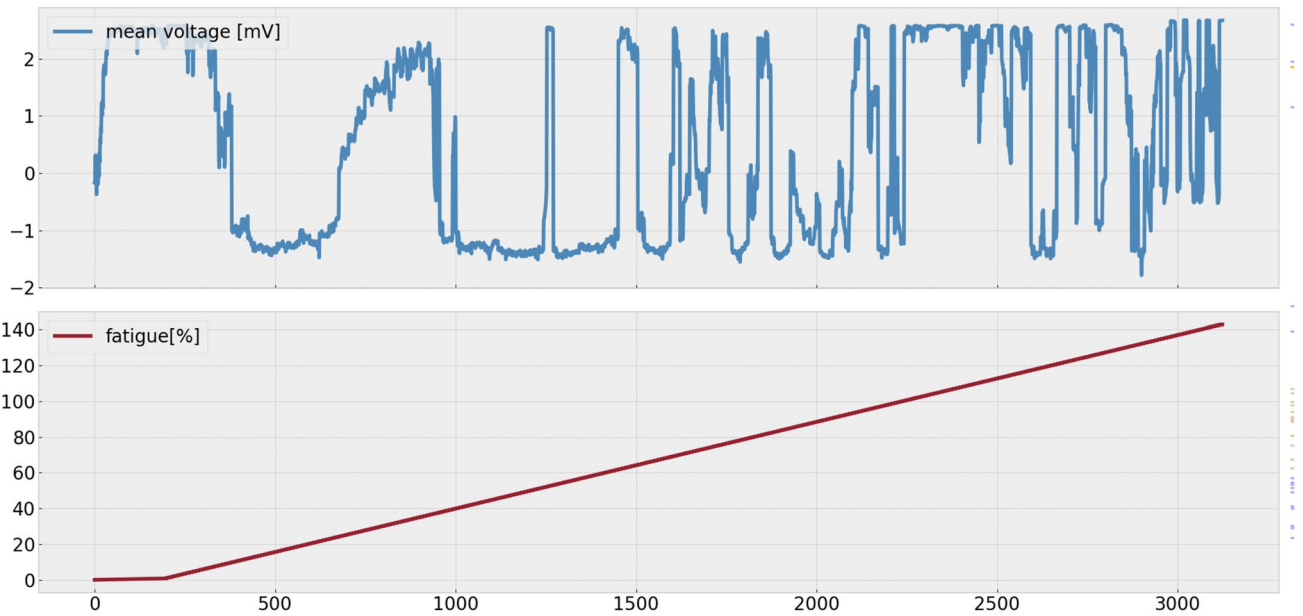


Fig. 11 Voltage measured in the fabric (top) and fatigue state (bottom) in a bad scenario: the linear increase of the fatigue corresponds with oscillating voltage curve

Fig. 12 Evaluation results of Random Forest classification for a bad experiment: bad precision values across all classes, as well as a poor precision* value. There are a large number of misclassifications, as indicated by a dense confusion matrix with few zero entries

	precision	recall	f1-score	support	<u>confusion matrix:</u>
0	0.44	0.54	0.48	68	[[37 0 11 3 3 1 2 0 1 10]
1	0.36	0.40	0.38	53	[0 21 1 5 11 8 2 4 1 0]
2	0.45	0.63	0.53	46	[4 1 29 4 0 2 3 3 0 0]
3	0.19	0.17	0.18	46	[3 11 3 8 5 7 4 4 1 0]
4	0.33	0.51	0.40	43	[2 5 0 3 22 2 2 5 2 0]
5	0.23	0.13	0.17	52	[11 9 2 11 6 7 2 3 1 0]
6	0.27	0.20	0.23	50	[10 7 8 3 7 1 10 2 2 0]
7	0.24	0.22	0.23	37	[2 2 3 4 5 1 7 8 4 1]
8	0.39	0.23	0.29	39	[4 3 5 0 6 2 4 5 9 1]
9	0.68	0.55	0.61	47	[12 0 2 2 2 0 1 0 2 26]]

precision: 0.36 precision*: 0.48

data from this sensor fabric. The predictions are distributed across almost all classes, which leads to a very low modified $precision^* = 0.48$.

Findings

Measurement quality classification

In summary, we can say that some of the measurements are immediately almost perfect and well-suited for learning. Others are unusable for our purposes, with no recognizable or plausible course of measurements. A third category includes measurements that, while generally usable and showing expected behavior, have data quality issues that may impair learning outcomes. These issues include noise, outliers or anomalies, and unusable sections at the beginning or end of the measurement series.

Feature engineering

An approach to further improve the model quality is to expand the number of features.

However, the servo-hydraulic testing machine used in this study provides controlled laboratory conditions that only allow the measurement of voltage values. Additional environmental data cannot be included.⁷

Under real-world operation, additional features must be considered to account for environmental influences, including those that measure weather parameters such as temperature and humidity, as well as cameras and pressure sensors to estimate traffic load. In particular, we expect temperature to influence both the mechanical behavior of the asphalt and the response of the sensor system. At this stage, however, we lack sufficient insights to assess the importance of these additional features.

Another strategy to improve the model is to include additional voltage measurements as features. We investigated how the feature space can be expanded by including previous voltage measurements (lag values) and the differences between data points in order to create a temporal context and capture short-term variations. However, this approach did not significantly improve the results in the “bad data” experiments and led to strong overfitting—yielding excellent training accuracy but poor cross-validation performance. Therefore, this approach was not pursued further in this study.

⁷ Although the temperature can be varied, it had no relevant impact on our laboratory experiments. The selected copper–nickel wire exhibits its temperature-independent electrical resistance, and the duration of the laboratory tests—in the order of minutes—is likely too short for minor temperature variations to influence the fatigue state of the asphalt specimens.

Furthermore, it could be exploited that several wires are contained in the smart sensor fabric. This design leads to data redundancy, as multiple wires measure the same section of road, which helps to detect and correct erroneous data, resulting in a better model. However, the servo-hydraulic testing machine used in our experiments does not support the use of multiple wires simultaneously. The implementation of multiple wires will become feasible after further advancements in the testing machine and during the subsequent deployment of the sensor system in real-world road applications.

