

# DOCTORAL (PhD) THESIS

ROLAND NAGY

DOI:10.18136/PE.2026.1005



University of Pannonia  
Veszprém  
2026





University of Pannonia  
Faculty of Engineering

Automotive and mechatronics applications of IMU sensors

DOCTORAL (PhD) THESIS  
Roland Nagy

Supervisor:  
István Szalai, DSc, professor

Doctoral School of Chemical Engineering and Material Sciences  
Veszprém 2026



Automotive and mechatronics applications of IMU sensors

The thesis was prepared for the award of a doctoral degree (PhD) within the framework of the  
Doctoral School of Chemical Engineering and Material Sciences at University of Pannonia

in the discipline of Materials science and technology

written by: Roland Nagy

Supervisor: István Szalai

I recommend the dissertation for acceptance: yes / no.

.....  
supervisor

I recommend the dissertation for peer review.

.....  
chair of the DDHC

The PhD-candidate has achieved ..... % at the public debate.

The composition of the Final Examination Committee:

chair:.....

reviewers:.....

members:.....

Veszprém, .....

.....  
chair of the committee

Qualification of degree: .....

Veszprém, .....

.....  
chair of the UDHC



## Abstract

Vibrations in road vehicles cause several harmful effects, health problems can occur for the passengers, and mechanical damage can occur to the vehicle components. Given the health, safety, and financial issues that arise, keeping the road network in good condition and detecting road defects as early as possible requires an extensive monitoring system.

Related to these issues, my research is concerned with a comprehensive methodology for automated road surface quality assessment using vehicle-mounted vibration sensors and machine learning techniques. The basic goal is the automated categorization of internally homogeneous and coherent quality road sections based on predefined road quality classes. Motivated by the deficiencies of the pavement condition and measurement methods, particularly on lower-tier road networks, the research investigates how multi-sensor systems with data-driven models can offer a scalable solution for road monitoring.

To support this aim, a custom multi-sensor measurement platform was developed, combining an IMU installed in the vehicle cabin and piezoelectric accelerometers mounted on the suspension. This setup allowed for simultaneous recording of cabin and axle-level vibrations under varied driving conditions. The raw sensor data underwent a dedicated preprocessing workflow, including noise reduction, spatial alignment, and spectral feature extraction.

A subsequent supervised machine learning model requires high-quality labeled data, which is typically unavailable at the necessary resolution. Therefore, I introduced a Machine-In-the-Loop annotation method novel in this field for automated pavement quality assessment. The clustering algorithm categorizes road sections by quality, significantly reducing reliance on manual labeling. The annotated dataset served as the basis for training the subsequent supervised classification models.

In addition, I paid attention to previously unexplored areas in the research, thereby examining the information content of sensors placed at different points of the vehicle and the necessity of multi-sensor measurement. Furthermore, I examined the negative impact of measurement signals from abrupt vehicle maneuvers on the accuracy of road quality measurement, and then, having ascertained this, I introduced a detection and correction method in order to reduce these negative effects. The methodology is validated through extensive real-world testing across diverse road conditions using multiple vehicle types.



## Kivonat

A közúti járművekben fellépő vibrációk számos káros hatást okozhatnak, gyakoriak az egészségügyi problémák az utasoknál, valamint mechanikai meghibásodásokat okozhatnak a járműben. Az egészségügyi, biztonsági és pénzügyi következmények miatt elengedhetetlen a közúthálózatban az úthibák mielőbbi felismerése, amelyhez kiterjedt monitoringrendszer szükséges.

Ehhez kapcsolódóan kutatásomban egy átfogó módszertannal foglalkozom, amely a járműre szerelt rezgésérzékelők és gépi tanulási technikák alkalmazásával teszi lehetővé az útfelületi minőség automatikus kiértékelését. Az alapvető cél az, hogy előre meghatározott útfelületi minőségi osztályok alapján automatikusan kategorizálhatóvá váljanak a belsőleg homogén és koherens minőségű útszakaszok. A kutatás a burkolatminőség és a mérési módszerek hiányosságaira alapozva vizsgálja, hogy a többszenzoros rendszerek kiegészítve az adatalapú modellekkel hogyan kínálhatnak skálázható megoldást az útfelügyeletre.

Ehhez egy egyedi fejlesztésű, több szenzorból álló mérőrendszert vezettem be, amely kombinálja a jármű kabinjában elhelyezett IMU szenzort és a felfüggesztésre rögzített piezoelektromos gyorsulásmérőket. Ezáltal a kabin- és tengelyszintű vibrációk egyidejű rögzíthetők különféle tesztelési körülmények között. A nyers szenzoradatok egyedileg fejlesztett előfeldolgozási folyamaton esnek át, amely zajcsökkentést, térbeli adatelőkészítést és spektrális jellemzők kinyerését tartalmazza.

A felügyelt gépi tanulási modellekhez jó minőségű, címkézett adatok szükségesek, amelyek azonban rendszerint nem állnak rendelkezésre megfelelő felbontásban. Ennek megoldására a területen újdonságnak számító Machine-In-the-Loop címkéző eljárást vezettem be. A klaszterező algoritmus az útszakaszokat minőség szerint klaszterezi, jelentősen csökkentve a manuális beavatkozás időigényét. A létrehozott címkézett adatbázis szolgált alapul a későbbi felügyelt klasszifikációs modellek tanításához.

Emellett a kutatás során olyan korábban nem vizsgált tényezőket tanulmányoztam, mint a jármű különböző pontjaira elhelyezett szenzorok információtartalma és a több szenzoros mérés szükségessége. Továbbá vizsgáltam a hirtelen járműmanőverekből származó jelkomponensek útfelületi minőségmérésre gyakorolt negatív hatását, majd ezen hatás megerősítését követően detektálási és korrekciós módszert vezettem. A módszertan érvényességét közutakon végzett hosszú távú mérésekkel igazoltam, változó minőségű útszakaszokon és többféle járműtípus alkalmazásával.





## Abstrakt

Vibrationer i vejgående køretøjer medfører flere skadelige effekter; passagerer kan opleve helbredsproblemer, og køretøjskomponenter kan pådrage sig mekaniske skader. På baggrund af de sundhedsmæssige, sikkerhedsmæssige og økonomiske konsekvenser, som vibrationer kan give anledning til, er det nødvendigt med et omfattende overvågningssystem for tidlig detektion af vejdefekter.

I relation til disse problemstillinger omhandler min forskning en omfattende metodologi vejoverfladers kvalitet ved hjælp af køretøjsmonterede vibrationssensorer og maskinlæringsteknikker. Det overordnede mål er en automatiseret kategorisering af internt homogene og sammenhængende vejstrækninger baseret på foruddefinerede vejklasser. Motiveret af manglerne ved eksisterende metoder til vurdering af belægningstilstand undersøger forskningen, hvordan multisensorsystemer kombineret med datadrevne modeller kan udgøre en skalerbar løsning til vejinfrastrukturens overvågning.

Til dette formål blev en skræddersyet multisensor-måleplatform udviklet, bestående af en IMU installeret i kabinen samt piezoelektriske accelerometre monteret på køretøjets affjedring. Denne opsætning muliggjorde samtidig registrering af vibrationer i bil kabine og akselniveau. De rå sensor-data gennemgik en dedikeret præprocesseringspipeline, herunder støjreduktion, rumlig justering og spektral feature-ekstraktion.

En efterfølgende superviseret maskinlæringsmodel krævede høj-kvalitets annoterede data, som typisk ikke er tilgængelige i den fornødne opløsning. Derfor introducerede jeg en Machine-In-the-Loop-annoteringsmetode til automatiseret vurdering af vejbelægning. En klyngealgoritme kategoriserer vejstrækninger efter kvalitet og reducerer dermed afhængigheden af manuel annotering betydeligt. Det annoterede datasæt dannede grundlag for træningen af de efterfølgende superviserede klassifikationsmodeller.

Derudover blev der fokuseret på hidtil uudforskede aspekter, herunder undersøgelse af informationsindholdet fra sensorer placeret på forskellige punkter af køretøjet samt nødvendigheden af multisensoriske målinger. Ydermere blev de negative effekter af målesignaler, der stammer fra pludselige køremanøvrer, analyseret i relation til nøjagtigheden af vejtilstandsmålingerne. Efter at have identificeret denne problematik, blev en metode til detektion og korrektion udviklet for at reducere de negative effekter. Metodologien er blevet valideret gennem omfattende afprøvning i virkelige kørselsforhold på forskellige vejtyper med anvendelse af flere køretøjsmodeller.





# Contents

|                                                                          |           |
|--------------------------------------------------------------------------|-----------|
| <b>List of Symbols</b>                                                   | <b>1</b>  |
| <b>1 Introduction</b>                                                    | <b>5</b>  |
| <b>2 Fundamentals and literature research</b>                            | <b>7</b>  |
| 2.1 Road condition . . . . .                                             | 7         |
| 2.1.1 Causes of road discontinuities . . . . .                           | 7         |
| 2.1.2 Effects of road defects . . . . .                                  | 8         |
| 2.1.3 State-of-the-art of road condition monitoring . . . . .            | 9         |
| 2.1.4 Metrics of pavement quality . . . . .                              | 13        |
| 2.2 Vehicle body motion characteristics . . . . .                        | 14        |
| 2.2.1 Coordinate systems in vehicle dynamics . . . . .                   | 14        |
| 2.2.2 Coordinate Systems . . . . .                                       | 15        |
| 2.2.3 Vehicle dynamics and models . . . . .                              | 17        |
| 2.2.4 Vehicle vibration . . . . .                                        | 19        |
| 2.3 Review of previous vibration-based road monitoring systems . . . . . | 22        |
| 2.3.1 Data acquisition . . . . .                                         | 23        |
| 2.3.2 Data mining . . . . .                                              | 24        |
| 2.3.3 Significance of additional factors and generalization . . . . .    | 31        |
| 2.3.4 Inertial sensor-based vehicle maneuver detection . . . . .         | 33        |
| <b>3 Concept of the novel system</b>                                     | <b>36</b> |
| 3.1 Aim of this work and research objectives . . . . .                   | 36        |
| 3.2 Novel methods for road condition measurement . . . . .               | 38        |
| 3.3 Hardware and embedded software structure of the system . . . . .     | 41        |
| <b>4 Implementation of the software framework</b>                        | <b>48</b> |
| 4.1 Structure of the algorithm . . . . .                                 | 48        |
| 4.1.1 Detailed approach and setup . . . . .                              | 49        |
| 4.1.2 Notation scheme . . . . .                                          | 51        |
| 4.2 General data preprocessing . . . . .                                 | 52        |
| 4.3 Aggressive maneuver detection and correction . . . . .               | 55        |
| 4.3.1 Overview of the importance of maneuver detection . . . . .         | 55        |
| 4.3.2 Maneuver labelling . . . . .                                       | 57        |

|          |                                                                                |            |
|----------|--------------------------------------------------------------------------------|------------|
| 4.3.3    | Data processing . . . . .                                                      | 59         |
| 4.3.4    | Maneuver classification model . . . . .                                        | 62         |
| 4.3.5    | Aggressive maneuver correction . . . . .                                       | 64         |
| 4.4      | Feature extraction for pavement classification . . . . .                       | 66         |
| 4.4.1    | Power Spectrum Analysis . . . . .                                              | 67         |
| 4.4.2    | Time Series Segmentation and Feature Generation . . . . .                      | 69         |
| 4.5      | Machine learning-based clustering and classification . . . . .                 | 70         |
| 4.5.1    | Machine-in-the-Loop annotation . . . . .                                       | 71         |
| 4.5.2    | Unsupervised machine learning models . . . . .                                 | 73         |
| 4.5.3    | Supervised machine learning classification . . . . .                           | 75         |
| 4.6      | Performance measures . . . . .                                                 | 77         |
| 4.6.1    | Measuring the quality of clustering . . . . .                                  | 77         |
| 4.6.2    | Classification quality metrics . . . . .                                       | 79         |
| 4.7      | Reverse geocoding and map matching . . . . .                                   | 81         |
| 4.7.1    | Theory and implementation of map matching . . . . .                            | 81         |
| 4.7.2    | Address assignment from coordinates with Reverse Geocoding . . . . .           | 82         |
| 4.7.3    | Structure of the final output of the algorithm . . . . .                       | 82         |
| <b>5</b> | <b>Results and discussion</b>                                                  | <b>84</b>  |
| 5.1      | Measurement conditions . . . . .                                               | 84         |
| 5.2      | Data preprocessing in practice . . . . .                                       | 86         |
| 5.3      | Maneuver detection and correction . . . . .                                    | 88         |
| 5.3.1    | Importance of maneuvers . . . . .                                              | 88         |
| 5.3.2    | Evaluation of the detection algorithm . . . . .                                | 90         |
| 5.4      | Data annotation and road unevenness estimation . . . . .                       | 93         |
| 5.5      | Supervised classification model with labeled data . . . . .                    | 97         |
| 5.6      | Evaluation of the effects of sensor location and maneuver correction . . . . . | 98         |
| 5.6.1    | Approach of the investigation and evaluation . . . . .                         | 99         |
| 5.6.2    | Maneuver detection and correction experiments . . . . .                        | 101        |
| 5.6.3    | Sensor placement comparison . . . . .                                          | 102        |
| 5.6.4    | Integrated modeling and feature importance analysis . . . . .                  | 102        |
| 5.7      | Final production model and long-term measurements . . . . .                    | 106        |
| <b>6</b> | <b>Summary</b>                                                                 | <b>112</b> |
|          | <b>Theses</b>                                                                  | <b>115</b> |

|                                    |     |
|------------------------------------|-----|
| Publications related to theses     | 120 |
| Publications not related to theses | 122 |
| References                         | 123 |
| Acknowledgement                    | 130 |

## List of Symbols

|                                                                            |                                                                                        |
|----------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| $\delta$                                                                   | Cliff's Delta non-parametric effect size measure                                       |
| $X_E, Y_E, Z_E$                                                            | Earth-fixed coordinate system: horizontal plane $(X_E, Y_E)$ and vertical axis $(Z_E)$ |
| $\alpha_i$                                                                 | SVM coefficient for support vector $i$                                                 |
| $\beta_0$                                                                  | Intercept in the SVM model                                                             |
| $\boldsymbol{\mu}_k$                                                       | Mean vector of component $k$                                                           |
| $\delta x$                                                                 | Distance between spatial intervals after resampling                                    |
| $\Lambda$                                                                  | Diagonal matrix of PCA eigenvalues                                                     |
| $\lambda_k$                                                                | The $k$ th eigenvalue in PCA                                                           |
| $\mathbf{c}$                                                               | Clustering output vector, $\mathbf{c} = (c_1, \dots, c_n)^T \in \mathbb{R}^n$          |
| $\mathbf{x}'$                                                              | Infinity norm-normalized version of $\mathbf{x}$                                       |
| $\mathbf{x}$                                                               | Row vector of features for a single sample                                             |
| $\mathbf{x}_i$                                                             | Feature vector of the $i$ th observation                                               |
| $\mathbf{y}$                                                               | Classification output vector, $\mathbf{y} = (y_1, \dots, y_n)^T \in \mathbb{R}^n$      |
| $\mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ | Multivariate normal distribution for component $k$                                     |
| $\mathcal{T}$                                                              | Set of change point locations                                                          |
| $\omega_x, \omega_y, \omega_z$                                             | Angular velocity around the X, Y and Z-axis, respectively                              |
| $\pi_k$                                                                    | Mixing coefficient (prior probability) for component $k$                               |
| $\psi_V, \theta_V, \phi_V$                                                 | Yaw, pitch, and roll angles around axes $Z_V, Y_V$ , and $X_V$ , respectively          |
| $\text{pen}(\mathcal{T})$                                                  | Penalty term for complex segmentations                                                 |
| $A$                                                                        | Number of distinct classes in supervised learning                                      |
| $a$                                                                        | Average intra-cluster distance                                                         |

|                               |                                                                                     |
|-------------------------------|-------------------------------------------------------------------------------------|
| $a_x, a_y, a_z$               | Vehicle linear accelerations along the $X_V$ , $Y_V$ , and $Z_V$ axes, respectively |
| $a_{zl}, a_{zr}$              | Z-axis accelerometer data from the left side and right-side, respectively           |
| $b$                           | Average distance to the nearest other cluster                                       |
| $c_i$                         | Cluster assignment for sample $x_i$                                                 |
| $c_s$                         | Damping coefficient of the shock absorber                                           |
| $c_{tr}$                      | Damping coefficient of the tire                                                     |
| $d$                           | Degree of the polynomial kernel                                                     |
| $d_{\text{spatial}}$          | Spatial position for spectrogram evaluation                                         |
| $F(t)$                        | Excitation force acting on the wheels due to surface irregularities                 |
| $f$                           | Frequency                                                                           |
| $G$                           | Center of gravity of the whole vehicle                                              |
| $G_s, G_{uf}, G_{ur}, G_{us}$ | Centers of gravity for sprung, front/rear unsprung, and total unsprung components   |
| $j$                           | Imaginary unit, $j^2 = -1$                                                          |
| $k(x_i, x_{i'})$              | Kernel function between $x_i$ and $x_{i'}$                                          |
| $K$                           | Number of distinct clusters in unsupervised learning                                |
| $k_s$                         | Suspension spring stiffness                                                         |
| $k_{tr}$                      | Equivalent tire spring stiffness                                                    |
| $L$                           | Likelihood of the proper clustering of the unsupervised model                       |
| $lat, lon$                    | Latitude and Longitude coordinate                                                   |
| $m$                           | Total mass of the vehicle                                                           |
| $m_s, m_{uf}, m_{ur}, m_{us}$ | Masses of sprung, front unsprung, rear unsprung, and total unsprung components      |
| $N$                           | Total number of samples                                                             |



|                             |                                                                                                                    |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------|
| $n$                         | Number of distinct data points or observations                                                                     |
| $p(\mathbf{x}_i)$           | Likelihood of $\mathbf{x}_i$ under the GMM                                                                         |
| $P$                         | Signal power in frequency range $[f_1, f_2]$                                                                       |
| $p$                         | Number of measured variables                                                                                       |
| $S(d, f)$                   | Spectrogram matrix; power at position $d$ and frequency $f$                                                        |
| $S$                         | Principal component score matrix                                                                                   |
| $s$                         | Silhouette coefficient                                                                                             |
| $S_i$                       | Signal segment $i$ from the PCA score matrix                                                                       |
| $S_s$                       | Subsystem of the sprung mass (vehicle body)                                                                        |
| $S_{\text{sum}}(d)$         | Total spectral power at location $d$                                                                               |
| $S_{uf}$                    | Subsystem of the front unsprung mass                                                                               |
| $S_{ur}$                    | Subsystem of the rear unsprung mass                                                                                |
| $T(\omega)$                 | Power spectral density as function of angular frequency $\omega$                                                   |
| $T_{\text{sum, segmented}}$ | Segmented spectral energy features                                                                                 |
| $V(\mathcal{T})$            | Segment fit cost function                                                                                          |
| $V$                         | Matrix of eigenvectors (loadings) from PCA, $V \in \mathbb{R}^{m \times m}$                                        |
| $v$                         | Vehicle speed                                                                                                      |
| $X(\omega)$                 | Fourier transform of $x$ at frequency $\omega$                                                                     |
| $x$                         | Processed input matrix of size $n \times p$                                                                        |
| $x^{(\text{min})}, x_K, x'$ | Minority class sample, its neighbor, and the synthetic SMOTE sample                                                |
| $X_k$                       | Discrete Fourier Transform (DFT) coefficient $k$                                                                   |
| $X_R, Y_R$                  | Road plane coordinate along the X-axis and Y-axis                                                                  |
| $X_V, Y_V, Z_V$             | Vehicle-fixed coordinate along the longitudinal (forward), lateral (left) and vertical (upward) axis, respectively |

$X_W$  Wheel coordinate system, for the X, Y and the Z-axis, respectively

$x_{raw}$  Raw data matrix of size  $n \times p$

$y_i$  Class label for sample  $x_i$

$z_0, z_1, z_2$  Vertical elevation of the surface profile, the sprung mass and the unsprung mass, respectively

# 1 Introduction

Unevenness and defects in road surfaces have a major negative impact on passenger vehicles, their occupants and vehicle traffic. Most of the vibrations in passenger vehicles are caused by road surface irregularities, which can reduce the lifetime of the mechanical and electronic components of the vehicle, and high frequency vibrations also have a negative physiological impact on the occupants [1–4]. Furthermore, a significant proportion of road accidents in recent years have been attributed to these road defects [1, 5, 6].