Fortunately, in our laboratory experiments, neither imbalanced data nor noisy labels were observed. As shown in Figs. 7, 9 and 11, the measured fatigue calculated from strain measurements increases linear. The calculated fatigue labels are therefore considered reliable, as a linear increase corresponds to the expected linear deterioration under constant test conditions.

Significance of sensor data

The results presented in Sect. 4 demonstrate that the effectiveness of our approach is highly dependent on data quality. The sensor fabric may be improperly installed in the asphalt or degrade over time, leading to unreliable data. Due to these challenges in installing the fabric in the test specimen, a high percentage of experiments did not yield useful results.⁸

Therefore, distinguishing reasonable sensor data from untrustworthy data is a key issue of our road monitoring approach:

1. Only significant data should be used to train the classification model, ensuring accurate learning of the relationship between measured voltages and the corresponding road degradation states. If the training data is not correct, the quality of the classifications will suffer.
2. In real-life operation, only a significant dataset from properly functioning wires should be used to classify road conditions, enabling accurate diagnostic results. Improperly functioning sensor fabrics are useless for road section classification purposes.

⁸ Some damage to the sensor wires is to be expected and is unavoidable. However, for the presented approach, it is sufficient if a certain percentage of properly functioning sensors provide reasonable data quality. Note that this percentage is expected to improve significantly as the sensor wire material and diameter, wire coating, fabric design and material, hybrid fabric manufacturing process, and the installation procedure and guidelines in asphalt are further optimized.

In the following, we present a *time series-based approach* for automatically assessing the significance of the sensor data generated by a specific wire within the fabric.

Time series of road sensor data

The measurement data from a particular wire can be viewed as a *time series*, consisting of a sequence of temporary ordered observations [9]. As in most time series, values that are closer together tend to have a high correlation [10, 11]. This characteristic can be leveraged to identify outliers and anomalies in sensor data. In general, time series can exhibit a variety of patterns [12]:

- A *trend* in a time series refers to the long-term movement or direction in the data over time. In a road monitoring system, the road condition is continuously deteriorating. This behavior should result in increasing voltage values over time, indicating a *positive trend*.
- *Seasonality* refers to recurring patterns or cycles with a fixed period (e.g., daily, weekly, monthly) due to seasonal factors. The sensor fabric data may vary due to seasonal phenomena, such as temperature changes. However, the provided datasets in Sect. 4 are derived from laboratory experiments conducted under constant temperatures. In the future, in real-life operation the temperature will influence the measurement values.

STL decomposition

STL decomposition (Seasonal-Trend decomposition using Loess) is a versatile and robust method for breaking down a time series into the trend and the seasonal component. When the trend and seasonal components are removed from the data, the remaining residual component captures everything not influenced by trends, cycles, or seasonal effects [12–14].

STL uses *Loess* (locally estimated scatterplot smoothing) as a regression method to flexibly estimate the trend and seasonal components. It can operate with an additive model, meaning that at each time instant t , the data point y_t can be calculated as the sum of the corresponding trend (T), seasonal (S), and residual (R) components as expressed by the following equation.

$$y_t = T_t + S_t + R_t \tag{7}$$

Although we do not expect true seasonal effects in the voltage data measured by the sensor fabric, STL decomposition remains a valuable tool for de-noising the signal and separating the long-term trend from short-term fluctuations—even if those fluctuations are not strictly periodic. To

achieve this, we selected a large seasonal window, defined as one-tenth of the total number of data points. In this context, the seasonal window functions more as a smoothing parameter rather than representing a physical cycle length. This allows the STL algorithm to treat short-term variations as “seasonal” noise and effectively filter them out, thereby isolating the underlying trend.

Figure 13 shows an example of STL decomposition applied to a dataset generated by a specific wire [9]. It depicts the original data of the voltage increase (at the top) along with the three corresponding components T , S , and R .

The trend component contains the largest share of the data in the STL decomposition, and increases steadily, resembling the linear progression of voltage curve. As expected, the seasonal component, with values ranging from -0.0011 to 0.0011 , contributes only a very small portion. After subtracting the trend and seasonal components, a residual remains as shown at the bottom. At the end of the curve of the original data, increasingly larger leaps of the voltage increments results in a higher variance of the residual component.

Trend strength

The *strength of the trend* in a time series is a measure of the extent to which the trend component contributes to the overall variability of the data. A strong trend means that a significant portion of the data’s variability can be explained to the trend component, whereas a weak trend suggests that other factors, such as seasonal patterns or noise, are more dominant [15]. The strength of the trend F_T can be calculated by Eq. 8:

$$F_T = \max(0, 1 - \frac{\text{variance}(R_t)}{\text{variance}(T_t + R_t)}) \tag{8}$$

A value close to 1.0 indicates a strong trend, while a value near 0 indicates a weak or non-existent trend. The dataset depicted in Fig. 13 exhibits a trend strength of almost 1.0, indicating that the data is entirely explained by the trend component, with no residual variability [9]. This suggests a clear progression over time and may imply weak or negligible effects from other factors such as noise or seasonality.

Autocorrelation function

The *autocorrelation function (ACF)* measures the similarity between a time series and a lagged version of itself shifted by k values (*lags*) over successive time intervals. It quantifies how the values of a time series at one time point are related to values at previous time points [16]. The ACF $\rho(k)$ for lag k can be calculated by Eq. 9.