For these reasons, the early detection and repair of surface defects on the road network is a priority. Repairing road defects as soon as possible is also important from a financial point of view, as the cost of reconstruction increases exponentially with the degradation of the road surface [7, 8]. The detection of defects that occur requires extensive and continuous measurements due to the continuous changes in the surface, which requires a road quality measurement system that is simple in design, highly accurate and widely usable.

The primary goal of my work is to introduce and examine the application of novel methods in this field that could potentially provide a better solution from many aspects for general road surface quality measurement and classification. This novel system should be more flexible than existing and widely used solutions, and should be operated at lower cost but with good accuracy. The basic principle of the developed solution was the application of advanced data processing techniques in the data processing and evaluation process, which allows to significantly reduce the number of hardware components. The final solution provides a scalable, cost-effective, and high-resolution road quality assessment with minimal manual intervention for road infrastructure monitoring.

My work included the investigation of the sources of errors relevant for road surface quality measurement and classification, the applicability of automated data annotation solutions based on clustering, and finally the information content of sensors placed at different locations on the test vehicle and their impact on the measurement results.

The dissertation will first provide a summary of the state of the art with a literature review (see Section 2), in which I will discuss the methods and metrics currently used as well as their shortcomings. I will also discuss the dynamic and vibration properties of vehicles, followed by a brief summary of advanced data processing methods. I will then present the implementation of the custom-developed measurement system (see Section 3). Following the hardware unit, I will detail the architecture of the software framework (see Section 4),

in which I will describe the data pre-processing, feature extraction and machine learning solutions for database labelling and classification. Then, I present the measurement results and their evaluation on a real-world road network (see Section 5). Finally, I summarize the results and observations obtained in this study (see Section 6). The dissertation concludes with a thesis-like summary of the results.

## 2 Fundamentals and literature research

The section of literature research provides an overview of prior research relevant to my work. I discuss key advancements in vehicle body motion analysis and vibration-based road quality measurement techniques. This is followed by an overview of data mining methods, including feature extraction methods, clustering-based data annotation techniques and supervised machine learning-based classification.

### 2.1 Road condition

The road profile is mainly characterized by the texture wavelength [9]. The wavelengths physically illustrate the different lengths of periodically repeating structures of the profile. The road texture is categorized based on the range of these wavelengths. It is classified as microtexture when the wavelength is less than 0.5 mm, as macrotexture when the wavelengths range from 0.5 mm to 50 mm, and as megatexture when the wavelengths fall between 50 mm and 0.5 m [9]. Additionally, the profile texture with wavelengths from 0.5 m to 50 m is referred to as unevenness. Each of these categories plays an important role in defining the interaction between the road surface and vehicles, influencing aspects such as tire grip, rolling resistance, ride comfort, and noise generation.

Microtexture primarily affects tire-road adhesion at low speeds and wet conditions, as it determines the roughness at the microscopic level, which in turn influences friction. Macrotexture, on the other hand, has a significant impact on water drainage and rolling resistance, making it an important factor for skid resistance and fuel efficiency [10]. Megatexture can lead to discomfort for vehicle occupants and increased wear on vehicle components, particularly in heavy transport applications. Unevenness, which encompasses the longest wavelengths in road surface texture, is a key factor in ride quality, affecting dynamic vehicle responses, including vibrations and handling stability [1–3].

#### 2.1.1 Causes of road discontinuities

Beyond these regularly occurring textures, the road profile can be overlaid by road discontinuities. These can be intentionally incorporated into pavement design, including features like expansion joints, railway and tramway crossings, manholes, and speed bumps. Alternatively, pavement discontinuities may arise inadvertently due to pavement failures such as potholes, cracks, and surface deformations. With a systematic maintenance, the life span of asphalt layers is 10 to 15 years and 15 to 20 years for binding layers. Concrete surfaces

can last up to 30 years and base layers 40 to 50 years [11].

However, the nominal lifetime can be greatly reduced by poor construction or by environmental factors and traffic loads during normal usage. Over time, binders lose their binding strength, and continuous tire-road contact contributes to surface wear. The extreme weather conditions also significantly affect road durability. Prolonged heat causes asphalt and concrete roads to expand, leading to cracks and deformations. In extreme cases, concrete plates stretch, reducing joint space and causing "blow-ups" [12]. Asphalt, due to its bitumen content, softens under heat, resulting in surface deformations [12]. Cold weather further accelerates damage, as water seeps into cracks, freezes, and expands, leading to potholes through repeated freeze-thaw cycles and traffic stress [12]. Lastly, heavy-load vehicle traffic, particularly axle load, plays a crucial role in road deterioration. Axle load refers to the weight a single axle carries, which increases with both vehicle and cargo weight. Studies show that a linear increase in axle load results in road damage to the fourth power [13].

### **2.1.2 Effects of road defects**

Many studies [1, 2, 14–16] have investigated the influences of bad road conditions. The negative effects of these discontinuities can be summarized as

- reduced lifetime of mechanical components,
- potential long-term health risks,
- worsening vehicle dynamics, as more difficult handling,
- influencing ride smoothness and travel comfort,
- degradation in suspension system performance,
- and the potential safety risks.

Many health problems can be associated with vibrations in the vehicle. Spinal pain, arthritis, and musculoskeletal problems are common, and in addition, body vibrations can negatively affect driver and passenger comfort levels and increase stress [1–3].

Uneven road surfaces also affect the dynamics of vehicle movements. Traction conditions are primarily determined by terrain characteristics, with implications for longitudinal and lateral wheel slips. Furthermore, surface irregularities alter the forces acting on the wheel,

which changes the handling of the vehicle in a negative direction. These can have a direct impact on the accidents that occur, with 50% of fatal road accidents in France in 2006 being linked to poor road surfaces, while in Germany in 2016, 4 accidents were caused solely by poor road surfaces [1, 5, 6].

Moreover, as surface degradation progresses, the cost of pavement repairs rises exponentially [7, 8]. The deterioration curve illustrates how surface condition declines over time, often accelerating as minor issues compound into more significant damage. This exponential trend underscores the importance of timely maintenance to avoid the higher costs associated with advanced-stage repairs.

Related to the health, safety, and financial costs listed above, it is essential to keep the road network in good condition and to repair any road defects as soon as possible. Consequently, the knowledge of the actual condition of the road infrastructure is a significant source of information. However, condition metrics are dynamic parameters that change over time, so continuous monitoring of road quality conditions is necessary.

### **2.1.3 State-of-the-art of road condition monitoring**

There are several methods available to measure road quality. The condition monitoring process typically consists of multiple steps, including data collection, defect identification, defect assessment, and the calculation of a road condition index on the principle of a regression or classification method [17]. However, the specific implementation and details of each step vary across countries and studies.

The measurement requirements can vary widely and are set by the client, usually the local road management company. These requirements are mainly defined by the expected measurement accuracy, the evaluation method and the spatial resolution. The evaluation resolution may be a continuous scale rating or a rating according to predefined quality classes. The spatial resolution often implies the definition of segments of a specific length, which usually ranges from 10 to 100 meters among the countries [18]. For these, segments of a given length are to be evaluated as individual segments with a single score. Segmentation is a commonly used method for general purpose road network monitoring, where the goal is to obtain a comprehensive picture and identify critical sections in a relatively quick and cost-effective way.

The measurement systems, which have been applied to monitor the road infrastructure are

mainly based on

- visual inspection by human surveyors,
- road profilometers, including laser distance sensors and LIDARs,
- cameras,
- or inertial vibration sensors.

The following describes the principles and properties of these most commonly used measurement methods in more detail.

### **Visual inspection based survey**

One of the primary methods for road condition monitoring is manual inspection, which is carried out by accredited surveyors who either walk along the roadway or conduct assessments from a moving vehicle. This approach involves visual evaluation and, in some cases, the use of simple measurement tools. The main advantage of manual inspection is its ability to provide detailed qualitative insights, particularly for identifying complex surface irregularities and defects that may be challenging to detect through automated methods. Additionally, surveyors can account for contextual factors, such as environmental influences or localized structural issues, that may not be fully captured by sensor-based systems. However, manual inspections are labor-intensive, time-consuming, and susceptible to human error or subjective interpretation, leading to inconsistencies in the assessment process. Furthermore, this method poses safety risks to surveyors, particularly in high-traffic areas, and may not be feasible for large-scale monitoring due to its relatively low efficiency [19, 20]. Despite these limitations, manual inspection remains a widely used approach, due to its cost-effectiveness and simplicity [18].

### **Camera and road profilometer based monitoring**

The most accurate solution for road monitoring is offered by specially adapted measuring vehicles with a wide range of sensors, including tactile and laser profilometers, as well as camera-based systems [21], as shown in Fig. 1. Mobile tactile profilometers, such as the longitudinal profile analyzer, employ trailers with suspended wheels that remain in continuous contact with the road surface, transmitting vertical movement data to a recording unit [22]. These systems provide highly accurate reference profiles but are limited by their relatively low measurement speed and susceptibility to wear over time. In contrast, 3D



laser road profiling systems use laser line scanners to detect road defects at high speeds [23, 24], reaching up to 100 km/h [25]. These systems offer precise and efficient defect identification, and recent advancements in data processing have significantly reduced computation times for 3D image optimization. However, their high cost—up to 220,000 USD [26] per vehicle—and the need for trained operators present limitations. Additionally, the presence of debris, salt, or snow on the road can interfere with accurate defect detection.