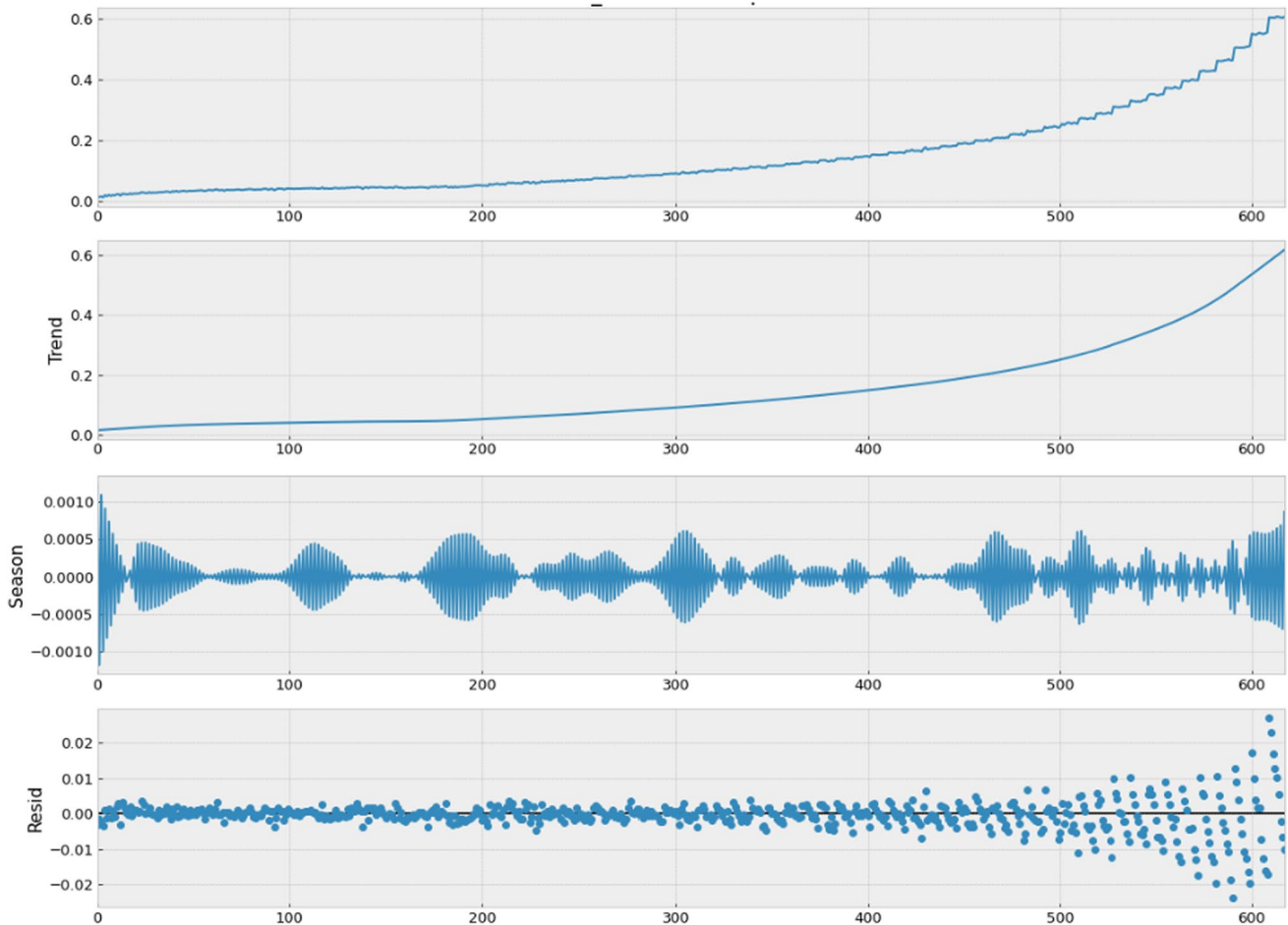


Fig. 13 STL decomposition of sensor data from a very good experiment (curve at the top) into trend, seasonal and residual components [9]. The trend curve is smoother as that from the original sensor data.

Seasonal and residual data are orders of magnitudes smaller and account only for a small proportion

$$\rho(k) = \frac{\sum_{t=k+1}^N (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2} \tag{9}$$

with the total number of observations N , and the mean of the time series \bar{y} .

A high ACF value at lag k indicates that the data points y_t and y_{t-k} are strongly correlated. Obviously, the ACF is 1.0 for $k = 0$ and is getting smaller with greater lags. This is due to the fact that values closer in time are more similar than values further away in time [12].

The *significance borders* in the context of an ACF are the thresholds used to determine whether a given autocorrelation value at a specific lag is statistically significant. A value is not significant if it lies in the range $\pm \frac{1.96}{\sqrt{N}}$, where N is the length of the time series [10].

Figures 14, 15 show the ACF of two datasets from two of our experiments. Due to the the different number of

measurement points, the number of the considered lags was adjusted to the number of measurement points.⁹

Figure 14 shows the ACF with the significance borders of the dataset from an exemplary experiment. As always, the correlation for lag 0 is 1.0 because the time series is compared with itself. Thereafter, the correlation values remain high and decrease steadily as the size of the shift increases. In the time series considered here, a trend can therefore be assumed [9].

The ACF in Fig. 15 presents a slightly different scenario. The correlation diminishes significantly more rapidly, dropping to just 0.16 at a lag size of 70. Furthermore, the correlation does not decrease uniformly: between lags 26 and 48, there is almost no change, while from lag 106 onward, the correlation begins to increase again. Based on this ACF, it can be inferred that the data lacks both trend and seasonality [9].

⁹ We considered $N/20$ lags for a time series with N data points.