Camera-based systems provide alternative techniques to identify road hazards and defects. These methods are particularly cost-effective, as modern vehicles are often equipped with cameras, allowing for extensive road network coverage. Machine learning techniques have further enhanced the accuracy of defect detection from images. However, camera-based approaches have notable drawbacks, including their reliance on favorable weather conditions; rain, fog, or snow can compromise image quality. Furthermore, vehicle vibrations and suspension movement can distort images, requiring reduced driving speeds—sometimes as low as 10–15 km/h—to maintain data accuracy [27]. One approach is to utilize high-resolution cameras with sufficient frame rates and image stabilization to ensure accurate measurements at higher speeds. However, this increases hardware costs and highlights the trade-off between image quality and the substantial data storage requirements.

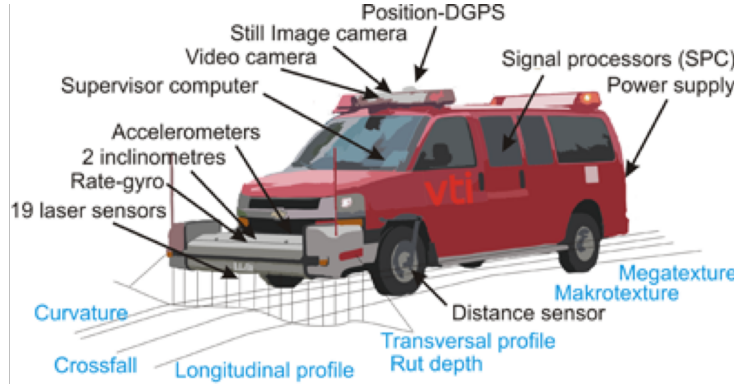


Figure 1: Conventional multi-sensor road-quality monitoring vehicle setup [28]

In conclusion, the drawbacks of these method are the high initial cost and equipment requirements, and the need for a trained operator to manage the system. These drawbacks make the use of these complex systems less widespread, mainly along major routes [6, 15]. Furthermore, in most cases, a complete and high accuracy 3-dimensional mapping of the road surface would not be necessary, as a preliminary general picture of the road surface would allow to screen the road sections that need to be upgraded in the near future. For these smaller sections, full mapping would be more justified at a later stage. This monitoring approach could bring significant cost savings and facilitate coordination and

planning.

### **Inertial sensor based monitoring**

The continuous reduction in computing costs, coupled with advancements in data processing techniques, has facilitated the development of alternative, cost-effective road condition monitoring methods. These approaches leverage vehicle-mounted inertial sensors and have been the subject of numerous recent studies [15, 16, 29]. The fundamental principle underlying these systems is that road surface irregularities induce vibrations that propagate through the vehicle structure, causing displacements along three axes. By capturing these vibrations, the condition of the road can be inferred.

The core component of such systems is the Inertial Measurement Unit (IMU), a compact suite of sensors capable of measuring linear accelerations, angular velocities and strength of magnetic field along multiple axes. These sensors are typically mounted on the vehicle chassis or suspension system to record the dynamic response of the vehicle to road roughness. The collected signals contain valuable information regarding road anomalies, including potholes, cracks, and uneven surfaces. However, interpreting these signals accurately requires sophisticated data analysis techniques, as raw sensor outputs are influenced by various external factors such as vehicle speed, suspension type, and tire characteristics.

To enhance the reliability of inertial sensor-based monitoring, advanced machine learning algorithms are employed. Supervised machine learning models play an important role in aggregating multiple variables to distinguish between different types of road defects and assess overall pavement quality.

Compared to traditional profilometers, inertial sensor-based monitoring offers several advantages. These systems are significantly lower in cost, can be easily deployed on standard vehicles, and enable large-scale, real-time data collection and evaluation. Furthermore, their scalability allows integration into crowdsourced monitoring frameworks, where data from multiple vehicles contribute to a continuously updated road condition database, as this method is not limited to decent weather and daytime light conditions [18].

However, there are some possible limitations, including sensitivity to variations in vehicle dynamics and to different possible sensor placements. Furthermore, another main challenge for these vehicle vibration-based systems is to train the machine learning models, as reference road quality labels are usually missing or not available at all for public road

networks [6, 18, 30, 31]. Even the most recent papers have to rely on supervised training and manual labeling [1, 6, 17, 31, 32], however, this approach does not provide an adequate solution in most cases, due to its extremely high time requirement and therefore cost, the large number of data points and its general inaccuracy, as consistent manual labeling is difficult to maintain, with a high potential for subjectivity. Manual annotation is also not flexible and very difficult to modify later, for example when a new road quality class is required.

#### **2.1.4 Metrics of pavement quality**

Road roughness is the elevation profile along the travelled wheelpath and is defined as "the deviations of a pavement surface from a true planar surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamic loads, and pavement drainage" [33]. Roughness has been recognised as one of the main criteria used to judge the surface quality condition of a road network. The assessment of road surface defects varies across countries but generally includes common issues such as cracks, rutting, shoving, distortion, potholes, raveling, and bleeding. These defects are evaluated through various methods, including on-site manual inspections, photographic analysis, and data collected from measurement vehicles. Additionally, many transportation agencies employ defect severity indices, which quantify the extent and impact of each defect. Based on these assessments, an overall road condition index is computed, integrating defect severity and density. However, the weighting and calculation methodologies differ across transportation departments, reflecting regional standards and priorities.

Several different parameters can express the road surface irregularity, among these the most used is the International Roughness Index (IRI). IRI expresses the longitudinal road profile as a slope statistic in meters per kilometer. Essentially the index is computed independently for each wheel track, and the reference speed is 80 km/h. IRI was demonstrated to summarize the roughness qualities that affect overall vehicle operating cost, ride quality, and dynamic wheel loads [34, 35]. The IRI is typically measured using road profilometers and is categorized into different International Quality Levels (IQL), usually ranging between three distinct categories (IQL-1 up to IQL-3) [18].

However, there are some disadvantages of IRI. Even the highest-precision IQL-1 systems provide roughness measurements at intervals of 10 to 20 meters, while lower-precision IQL-3 systems report at intervals exceeding 100 meters. This means that certain obstacles, such

as separated or minor potholes or irregularities, are not detected [6]. Furthermore, it was pointed out that the IRI is only slightly sensitive to human vibration perception since the IRI does not reflect wavelength contents of road elevations but rather profile slopes [36]. Furthermore, IRI has decreased validity at different speeds since the vehicle body has various bounces according to suspension characteristics and damping ratios. Ahlin et al. and Praznowski et al. [2, 37] say that the vertical and longitudinal position of the wheel changes during vibration. As a result, at constant vehicle speed, the forces due to the irregularity of the road will accelerate the body of the vehicle both longitudinally and vertically.

An indicator that has a stronger correlation between road unevenness and vehicle or human body ride vibration response would be more appropriate. Mucka and Granlund [36] proposes the Power Spectral Density (PSD), as it reflects better to road irregularities. Similarly, Loizos and Plati [38] says that IRI generally provides less information concerning the characteristics of the measured road profile than the PSD [39–42].

## **2.2 Vehicle body motion characteristics**

The dynamic response of a vehicle is governed by a complex interaction between road surface irregularities, vehicle suspension systems, and external forces. Understanding vehicle body motion characteristics is essential for evaluating ride comfort, road-induced vibrations, and structural loads affecting both the vehicle and its occupants. These characteristics are influenced by multiple factors, including vehicle's vibration behavior, dynamic properties, and the underlying road profile.

Vehicle motion analysis typically relies on well-defined coordinate systems to describe translational and rotational movements accurately. The study of vehicle dynamics encompasses mathematical models that describe motion behavior under various driving conditions, integrating factors such as mass distribution, suspension properties, and external disturbances.

This subsection provides an overview of the key aspects of vehicle body motion, starting with the definition of coordinate systems, followed by an examination of vehicle vibrations, dynamic modeling approaches, and road surface representations.

### **2.2.1 Coordinate systems in vehicle dynamics**

The accurate analysis of vehicle motion requires the definition of multiple coordinate systems, each serving a specific role in kinematic and dynamic modeling. The primary co-

ordinate frames used in vehicle dynamics include the earth-fixed coordinate system, the road plane coordinate system, the vehicle body coordinate system, and the wheel coordinate system. These coordinate systems provide the necessary mathematical framework for describing the translational and rotational motion of the vehicle, ensuring consistency in motion equations and facilitating interaction between different vehicle components.

### 2.2.2 Coordinate Systems

In the study of vehicle motion, the use of well-defined coordinate systems is essential for accurately describing translational and rotational movements. Several coordinate frames are commonly used in vehicle dynamics, including the earth-fixed coordinate system, the road plane coordinate system, the vehicle body coordinate system, and individual wheel coordinate systems. These frames serve distinct purposes in motion analysis, ranging from global positioning to localized movement descriptions.

The earth-fixed coordinate system  $(X_E, Y_E, Z_E)$  is an inertial reference frame in which  $X_E$  and  $Y_E$  span a horizontal plane perpendicular to the direction of acceleration of gravity  $g$ , while  $Z_E$  is aligned with the vertical axis. This system is commonly used in absolute positioning with map-matching and geocoding algorithms, and large-scale motion analysis. The road plane coordinate system  $(X_R, Y_R)$  represents the road surface in the vicinity of the vehicle. Since road surfaces are often uneven, each tire contact area may correspond to a different local road plane with its own surface profile. This coordinate frame is particularly useful for modeling road-induced vibrations and surface irregularities.

In order to take into account the dynamic couplings, vertical load transfer and influence of suspensions, the vehicle can be broken down into two supposedly undeformable masses: the sprung mass and the unsprung mass. The sprung mass includes the vehicle body, engine, passengers and so on., while the other connected parts, including the wheels, suspensions, brakes and so on., are the unsprung mass, as Fig. 2 shows [43].

In addition, to take into account the differences between the influence of the front axle and the rear axle, the unsprung mass can be also decomposed into two supposedly undeformable masses. We then have  $\sum = S_s + S_{uf} + S_{ur}$ , where  $\sum$  is the overall vehicle of a mass  $m$  and a Center of Gravity (CoG)  $G$ ,  $S_s$  is the sprung mass of a mass  $m_s$  and a CoG  $G_s$ ,  $S_{uf}$  the front unsprung mass of a mass  $m_{uf}$  and a CoG  $G_{uf}$ , and  $S_{ur}$  is the rear unsprung mass of a mass  $M_{ur}$  and a CoG  $G_{ur}$ , with an overall CoG  $G_u$  and mass  $m_u$  for the unsprung mass [43].

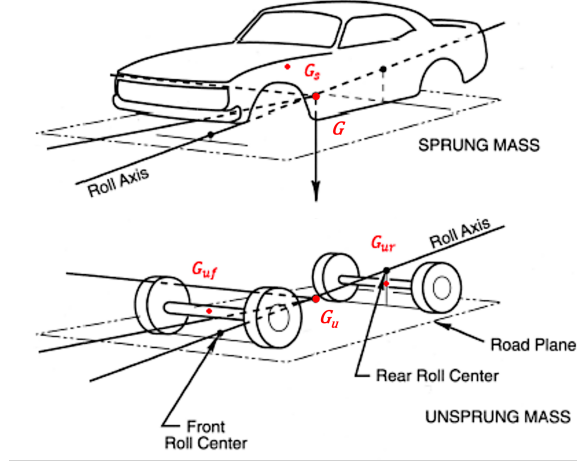


Figure 2: The sprung and unsprung masses decomposition [43]

The vehicle coordinate system  $(X_V, Y_V, Z_V)$  is fixed to the vehicle's body, with its origin typically placed at a designated vehicle reference point. The  $X_V$ -axis points forward along the longitudinal centerline of the vehicle, the  $Y_V$ -axis extends laterally to the left, and the  $Z_V$ -axis is directed upward. The reference point is often chosen as the center of gravity of the sprung mass ( $G_S$ ), as this simplifies dynamic modeling and force analysis. The vehicle motion is characterized by linear accelerations  $(a_x, a_y, a_z)$  along the three axes, as well as angular velocities  $(\omega_x, \omega_y, \omega_z)$  about them. Rotational movements are further described using Euler angles: the yaw angle  $\psi_V$  (rotation around  $Z_V$ ), the pitch angle  $\theta_V$  (rotation around  $Y_V$ ), and the roll angle  $\phi_V$  (rotation around  $X_V$ ). It is important to note that the pitch and roll angles are measured relative to the earth-fixed coordinate system rather than the road surface, meaning a stationary vehicle on an inclined plane will exhibit a nonzero pitch angle.

Each wheel is assigned its own wheel coordinate system  $(X_W, Y_W, Z_W)$ , centered at the wheel's center of gravity. The longitudinal, lateral, and vertical accelerations within this frame are denoted as  $(a_{W_x}, a_{W_y}, a_{W_z})$ , while wheel rotation, position, and camber angles are represented by  $\theta_W$ ,  $\psi_W$ , and  $\gamma_W$ , respectively.

The above described notations and coordinate system are shown in Fig. 3. By establishing these coordinate systems, the kinematic and dynamic behavior of the vehicle can be rigorously analyzed, enabling accurate modeling of vehicle-road interactions, sensor-based motion estimation, and control system development.

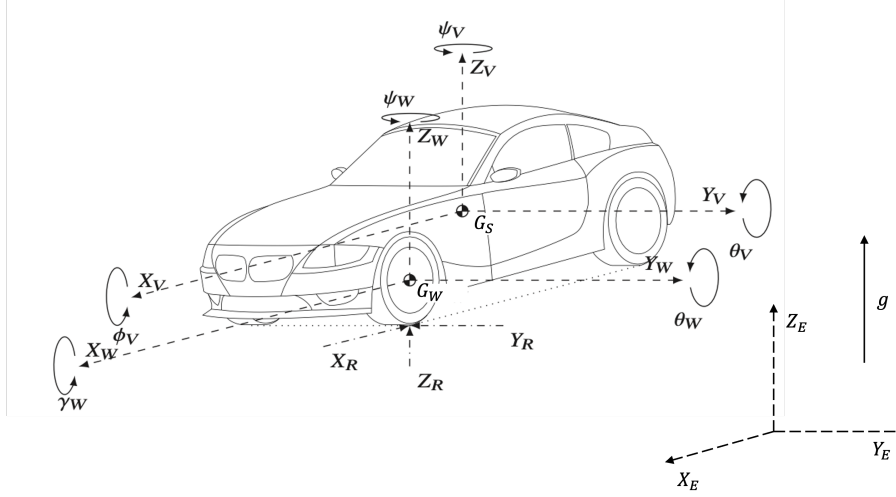


Figure 3: Vehicle with three translational and three rotational degrees of freedom [18]

### 2.2.3 Vehicle dynamics and models

A vehicle's dynamic behavior depends on multiple degrees of freedoms (DOFs), including body accelerations, roll and pitch motions, and dynamic wheel load fluctuations. Suspension systems aim to minimize these effects, ensuring both ride comfort and handling stability. The main factors influencing the vibration behavior of a vehicle are the mass, stiffness and damping distribution in the structure and the vibration excitation of the structure [44]. The primary objectives of a well-designed vertical dynamic system include minimizing vehicle body accelerations, reducing roll and pitch motions, and limiting fluctuations in dynamic wheel loads, while maintaining consistent vibration characteristics independent of external loading conditions. Vertical forces in the system primarily arise from the road surface irregularities, which then propagate through the vehicle body through the suspension springs and dampers, which regulate its movement in response to road irregularities [18].

To study the dynamic properties of vehicles in more detail, various models have been developed to investigate the ride quality of ground vehicles. The most common is the 2 degrees-of-freedom quarter-car model, where the effects of sprung and unsprung masses, spring stiffness, and shock absorbers are investigated, as shown in Fig. 4. Thereby, the mass stores kinetic energy and the spring potential energy, while the damper serves as the dissipative element. The suspension usually includes a number of additional components, however, in vibration models often only springs and dampers are considered. The suspension is characterized by its spring stiffness  $k_s$  and damping coefficient  $c_{sh}$ , while the tire is modeled with an equivalent stiffness  $k_{tr}$  and damping coefficient  $c_t$ . Road irregularities,

represented by the surface profile elevation  $z_0$ , directly influence the system's response. In vehicle dynamics, three different orientations are defined: longitudinal, lateral, and vertical. Longitudinal dynamics refer to the behavior of the vehicle in a straight line, encompassing aspects such as acceleration, braking, and straight-line stability. Lateral dynamics focus on the vehicle's behavior during turning, while vertical dynamics relate to movements caused by factors such as speed bumps, surface roughness, or pavement defects [14].

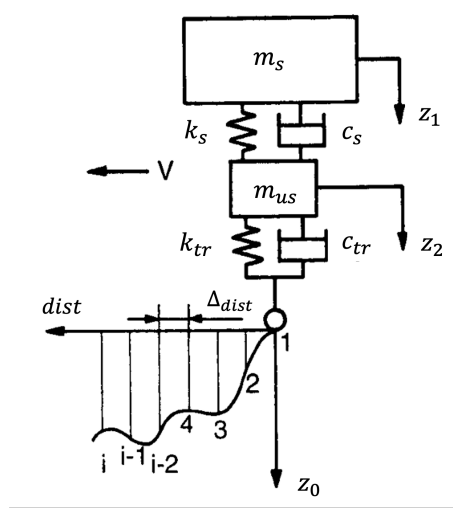


Figure 4: Quarter-car model with the variation of surface profile [14]

By applying Newton's second law to the sprung and unsprung mass separately, the dynamic response of a two-degree-of-freedom vehicle suspension system under surface excitation can be obtained. For vibrations excited by surface undulation, the equations of motion are as follows, first, for the sprung mass:

$$m_s \ddot{z}_1 + c_{sh}(\dot{z}_1 - \dot{z}_2) + k_s(z_1 - z_2) = 0 \quad (1)$$

and for the unsprung mass:

$$m_{us} \ddot{z}_2 + c_{sh}(\dot{z}_2 - \dot{z}_1) + k_s(z_2 - z_1) + c_t \dot{z}_2 + k_{tr} z_2 = F(t) = c_t \dot{z}_0 + k_{tr} z_0 \quad (2)$$

where  $F(t)$  is the excitation acting on the wheels and induced by surface irregularities. If  $z_0$  is the elevation of the surface profile and  $\dot{z}_0$  represents the vertical velocity of the tire at the ground contact point, which is the slope of the profile multiplied by the forward speed of the vehicle, then the excitation due to surface undulation may be expressed by  $c_t \dot{z}_0 + k_{tr} z_0$ , as shown in Eq. 2. If the excitation of the system is known, then, in principle,



the resulting vibrations of the sprung and unsprung mass can be determined by solving Eq. 1 and Eq. 2 [14].

According to this, the vehicle vibrations mainly depend on the road roughness, which can be used to infer the quality of the road surface [14, 37, 45, 46]. Vibrations due to road defects primarily result in vertical displacements, however, longitudinal vibrations also occur at the wheels. As a result, road surface irregularities can induce displacements along all three axes, therefore, all three axes must be considered simultaneously in the road quality assessment for a good measurement accuracy.