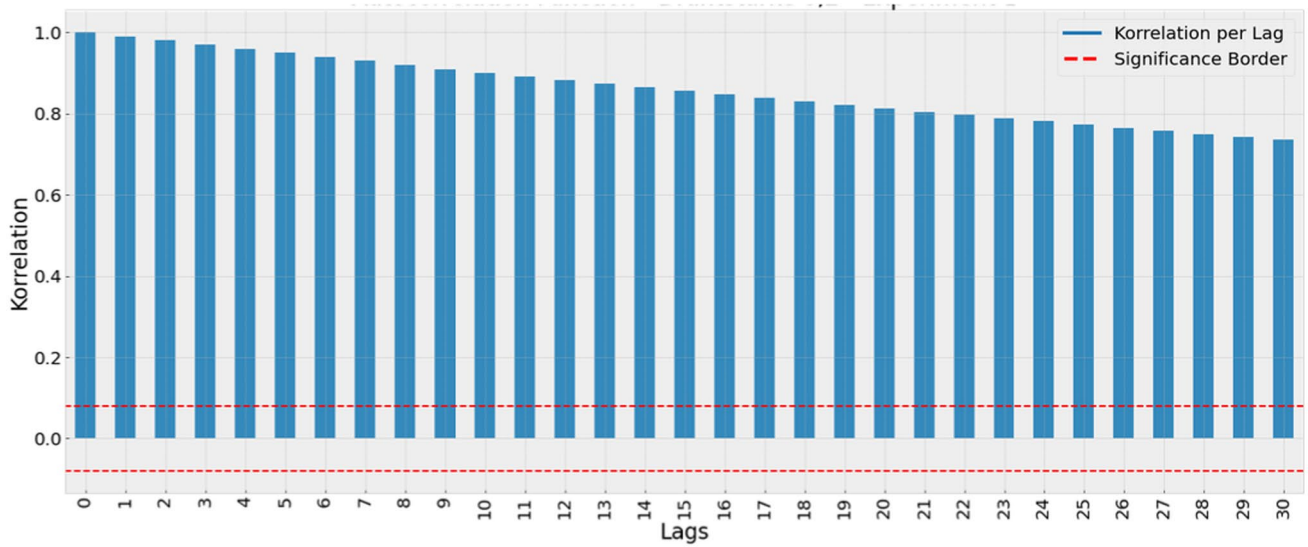


Fig. 14 ACF values from a dataset of a very good experiment: steadily decreasing ACFs over growing lags suggest strong serial dependence, such as a trend [9]

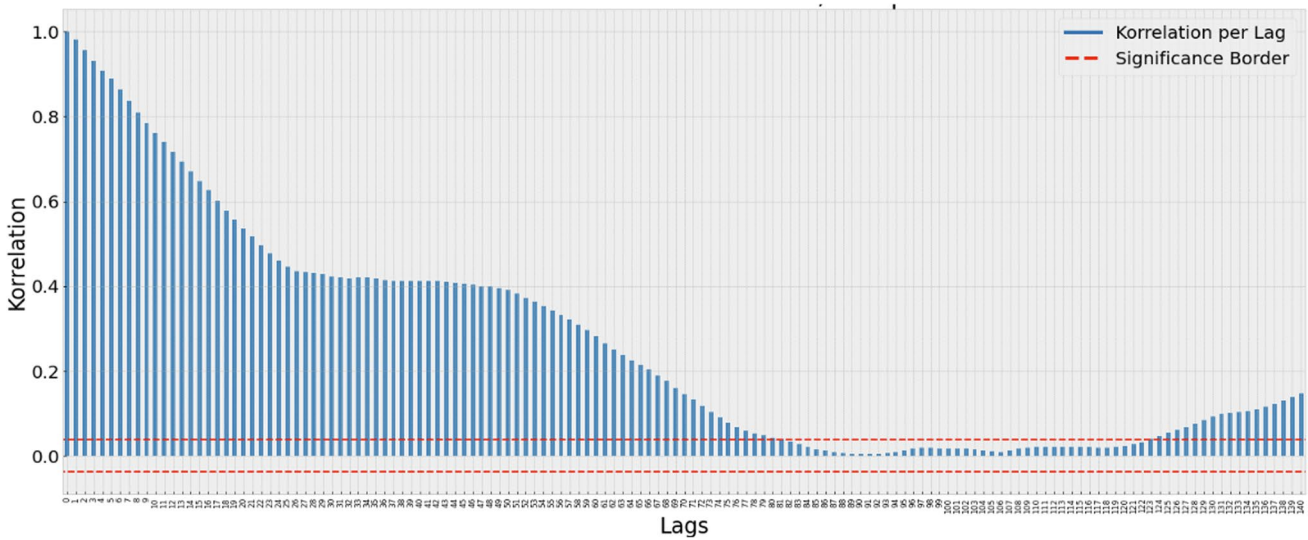


Fig. 15 ACF values from a dataset of a bad experiment: rapidly declining ACFs over increasing lags suggest low serial dependence without a sustained trend [9]

Figure 14 shows that all ACF values of the first experiment are significant. However, the time series presented in Fig. 15 contains only about 71% significant ACF values indicating that there is no trend.

Evaluating the significance of measurement data

In this subsection, we establish criteria for automatically distinguishing between significant and non-significant datasets. Each dataset, corresponding to a specific wire, can be characterized by graph-related indicators such as *time-series metrics* and the results of the Machine Learning experiments. Table 2 provides a summary of the results from all 32 conducted experiments

described in Sect. 4, with each experiment identified by its corresponding number in the first column. The subsequent columns are displayed:

- The left-hand side of the table contains columns with time series and graph-related indicators. These values can be calculated during system operation by monitoring the data measurement trajectory, without relying on any Machine Learning methods. Therefore, we want to define criteria based on these measurements to decide whether a particular wire can be used to diagnose the fatigue condition of the road.

Table 2 Time series indicators and Machine Learning results for conducted experiments. Experiments with good data quality are marked with * (in the first column)

No.	F_T	PACFS [%]	MaxNegGap [%]	Precision	Precision*	CDist
1*	0.99	100	9.63	0.99	1.00	1
2	0.94	100	64.34	0.72	0.91	6
3	0.96	100	49.87	0.74	0.85	5
4	0.57	100	96.07	0.35	0.53	9
5	0.85	75	98.30	0.11	0.31	6
6	0.98	100	46.12	0.47	0.73	5
7	0.99	100	98.36	0.70	0.89	2
8	0.87	100	83.85	0.25	0.47	9
9	0.91	100	66.93	0.34	0.64	7
10	0.92	100	50.36	0.77	0.89	4
11	0.45	100	97.34	0.16	0.36	9
12*	0.98	100	25.79	0.65	0.96	2
13	0.95	100	97.51	0.35	0.59	8
14	0.56	100	84.32	0.19	0.40	9
15	0.70	100	73.58	0.38	0.50	8
16	0.09	73	88.29	0.15	0.40	9
17*	0.95	100	04.02	0.98	1.00	1
18	0.24	99	98.28	0.13	0.34	9
19	0.97	100	82.39	0.72	0.96	2
20	0.32	89	97.65	0.40	0.58	9
21	0.62	100	63.64	0.26	0.55	7
22	0.56	100	98.16	0.36	0.47	9
23	0.49	100	98.01	0.36	0.61	9
24	0.55	100	98.79	0.32	0.44	9
25*	0.99	100	14.73	0.93	1.00	1
26	0.84	100	49.42	0.33	0.73	5
27	0.56	100	98.48	0.18	0.39	9
28	0.46	100	90.91	0.18	0.41	9
29	0.68	100	98.08	0.29	0.53	9
30	0.59	100	98.61	0.27	0.52	9
31	0.76	100	50.18	0.51	0.76	8
32	0.03	79	95.18	0.21	0.36	9

Trend strength (F_T): Using a STL decomposition, the *trend strength* of the underlying time series was calculated according to Eq. 8. A meaningful voltage curve should exhibit a continuous increase and demonstrate a strong trend, reflected by an F_T value close to 1.0. The results presented in Table 2 show trend values ranging from 0.03 to 0.99. Notably, a strong trend emerges as a critical prerequisite for achieving high prediction accuracy.