To determine the natural frequencies of the distinct masses, the two-degree-of-freedom system also can be used. The natural frequencies that occur on the unsprung mass is usually an order of magnitude higher than that of the sprung mass. Additionally, for passenger cars, the damping ratio provided by shock absorbers generally falls within the range of ratio 0.2 to 0.4, while the damping effect of the tire is relatively insignificant. These differences and the distinct separation between the natural frequencies of the sprung and unsprung masses have significant implications for the vibration isolation characteristics of the suspension system. For instance, when the wheel encounters a bump, the impulse primarily excites the unsprung mass, setting it into oscillation. After passing over the bump, the unsprung mass continues in free oscillation at its natural frequency. In contrast, the sprung mass experiences excitation from the vibrations of the unsprung mass rather than directly from the road surface. Since the natural frequency of the unsprung mass is an order of magnitude greater than the sprung mass's, the resulting oscillations of the sprung mass remain relatively small in amplitude.

#### **2.2.4 Vehicle vibration**

Vehicle vibrations arise primarily from road surface irregularities, which act as external excitations to the vehicle's oscillation system. Since these irregularities vary in amplitude and wavelength and occur at non-uniform intervals, the road surface can be characterized as a stochastic excitation source. In addition to road-induced vibrations, internal forces generated by the powertrain, drivetrain, and rotating wheels contribute to the overall vibrational behavior of the vehicle. Furthermore, dynamic pitch and roll movements during longitudinal and lateral maneuvers introduce additional excitations, affecting ride comfort and handling stability.

Tires, being the sole contact point between the vehicle and the road, play a fundamental

role in transmitting forces and compensating for surface irregularities. The interaction between the tire and the road surface is governed by two primary physical effects: adhesive friction and hysteresis friction. Adhesive friction arises from molecular interactions at the contact patch, while hysteresis friction results from the viscoelastic properties of the tire material, leading to energy dissipation as the tire deforms over surface asperities. These mechanisms influence both traction performance and vibrational characteristics.

Excitation from the vehicle when it passes over an undulating surface consists of a broad spectrum of frequencies. High-frequency inputs can be effectively isolated due to the low natural frequency of the sprung mass. However, low-frequency excitations may be transmitted to the vehicle body with minimal attenuation or even amplification, as the transmissibility ratio increases when the excitation frequency approaches the natural frequency of the sprung mass.

One of the most critical parameters in assessing the overall performance of a suspension system is the vibration isolation between the sprung and unsprung masses. This can be evaluated by analyzing the response of the sprung mass (output) to excitation from the ground (input). The transmissibility ratio or transfer function serves as a fundamental measure for characterizing the vibration isolation performance of a linear suspension system.

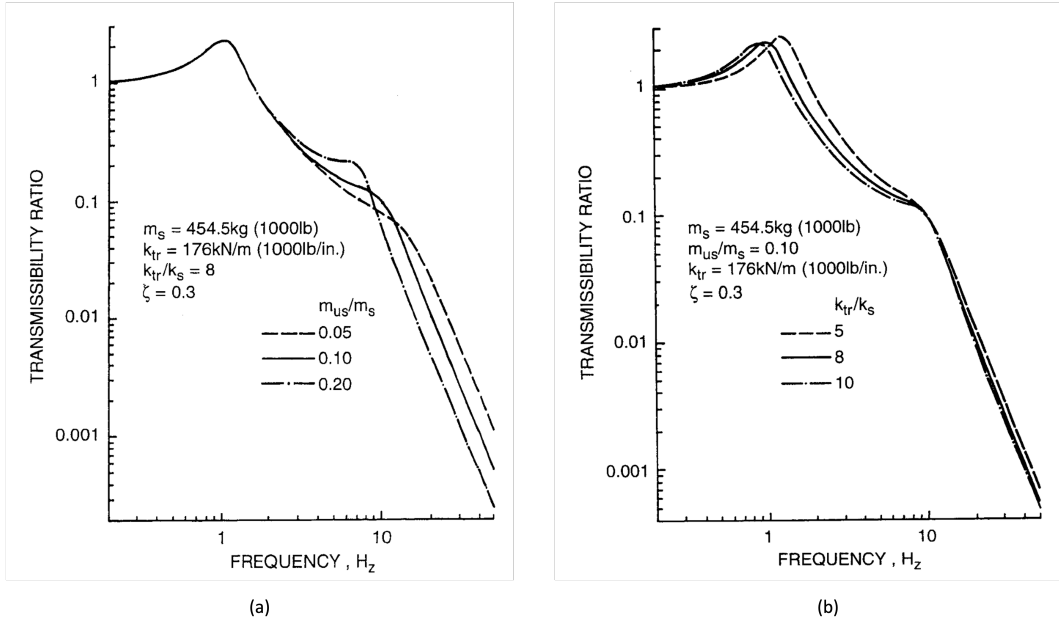


Figure 5: Transmissibility ratio as a function of frequency for the sprung mass of a quarter-car model with different ratios of unsprung to sprung mass, for a general passenger vehicle [14]

Fig. 5(a) illustrates the influence of the ratio of the unsprung mass to the sprung mass  $m_{us}/m_s$  on the transmissibility ratio in a two-degree-of-freedom system. In the frequency range below the natural frequency of the sprung mass, the unsprung mass has minimal effect on the vibration response of the sprung mass. However, when the excitation frequency approaches the natural frequency of the unsprung mass, a lower unsprung mass results in a lower transmissibility ratio. This implies that, given the same level of excitation, the sprung mass experiences reduced vibration when the unsprung mass is lighter. Conversely, in frequency ranges above the natural frequency of the unsprung mass, a lighter unsprung mass leads to a slightly higher transmissibility ratio. The breakpoints observed around 10 Hz can be attributed to the natural frequency of the vehicle suspension system. Below this frequency, the sprung and unsprung masses move predominantly together, whereas above it the suspension increasingly isolates the vehicle body from wheel-induced vibrations, leading to a characteristic change in the transmissibility ratio.

In addition, Fig. 5(b) illustrates the influence of the ratio between the equivalent tire stiffness  $k_{tr}$  and the suspension spring stiffness  $k_s$  on the system's transmissibility ratio. For a given tire stiffness, an increased ratio of  $k_{tr}/k_s$  corresponds to a reduction in suspension spring stiffness. In the frequency domain below the natural frequency of the sprung mass, a lower ratio of  $k_{tr}/k_s$  results in a decreased transmissibility ratio, indicating improved damping of vibrations. Within the frequency range bounded, a reduction in suspension stiffness enhances vibration isolation effectiveness. Conversely, in the frequency domain beyond the natural frequency of the unsprung mass, the suspension spring stiffness exerts a negligible influence on the vibratory response of the sprung mass, making the transmissibility ratio largely insensitive to variations in  $k_{tr}/k_s$ . Based on these, employing a softer suspension spring enhances vibration isolation within the mid- to high-frequency range. However, this benefit is accompanied by a potential trade-off, as it may lead to increased transmissibility in the low-frequency region, particularly below the natural frequency of the sprung mass [14].

The propagation of vibrations in a vehicle can be investigated and assessed in a number of other ways. Among the most relevant parameters are the relations between the up and down linear motion (bounce) and the angular motion (pitch) of the vehicle body.

Due to the significant separation between the natural frequencies of the sprung and unsprung masses, the bounce and pitch motions of the vehicle body and the motion of the wheels can often be considered approximately independent. However, in practical scenar-

ios, these motions are inherently coupled, meaning that an impulse exerted on the front or rear wheels will excite both bounce and pitch oscillations simultaneously. To further describe these vibration characteristics, the concept of the oscillation center can be introduced. For the bounce and pitch modes, the oscillation centers are located along the vehicle's longitudinal axis, ahead of and behind the vehicle's center of gravity, depending on the distribution of sprung masses and the stiffness of the suspension springs. In general, a road input at the front or rear wheel will cause a moment about each oscillation center and therefore excite both bounce and pitch oscillations. In other words, the body motion will be the sum of the oscillations about the two centers [14, 47].

Beyond the above interactions, environmental factors can also influence vehicle vibrations. Variations in ambient temperature affect the viscoelasticity of rubber, altering tire stiffness and friction properties. Snow, water, and other road contaminants modify the tire-road contact dynamics, introducing additional vibrational effects. Moreover, material aging and chemical degradation over time alter the mechanical properties of vehicle components, influencing their response to vibrational inputs [18].

The consequences have particular relevance for sensor placement in vibration-based road quality measurement systems, where three primary options exist: measurement on the sprung mass, or on the unsprung mass, or on both. However, given that the transmissibility ratio strongly depends on the excitation frequency, vehicle parameters, and suspension characteristics, measurements taken from the sprung mass may be significantly influenced by the complex coupled dynamics of the system. This can introduce distortions, leading to uncertainty and reduced accuracy in road surface assessments. Therefore, careful consideration must be given to sensor placement to ensure reliable and consistent measurements in real-world applications. A well-optimized sensor configuration should minimize the impact of vehicle dynamics on the recorded data, thereby enhancing the fidelity of vibration-based road quality evaluations.

### **2.3 Review of previous vibration-based road monitoring systems**

Many works [1, 15, 16, 21, 48–50] related to vehicle vibration-based road quality measurement in the literature perform the measurement and data processing according to different methodologies. Road condition monitoring works usually focus on detecting a single road feature, like a pothole, rather than a general assessment and classification of a section. Most of the work has focused on terrain classification to identify the type of pavement,

which can differentiate, for example, gravel or asphalt pavement. Fewer pothole detection and road profile reconstruction studies have been carried out. For data acquisition, inertial sensors or z-directional accelerometers are most often used, while the data mining methods used for data processing have varied considerably [51, 52].

In this subsection I provide a summary of previous studies on vehicle vibration-based road quality measurement. The measurement methods and sensors used are described, followed by the common methods and workflows used for data processing. I will then discuss evaluation methods, often based on machine learning. Finally, I will summarise the factors that have been investigated in the studies that affect the accuracy of measurement and the correction methods applied.