Percentage of ACF Significance (PACFS): Based on the autocorrelation function, as defined in Eq. 9, we compute the $ACF(k)$ values for all lags $k \in [1, N/20]$, where N is the total number of data points in the time series. Using the significance borders $\pm \frac{1.96}{\sqrt{N}}$, we define PACFS as the percentage of these $ACF(k)$ values that exceed the boundaries in absolute value, i.e., the proportion of lags for which the autocorrelation is statistically significant.

Table 2 shows *PACFS* values ranging from 79 to 100%. High *PACFS* values suggest the presence of structured patterns, indicating that current values are strongly dependent on recent values. Experiments with high precision are often associated with both a strong trend strength and a high *PACFS* value. This suggests a complex behavior characterized by the coexistence of a pronounced trend and a distinct periodic pattern.

Maximum negative gap (MaxNegGap): It is expected that the voltage values generated by the sensor fabric will increase continuously as the road pavement wears down. A decrease in voltage values, therefore, suggests a malfunctioning sensor. However, due to measurement inaccuracies, it is relatively common for a subsequent voltage value to be slightly smaller than its predecessor. Consequently, only a sustained decrease over an extended period is indicative of a faulty or unusable sensor. To identify such prolonged negative trends, we determined the *maximum negative gap (MaxNegGap)*

K_n in the graph, defined as the largest lag K_n with a negative trend, i.e. where two corresponding observations satisfy $x_i > x_{i+K_n}$. To ensure comparability of gap sizes across different experiments, we calculated *MaxNegGap* as a percentage of the total number of data points in each corresponding experiment. Small values indicate the expected course, where only few and short voltage decreases can be observed. Table 2 shows *MaxNegGap* values between about 4 and 98%. For instance, experiment number 19 in Table 2 has a relative negative gap size of about 82%, which means that a decreasing voltage interval covers almost all observations and therefore makes this experiment unusable.

- The columns on the right-hand side of the table contain performance indicators of the Machine Learning approach. These values require a Machine Learning model and cannot be determined during the operation of the monitoring system.

*Precision and precision**: As explained in Sect. 3.2, the well-known *precision* metrics indicate how well the voltage measurements from the sensor fabric can be used to determine the fatigue state. The proposed metrics, *prediction**, also considers a classification of an adjacent fatigue class as acceptable. As a result, $precision* \geq precision$ always holds. Experiment 12 in Table 2, illustrated by Figs. 9, 10, gives an example where *precision** is much greater than *precision*. Across all experiments, it can be seen that the values for *precision* and *precision** cover a wide range between 0.13 and 1.0.

Misclassification distances (CDist): Another criterion for evaluating the quality of the Machine Learning model is the distribution of forecast errors across classes. For example, the confusion matrix in Fig. 10 shows that some class 0 data points were misclassified as class 1 or class 2, resulting in a maximum distance of 2 for this fatigue class. In contrast, the confusion matrix from our poorly performing experiment (Fig. 12) reveals that class 0 data points were misclassified as class 9, leading to a maximum distance of 9. We have added a corresponding column in Table 2 to display *CDist*, the maximum distance between the correct and the misclassified class in the confusion matrix across all entries. In particular, *CDist* indicates the largest single misclassification, as opposed to the *prediction** indicator, which assesses the average performance of the model across all predictions.

First, with all the information provided in Table 2, we have to decide which of the experiments yield useful data. An obvious criteria are *precision* and *precision**: if the Machine Learning model achieves *precision** values greater than 0.9 we could assume that the classification works very good. Following this argument, experiments 1, 2, 7, 10, 12, 17, 19, 25 seem to provide useful data.¹⁰ But unfortunately, a high precision value is necessary but not sufficient for meaningful data. In a few experiments we could observe a *continuously decreasing voltage curve*, which does not make any sense, but could nevertheless be mapped perfectly to the fatigue states determined by the strain measurements as described in Section 2.4. This observed decrease in electrical voltage during these experiments can be attributed to the heterogeneous nature of asphalt. As a composite material, asphalt contains varying grain structures and voids, which can shift under load. These internal rearrangements, along with the closure of microcracks—particularly those oriented perpendicular to the sensor wire—can locally relieve or compress the wire.¹¹ This mechanical interaction affects the sensor's strain response, resulting in a downward trend in the measured voltage over time.

An illustrative case of this issue is provided by the model trained on data from experiment 19. Although it achieves a high *precision** value of 0.96, the corresponding voltage signal exhibits a continuous decline over time (see Fig. 16). This trend contradicts the expected physical response, indicating that the sensor wire used in this experiment could not be used for road state diagnosis and should be excluded from both model training and diagnostic evaluation. The same phenomenon is also observed in Experiments 2, 3, 7, and 10. While all of these exhibit strong predictive performance with *precision** values ≥ 0.85 , they also show pronounced anomalies in the form of large negative deviations, with *MaxNegGap* $> 49\%$.

As a result, the data from experiments 1, 12, 17 and 25 can be considered sufficiently reliable for determining the road condition. These experiments are marked with an '*' in the first column of Table 2.

To assess the usability of the measured data during the operation of the monitoring system, we rely on the three available key indicators: *trend strength* (F_T), *Percentage of ACF Significance* (*PACFS*) and *maximum negative gap* (*MaxNegGap*).

Based on the analysis of Table 2, we observe that all valid datasets satisfy the following conditions: $F_T \geq 0.95$, $PACFS = 100\%$ and $MaxNegGap < 26\%$.¹² Moreover, since none of the invalid datasets meet all three

¹⁰ Experiments 7 and 10 are also in this range and deliver $precision* \geq 0.89$

¹¹ Since these are local phenomena no conclusions can be drawn from them regarding the state of degradation.

¹² Note that all of the three criteria need to be considered. For instance, it is possible for a dataset to exhibit a low *MaxNegGap* but also a low trend strength, although this does not occur in the present data.

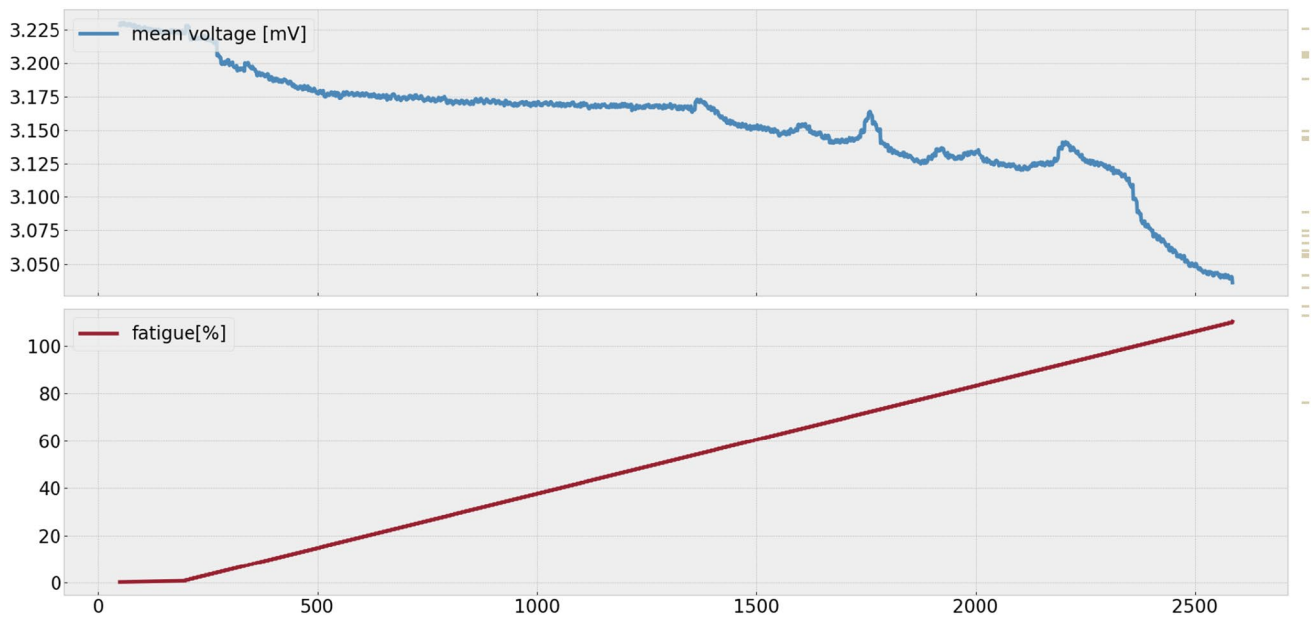


Fig. 16 Example of a continuously decreasing voltage curve (experiment 19): although this provides high precision* = 0.96, it contradicts the expected physical reaction

of these thresholds, we consider the above conditions to be both necessary and sufficient for identifying valid measurement data. However, due to the limited number of high-quality datasets currently available, we propose a more relaxed set of criteria to classify a dataset as useful, as defined in Eq. 10:

$$F_T > 0.90 \wedge PACFS > 90\% \wedge MaxNegGap < 30\% \quad (10)$$

While the 30% threshold for the gap size is somewhat arbitrary; our observations indicate that clearly unsuitable experiments consistently exceed a *MaxNegGap* of 40%. As more experimental data become available, these thresholds can be tightened to ensure suitable data. It is important to note that in practical deployment, the system operates with a large number of sensor wires. Therefore, it is sufficient if only a small subset of them works reliably (and can be identified).

Related work: AI in road quality assessment

Several studies have been conducted to develop innovative road quality assessment methods in recent years. These include several proposals that use AI technologies for data analytics.

The sensing technologies in the area of AI-based road monitoring can be roughly categorized into two classes, namely *vision-based* and *sensor-based* [17, 18]:

1. *Vision-based sensing technologies* include vehicle embedded cameras, RGB sensors, LiDAR, satellite imagery, drone-acquired images etc.
2. *Sensor-based sensing technologies*, especially vibration-based, include vehicle mounted accelerometers, gyroscopes, smartphone inertial sensors, wireless vibration meters etc.

Several studies focus on only one of the above sensing categories, but there are also studies that use both categories in combination, e.g. in [19]. A review on smartphone applications for pavement condition monitoring is given by Al-Sabaei et al. [20].

Depending on the type of measuring device and the sensing method, two forms of monitoring can be distinguished [17]:

1. *Static road condition monitoring* uses sensing devices that are installed at a specific location for a specific period of time.
2. *Dynamic road condition monitoring* is based on sensors that can be mounted on or in a moving vehicle to collect data about the road sections traveled while driving.

A comparative evaluation of smart sensors employed in road inspection can be found in Ranyal et al. [2]

Most AI-assisted approaches to date have focused on evaluating *road surface quality* using vision-based or vibration-based methods, rather than assessing the underlying road layers. For instance, Kobadagi et al. [21] provides a survey on various methods of potholes detection on road

pavements. Jan et al. [22] presents a survey specifically on crowdsensing approaches for road pavement condition monitoring. Several of these approaches rely on AI methods.

Due to recent advancements in data analytics technologies and the easy and inexpensive availability of high-performance computing resources, several innovative proposals have been published that employ Artificial Intelligence methodologies to assess, classify and localize structural road degradation using data collected with next-generation sensor technology [2]. For instance, Mancel et al. [23] developed an AI- and vision-based approach specifically for secondary road networks, for which conventional manual inspection techniques generally appear impractical from a technical and economic point of view.