### 2.3.1 Data acquisition

The focus of the objects to be detected in vibration-based road monitoring studies was often some predefined road feature, such as a pothole of a specific size, a raised bump, or a railway crossing, rather than a general assessment and classification of a section. In addition, in many cases, the measurements were performed under limited conditions, such as speed limits or in repetitive sections. Most of the work has focused on terrain classification to identify the type of pavement, which can differentiate, for example, gravel or asphalt pavement [17]. Fewer road profile reconstruction studies have been carried out, of which some of studies have performed a more comprehensive classification, where the objective was to classify the entire measured route into road quality classes after dividing it into minor sections.

Studies [1, 15, 16, 21, 48–50] have explored pavement roughness estimation systems based on vehicle chassis vibrations. These previous studies typically utilized a minimal number of sensors, often relying on an IMU either as a standalone device or integrated into a smartphone. These sensors were commonly positioned near the vehicle’s center of gravity or on the dashboard. The advantage of the standalone IMU sensor is higher measurement accuracy and more stable operation, as well as the possibility of stable axial fixation, which eliminates the need for additional axial correction.

The method of sensor placement divides the studies into two categories, with some [45] using vibration data on the unsprung mass or suspension  $m_{us}$  to gather data about road quality and more others [1, 6, 32, 37, 53–55] using sensors only on the sprung mass or vehicle body  $m_s$ . The latter can have the advantage that even a smartphone is sufficient

for data collection, so large amounts of data can be collected easily and cost-effectively. These results were generally good, but the tests on real-world data showed some loss of precision compared to laboratory data [6, 18].

Furthermore, the measurement in the cabin may have the disadvantage that the vibrations are significantly isolated from the suspension. Although the excited vibrations are transmitted to the bodywork, the damping elements absorb considerable energy; furthermore, the frequency of the vibrations changes, resulting in longer decay time oscillations in the cabin. The natural frequency of the unsprung mass is an order of magnitude higher than that of the sprung mass. At different frequencies, the system functions with varying damping rates, as detailed in Subsection 2.2.4. Based on the knowledge above, high frequency vibrations can be effectively filtered out with the suspension's damping elements since the natural frequency of the springs is in the lower region [14, 51]. According to health issues and passenger ride comfort, the acceptable amplitude of vibration is a function of frequency. As the frequency increases, the allowable amplitude decreases [14]. For this reason, the detection of higher frequencies is substantial, but as mentioned above, results may be distorted in the case of measurements exclusively in the cabin. Furthermore, previous studies [17, 32, 53] have often used a standard mobile phone to measure vibrations, which can cause additional measurement errors due to improper placement and uncertainties arising from the built-in sensor. Although a more complex measuring system is required for measurements on the suspension, the accuracy of the results presumably can be increased [56].

### **2.3.2 Data mining**

One of the main aspects of my work is to estimate and classify the quality of homogeneous and coherent road sections based on surface quality from raw vibration data. However, for these final evaluation algorithms to be sufficiently accurate, input features of high quality and high information content are essential. These points lead us to apply data mining methods, which support the processing of large data sets, the filtering of data errors and the extraction of useful and relevant information.

Data mining is a fundamental process in data-driven research and decision-making, and refers to a vast set of tools for understanding data [42]. These tools can be classified as preprocessing, feature extracting and supervised or unsupervised statistical learning models. A critical and the most time-consuming aspect of data mining are data modeling,

feature extraction and feature engineering, which aim to transform raw data into a more informative and meaningful representation, often referred to as information. This process involves mapping raw data to business-specific entities, extracting relevant features, and creating standardized workflows to ensure consistency and reproducibility. Since raw data is often noisy, incomplete, or unstructured, intermediate transformations are required before inputting the data into predictive models. These transformations in the programming level are implemented within a standardized data pipeline.

The actual data mining approach used in previous studies applied a few methods, however, for the later classification or regression tasks, a number of different solutions can be found. The data mining methods used in the previous studies are described below.

**Review of cleaning methods** Nguyen et al. [17] provided an overview of response-based road surface irregularity measurements, which generally categorized studies as pothole detection, road profile reconstruction or roughness estimation. The former are aimed at detecting only a single pothole or road defect, while the latter aim at providing a comprehensive statement of the level of roughness of the road surface for necessary maintenance.

These target variables basically determine the data cleaning and subsequent feature extraction steps. Previous works, like Chen et al. [26] have dealt with the replacement of null or missing values, while some have dealt with the filtering of stationary phases and sensor noise. Chao et al. [53] and his team were also the first who used noise filtering to clean the data from the IMU sensor, followed by an orientation correction, as the sensor was mounted on the dashboard rather than positioned axially.

Nuno et al. [1] not only recorded vibration data but also included the vehicle's velocity values, which enabled the exclusion of data measured while the vehicle was stationary. Douangphachanh and Oneyama [21] applied a high-pass filter to remove the gravitational component from all accelerometer axes, complemented by a low-pass filter to eliminate high-frequency noise.

Mikko et al. [54] provided a detailed overview of the data cleaning procedures applied. First, they implemented outlier detection on GPS coordinates, followed by a Kalman filter-based method for trajectory smoothing. Several variables were analyzed as a function of velocity; for this purpose, vehicle speed was estimated from coordinate values, corrected for observed latency and enhanced using fusion with acceleration data. A zero-speed correction by thresholding was then applied, where the identified stationary segments were

overwritten with zero velocity values.

**Review of feature extraction methods** Subsequently, frequency domain transformations and spectral analysis have been widely applied in many studies to extract informative features from time series signals. Mucka and Granlund [36] and Loizos and Plati [38] demonstrated that Power Spectral Density (PSD) shows a strong correlation between road surface irregularities and the vibration response of either the vehicle or human body. In addition to PSD, several studies utilized the Fast Fourier Transform (FFT) to examine frequency components [37, 53, 54].

Following frequency-domain analysis, segmentation or basic windowing techniques were frequently used to partition the dataset. Mikko et al. [54] used sliding windows of 2-second length with a 0.5-second overlap. Subsequently, several statistical features were extracted from the acceleration signals, including standard deviation, mean, variance, peak-to-peak amplitude, signal magnitude area, and others, calculated for each window. Chao et al. [53] aimed to define segments of approximately 10 meters in physical length, which ultimately translated to time windows of 1.3 seconds with a 50% overlap. This was followed by a series of threshold-based filters to eliminate non-potholed windows. One of the core components of this approach was the use of the root mean square value of the window as a preliminary metric to determine whether a segment contained anomalous data.

Nitsche et al. [57] utilized 35 features as model inputs, derived from the longitudinal, lateral, and vertical acceleration, along with the angular velocity of each wheel. Among the selected features were the mean, standard deviation, range, power spectral densities of the accelerations, and angular speed of both rear wheels. Lastly, Nguyen et al. [17] used a Principal Component Analysis for dimension reduction with good results.

**Number of distinct classes to assessed** The number of quality categories to be separated varies. Customer needs and current requirements are the decisive factors, however, these may change continuously. Road management companies typically define road quality classes of between 3 and 7 categories for road surface evaluation [14, 17, 42]. In addition, the useful information content provided by the identified clusters should be monitored to see to what extent they adequately isolate and express the information that is important to the road management company. In some cases, for example, too many clusters may be difficult to interpret if only a general picture of the road network is desired. A general four-class separation can be performed based on the following viewpoints:



- Class 1: The road section has continuous large-scale and deep defects. The road section contains sinkholes, subsidence, surface and deep potholes, and pavement edge cracks.
- Class 2: Minor defects persist along most road sections, with some more profound defects. Thus, it can contain continuous cracks, unevenness, subsidence, surface potholes, and possibly manhole cover level differences within residential areas.
- Class 3: The pavement is free of significant defects, but there are traces of previous repairs, cracks, and minor unevenness.
- Class 4: The road section is almost entirely new, with a defect-free surface.

For a classification with four distinct categories, the images of the road quality categories to be separated are shown in Fig. 6.

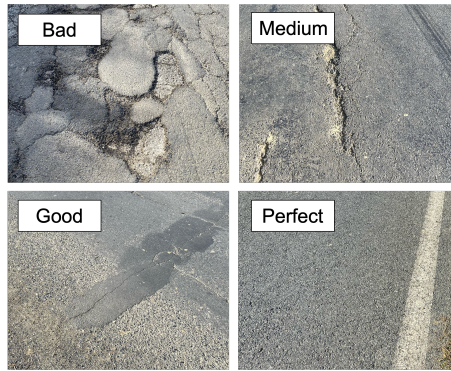


Figure 6: Photos of reference pavement classes, for a classification with four distinct categories

**Review of annotation methods for training data** One of the most important factor in ensuring the accuracy of any machine learning algorithm is the availability of a sufficiently large and correctly labeled training dataset. Without an adequate volume of training samples, the algorithm’s performance is significantly constrained.

While some studies do not specify their labeling methodology for real-world experiments, and others rely entirely on computer simulations where data can be automatically annotated, various approaches have been employed for labeling real-world data with differing levels of success. The most commonly used method for supervised model training is manual annotation, often combined with a threshold-based criteria system. However, this approach limits the spatial resolution of both measurement and evaluation, typically restricting classification to segment lengths between 10 and 100 meters. As a result, certain obstacles,

such as potholes, are not detected - the same problem as with IRI [6]

Tai et al. [58] employed a voice recorder to annotate training data during data collection with motorcycles in Taiwan. Their approach aimed to minimize the need for rider interaction. However, it resulted in significant inaccuracies, necessitating the development of a complex relabeling process to ensure the data's usability for their machine learning algorithm.

Jakob et al. [16] adopted a different strategy by using an in-vehicle computer along with a trained annotator seated in the passenger seat. The annotator manually labeled road defects in real-time by pressing keyboard keys corresponding to predefined defect categories. Additionally, they incorporated a loosely labeled dataset, where the general types and approximate frequencies of defects were known, but their exact count and locations remained unspecified.