The main AI techniques employed are Machine Learning methods based on manual feature extraction as well as automated Deep Learning methods that require large volumes of training data [2]:

- *Machine learning* models, including Random Forrest, Decision Tree, KNN, SVN, k-means clustering, e.g. used in [17, 24, 25]
- *Deep learning* models, including different kinds of convolutional neural networks (CNN), generative adversarial network (GAN), Long Short-Term Memory (LSTM), e.g. used in [18, 26–28]

For systematic reviews on AI-assisted road condition monitoring studies please refer to the reviews provided by Ranyal et al. [2], Cao et al. [29] and Leukel et al. [30].

In summary, road condition monitoring in general, particularly using AI methods, represents an active field of research today. We note that the published studies on AI-assisted road degradation monitoring have largely focused on assessing the quality of the pavement surface. These methods rely on an external perspective, typically using vibration data or various types of image data, and therefore are limited to analyzing superficial road conditions. In contrast, our work aims to evaluate the structural condition of the internal road construction. Since the asphalt base layer forms the structural foundation of a road, we have developed a novel hybrid sensorized fabric embedded directly into this layer. This enables the collection of continuous electrical voltage data which, through the application of Machine Learning models, provide unprecedented insights into the inner condition of the road. An approach not previously explored in the context of road condition monitoring.

Both road condition monitoring methods, surface level and base layer approaches, complement each other because they address different aspects of road degradation assessment.

Conclusion

Efficient monitoring of the condition of road infrastructure is essential for the provision of a reliable, resilient and sustainable mobility and transportation network. In this work, we presented a new data-driven assessment approach based on innovative fabric-based sensor material and Machine Learning-based diagnosis. In contrast to most other approaches, the sensor fabric is embedded into the asphalt base layer. The emitted sensor measurements are evaluated by a Machine Learning model so that the degradation state of the road section can be derived automatically and continuously.

The sensor system consists of a bio-based fabric made of flax fibers and sensor wires, which makes it inexpensive to manufacture and, thanks to its resistance to deformation, able to withstand the high stress loads during construction work. The sensor fabric thus meets two important requirements for large-scale deployment under real-world conditions. The innovative integration of advanced sensor fabric (smart measurement) and AI-driven data analysis (smart analytics) enables smart road condition monitoring in real time.

Extensive experiments demonstrate the feasibility and functionality of the approach. Several asphalt specimens with embedded sensor material were subject to controlled load stress tests under laboratory conditions. The measurements were used to train and evaluate the ML model with promising results. In addition, time series-based procedures were used to differentiate between significant and unusable sensor measurements.

The proposed approach significantly expands the research on AI-assisted road classification by developing a smart AI-based sensor system for the asphalt base layer.

Due to the revealed relevance of data quality in our approach, one focus of our future work will be the improvement of data quality of the measurement series, e.g., by means of outlier extraction, anomalies detection and noise reduction by smoothing.

An important aspect of our ongoing research is the investigation of the structural impact of embedding the sensor fabric in the road construction. First shear displacement tests between asphalt layers have already been performed, but further tests are necessary to be able to assess the effect on road durability.

Durability and longevity are important criteria for road construction. We expect our sensor system to have a long service life, as the copper-nickel material of the wire prevents corrosion. The flax carrier yarn serves only as carrier material and is explicitly allowed to degrade. The contact points between the wires and the electrical measuring system are potential corrosion points, but are protected as well as possible by appropriate isolation. Such possible installation

problems will be addressed in the follow-up project, which will conduct long-term studies on the installation conditions and the effects of the proposed sensor system in real-world operation. For this purpose, two real-life field test tracks are planned as the next steps:

In an initial field test track, various variants of the sensor fabric (especially different spacing between the wires) will be installed using heavy road construction machinery on a large-scale road section. The aim is to investigate the challenges of industrial road construction work and develop practical installation guidelines and procedures that ensure the robust and reliable installation of the hybrid sensor system in real-world construction environments. Moreover, additional sensors will also be installed to measure environmental influences and traffic congestion.

In addition, for a long-term test, a large section of a public asphalt road will be equipped with the fabric-based sensor material. This real test road section will serve to verify the laboratory results under realistic conditions with real traffic loads, real environmental conditions, on a real time scale, and subject to real road maintenance operations.

To address the anticipated challenges of properly embedding the sensor fabric into the asphalt, multiple parallel wires should be integrated within a section of the fabric. This design allows for the utilization of redundant data measurements from a single road section, enhancing robustness and reliability.

Another research direction focuses on predicting the future development of road conditions. This approach should also consider additional factors, such as the anticipated changes in environmental conditions and the projected traffic load, to provide more comprehensive insights.

Author contributions All authors contributed to the study conception and design. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the mFUND funding program of the Federal Ministry for Digital and Transport Affairs (Bundesministerium für Digitales und Verkehr, BMDV), Germany, through grants 19F1070 and 19F2286.

Data availability The datasets generated during and/or analysed during the current study are available at <https://doi.org/10.25625/R1CN7N>.

Code availability The source code of the system is available at <https://doi.org/10.25625/PQDAEQ>.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Jia M, Xu J, Gao C, Mu M, Guangxun E. (2023) Long-term cross-slope variation in highways built on soft soil under coupling action of traffic load and consolidation. *Sustainability* 15(1). <https://doi.org/10.3390/su15010033>
2. Ranyal E, Sadhu A, Jain K. (2022). Road condition monitoring using smart sensing and artificial intelligence: A review. *Sensors* 22(8). <https://doi.org/10.3390/s22083044>
3. Shadhar MH, Mohammed MM, Abdullah MH, Shather AH, Alalwan HA. (2023) Self-healing of concrete using bacteria: investigation of the impact of the process s conditions. *Innovative Infrastructure Solutions*. 8(4):115. <https://doi.org/10.1007/s41062-023-01079-9>
4. Shadhar MH, Salih BMM, Khaleel OR, Kadhim YM, Mohammed MM, Alalwan HA (2024) The influence of eggshell nanoparticles as a partial replacement of cement in concrete. *Innovative Infrastructure Solutions* 9(12):1–11. <https://doi.org/10.1007/s41062-024-01765-2>
5. Shadhar MH, Hassen DR, Kadhim YM, Araibi AS, Khaleel OR, Salih BMM, Samad AAA, Mohammed MM, Alalwan HA (2025) Experimental and numerical investigation of RC deep beams strengthened in shear using NSM-CFRP bars technique with and without steel stirrups. *Innovative Infrastructure Solutions* 10(3):1–25. <https://doi.org/10.1007/s41062-025-01915-0>
6. Radopoulou SC, Brilakis I (2016) Improving road asset condition monitoring. *Transportation Research Procedia* 14:3004–3012. <https://doi.org/10.1016/j.trpro.2016.05.436>. (Transport Research Arena TRA2016)
7. Argyroudis SA, Mitoulis SA, Chatzi E, Baker JW, Brilakis I, Gkoumas K, Vousdoukas M, Hynes W, Carluccio S, Keou O, Frangopol DM, Linkov I (2022) Digital technologies can enhance climate resilience of critical infrastructure. *Clim Risk Manag* 35:100387. <https://doi.org/10.1016/j.crm.2021.100387>
8. Herrmann J, Kayser S, Leopold M, Dunkel J (2024) Sensor integration in asphalt for data-based degradation monitoring. In: Akh-noukh A, Kaloush K, Souliman MI, Chang C (eds) *A pathway to safe, smart, and resilient road and mobility networks*. Springer, Cham, pp 103–116
9. Greve M (2022) Klassifizierung von Sensordaten mithilfe von Machine-Learning und Zeitreihenverfahren. Master's thesis, Hochschule Hannover, Department of Computer Science
10. Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2015) *Time series analysis: forecasting and control*, 5th edn. Wiley, Hoboken, NJ

11. Brockwell PJ, Davis RA (2002) Introduction to Time series and forecasting, 2nd edn. Springer, New York
12. Hyndman RJ, Athanasopoulos G. (2021) Forecasting: Principles and practice, 3rd Edition. OTexts, Melbourne, Australia. Accessed on 29 Dec 2024
13. Theodosiou M (2011) Forecasting monthly and quarterly time series using stl decomposition. *Int J Forecast* 27(4):1178–1195
14. Cleveland RB, Cleveland WS, McRae JE, Terpenning I (1990) Stl: A seasonal-trend decomposition procedure based on loess. *J Off Stat* 6(1):3–73
15. Hyndman RJ (2008) Time series decomposition and trend analysis. *Comput Stat Data Anal* 52(8):3989–4001
16. Bovas S, Kizilay NR (2020) A comprehensive review on autocorrelation function for time series. *J Appl Math* 2020:1–12
17. Shtayat A, Moridpour S, Best B, Abuhassan M (2023) Using supervised machine learning algorithms in pavement degradation monitoring. *Int J Transp Sci Technol* 12(2):628–639. <https://doi.org/10.1016/j.ijst.2022.10.001>
18. Tiwari S, Bhandari R, Raman B. (2020) Roadcare: a deep-learning based approach to quantifying road surface quality. In: Proceedings of the 3rd ACM SIGCAS conference on computing and sustainable societies. COMPASS '20, pp. 231–242. Association for Computing Machinery, New York, NY, USA <https://doi.org/10.1145/3378393.3402284>
19. Shtayat A. (2022) Dynamic monitoring and prediction of pavement degradation. Phd thesis, RMIT University
20. Al-Sabaeei AM, Souliman MI, Jagadeesh A (2024) Smartphone applications for pavement condition monitoring: A review. *Constr Build Mater* 410:134207. <https://doi.org/10.1016/j.conbuildmat.2023.134207>
21. Kodabagi M, Pavithra S, Bhavana N, Rakshitha H. (2022) Road condition monitoring and information system: Survey. In: 2022 Third international conference on smart technologies in computing, electrical and electronics (ICSTCEE), pp. 1–5. <https://doi.org/10.1109/ICSTCEE56972.2022.10099926>
22. Jan M, Khattak KS, Khan ZH, Gulliver TA, Altamimi AB (2023) Crowdsensing for road pavement condition monitoring: Trends, limitations, and opportunities. *IEEE Access* 11:133143–133159. <https://doi.org/10.1109/ACCESS.2023.3332667>
23. Mancel A, Nguyen TS, Uguet Canal N. (2023) Optimize pavement monitoring using artificial intelligence. In: 2nd International conference on asphalt 4.0. <https://www.h-a-d.hr/pubfile.php?id=1148>
24. Huang L-L, Lin J-D, Huang W-H, Kuo C-H, Huang M-Y. Application of automated pavement inspection technology in provincial highway pavement maintenance decision-making. *Applied Sciences* 14(15) (2024). <https://doi.org/10.3390/app14156549>
25. Rischioni LG, Babu A, Baumgartner SV, Krieger G (2023) Machine learning approaches for road condition monitoring using synthetic aperture radar. *IEEE J Sel Top Appl Earth Obs Remote Sens* 16:3070–3082. <https://doi.org/10.1109/JSTARS.2023.3258059>
26. Ahmed T, Ejaz N, Choudhury S (2024) Redefining Real-Time Road Quality Analysis With Vision Transformers on Edge Devices. *IEEE Transac Artif Intell* 5(10):4972–4983. <https://doi.org/10.1109/TAI.2024.3394797>
27. Mishra R, Gupta HP, Dutta T (2021) A road health monitoring system using sensors in optimal deep neural network. *IEEE Sens J* 21(14):15527–15534. <https://doi.org/10.1109/JSEN.2020.3005998>
28. Shakhovska N, Yakovyna V, Mysak M, Mitoulis S-A, Argyroudis S, Syerov Y. (2024). Real-time monitoring of road networks for pavement damage detection based on preprocessing and neural networks. *Big Data Cognit Comput* 8(10). <https://doi.org/10.3390/bdcc8100136>
29. Cao W, Liu Q, He Z (2020) Review of pavement defect detection methods. *IEEE Access* 8:14531–14544. <https://doi.org/10.1109/ACCESS.2020.2966881>
30. Leukel J, Scheurer L, Sugumaran V (2024) Machine learning models for predicting physical properties in asphalt road construction: A systematic review. *Construct Building Mater* 440:137397. <https://doi.org/10.1016/j.conbuildmat.2024.137397>