Seraj et al. [59] implemented a labeling method using a smartphone's microphone and camera. A single annotator, seated in the passenger seat, verbally identified road defects as the vehicle approached them. A software tool transcribed the voice annotations with millisecond-level accuracy. To enhance robustness, they conducted multiple drives over the same road segment at different speeds and angles of approach. However, challenges emerged in synchronizing the transcriptions with the smartphone's recorded data, leading to misalignments between labeled segments and actual road defects. As a result, nearly all annotations required manual correction.

Mikko et al. [54] recorded 25 km of road data in a single drive, capturing video alongside accelerometer and gyroscope measurements. They preprocessed GPS and acceleration signals and visualized the data by plotting a spectrogram of the summed band energies of accelerometer signals, overlaying speed information on the same graph. Labeling was conducted by two independent annotators who manually analyzed the recorded video footage to generate annotation files. While this method provided a detailed dataset, the authors acknowledged that labeling based on video analysis was both labor-intensive and prone to errors.

In the Nuno et al. [1] and Masino et al. [6] studies, labeling was obtained by physical push buttons connected to the measuring unit. If the driver passed a road defect or considered that the overall quality of the road section has changed, the physical button is pressed to label the road section in real time according to predefined categories.

In the studies mentioned above, several solutions were introduced to make manual labelling more efficient and accurate, thus the annotation could be performed in real time during the measurement, typically by an assistant person, however, labeling is still complicated and unstable, segment boundaries can be inaccurate and subjective judgement is to be expected in the final result. Furthermore, it is common deficiency in previous studies that labeling is difficult or not reversible after it has been completed during the measurement.

### **Review of machine learning models used for pavement classification**

After labeling, supervised machine learning algorithms have typically been used in previous work for final evaluation of the road surface. These subsequent statistical models contributes in discovering complex relationships within multi-variable datasets, improving predictive accuracy, and automating analytical processes. The statistical learning involves building a statistical model for clustering or estimating an output based on one or more inputs [60].

Supervised classification models are often used for the final evaluation. The range of applied models was broad, from the simple Support Vector Machine (SVM) to the Complex Neural Networks (CNN) [17].

Mikko et al. [54] employed support vector machines with a radial basis function kernel to classify road segments based on acceleration and GPS data. To address dataset imbalance, class weighting was applied, and the best model was selected using LIBSVM with 49 parameter configurations. The approach effectively distinguished severe anomalies from normal road surfaces, and speed dependence removal improved classification accuracy. A visualization framework provided intuitive assessment of predictions. However, the method struggled to differentiate minor and major defects. Furthermore, labeling training data from video was time-consuming and error-prone, thus, the authors suggested using unsupervised learning or clustering techniques to enhance performance.

Praznowski et al. [37] applied Bayesian classifiers to categorize road conditions based on vibration data, leveraging statistical and spectral analysis techniques such as the short-time Fourier transform and autocorrelation functions. The research emphasized that while this approach provided reliable insights into road roughness, it faced challenges related to sensor placement variability and data noise, requiring additional preprocessing steps.

Chao et al. [53] evaluated multiple classifiers, including logistic regression, Support Vector

Machines, and Random Forests (RF). The results indicated that RF provided the best classification performance, achieving 88.5% precision and 75% recall. The study highlighted that machine-learning classifiers outperformed traditional threshold-based methods but required extensive preprocessing, including filtering and reorientation of accelerometer data. While the approach demonstrated robustness across different road types, limitations included the reliance on manually labeled training data and potential inaccuracies due to sensor noise.

Raslan et al. [61] introduced a deep learning framework for road surface classification using data from Egyptian roads. The proposed model combines Bidirectional Long Short-Term Memory (BiLSTM) and Depthwise Separable Convolutions (DSC) to process both time and frequency domain features. Multiple models, including Random Forest, CNN, and Transformers, were evaluated, with the DSC-BiLSTM achieving the highest accuracy of 93%. The approach effectively reduced reliance on hand-crafted feature extraction, but challenges included computational complexity and the need for extensive labeled data. The study also proposed a road quality index estimator to assist authorities in road maintenance planning.

An unsupervised model-based approach that does not require labeling was proposed by [31]. Here, a K-Means clustering solution was used, however, the model was not used for labeling, but as a production tool for final path quality evaluation. The solution performed well, but drawbacks include the long-term instability and reliability of the unsupervised model in terms of classification results, the difficulty in interpretation of the model and the higher computational complexity compared to supervised models.

Overall, the reviewed studies indicate that the vehicle vibration-based methods provide reliable road profile estimations. However, their accuracy is sensitive to velocity variations. Feature extraction methods, such as spectral analysis and wavelet transforms, are commonly used to isolate road-induced vibrations from noise and vehicle-specific oscillations. Additionally, data fusion techniques combine information from multiple sensors to improve spatial localization and classification accuracy. Machine learning models played an important role in aggregating multiple variables to distinguish between different types of road defects and assess overall pavement quality. The average classification accuracy from the reviewed studies was on average between 85% and 93%. However, an important conclusion highlighted by many was the subjectivity and potential error factor of manual labeling, which should be improved in any industrial application.

### 2.3.3 Significance of additional factors and generalization

Due to the complex dynamics of vehicles, numerous additional factors must be considered in vibration-based road quality assessment. If the effects arising from these factors are not properly corrected, they can introduce significant errors into the measurements. In the following subsection, I examine the generalization methods applied in existing works, followed by a review on the factors identified in previous studies that have been observed to negatively impact measurement accuracy.

**Review of generalization methods for real-world evaluation** In real-world applications, the accuracy of measurements can be influenced by various external factors, making it essential to focus on generalization and real-world evaluation when developing new models. Generalization refers to a model's ability to perform well on unseen data, ensuring that it captures the underlying patterns rather than just memorizing the training data. The goal of generalization is to create robust models that maintain high performance across different conditions, minimizing the impact of noise, sensor variability, and environmental changes.

To evaluate generalization, experiments must incorporate a diverse range of vehicles, road segments, and an extensive road network to ensure comprehensive testing. However, many studies rely solely on computer-simulated data to assess algorithm accuracy, while others use a limited selection of vehicles on controlled test tracks, covering only a few kilometers of predefined road sections and focusing on specific outputs.

Seraj et al. [59] conducted tests using five different vehicle types with covering more than 45 kilometers public road. However, they did not specify whether measurement data from different vehicles were combined or if models were trained separately for each vehicle. Rajput et al. [31] placed vibration sensors on long-distance buses, allowing the measurements to be repeated several times over long distances, however, the paths to be monitored were very bounded and the interpretation of the results has to take into account the significantly different dynamic parameters of the buses. Jakob et al. [16] deployed seven identical taxis to monitor Boston's road infrastructure over ten days, while Menant et al. [62] utilized 12 similar probe vehicles to assess pavement quality across a 1000 km network in France. Additionally, Chen et al. [26] implemented a road surface monitoring system on 100 taxis in Shenzhen but did not disclose details about the vehicle types used. It should also be noted that the measurements were performed under limited conditions, such as

speed limits or in repetitive sections in many cases.

**Review of additional factors influencing measurement accuracy** One of the most significant factors is the distortion effect due to suspension, as outlined earlier. The measurement in the cabin may have the disadvantage that the vibrations are significantly isolated from the suspension. Although the excited vibrations are transmitted to the bodywork, the damping elements absorb considerable energy; furthermore, the frequency of the vibrations changes, resulting in longer decay time oscillations in the cabin. Overall, for cabin-only measurements, these should be given considerable attention. As a consequence, a suspension-based measurement approach can be beneficial, despite the fact that a more complex measuring system is required for measurements [56].

Ward and Iagnemma [45] have highlighted that vehicle speed significantly influences the amplitude of measured vibrations. A variation in vehicle speed can affect the frequency and amplitude of the measured wheel acceleration, this would introduce significant errors in the results without correction. Above 80 km/h, wheel accelerations may decrease. A possible approach is spatial resampling, which mitigates speed dependence by resampling the vibration data at constant spatial intervals instead of time intervals using linear interpolation. Instead of using acceleration signals directly in the time domain, the authors transformed them into the spatial domain by estimating the vehicle's displacement over time. This conversion allowed road profile features to be represented independently of speed variations. The process involved first calculating the cumulative travel distance from the GPS position data, then interpolating acceleration data at constant spatial intervals, ensuring uniform sampling regardless of speed fluctuations. By analyzing spatial frequency components rather than temporal signals, the classification algorithm became less sensitive to changes in vehicle speed, improving its robustness across different driving conditions.

In addition to the speed effect, Masino [18] have concluded a higher measurement accuracy on straight road sections. Although this effect has not been studied in depth before, it may be due to negative effects of vehicle maneuvers, as additional vibration signals from maneuvers causes signal distortions. Due to the high complexity of the forces acting on suspension elements, these components exhibit different characteristics during maneuvers compared to normal driving conditions. Under moderate driving or braking forces, the impact of these forces on suspension response is relatively minor and can often be neglected. However, during aggressive maneuvers, the primary vibration signals from road imperfections can become distorted. Consequently, the accuracy of vibration-based road

quality measurements may be reduced in such situations [14, 37].

While the driver should aim to maintain a consistent speed during measurements, sudden maneuvers may become unavoidable due to external traffic conditions. Additionally, since modern vibration-based systems do not rely on specialized vehicles, data collection can be performed using standard, unmarked vehicles. Consequently, external traffic dynamics play a more critical role, as it is not always possible to maintain a stable, self-determined speed or to perform smooth maneuvers.

The effects, detection, and isolation of these so-called aggressive maneuvers have not typically been addressed in previous studies, but could further improve the accuracy of vibration-based measurement systems. Based on these considerations, I will investigate the impact of vehicle maneuvers on the accuracy of road quality measurements, and in the next subsection I will present previous vehicle maneuver detection algorithms.

Wang [42] examined more other factors influencing vehicle accelerations and found that other vehicle-related factors typically have a medium (10-30%) or low (<10%) impact, according to their degree of influence. The former category includes the sprung mass and tire pressure, while the latter consists of suspension damping and the wheelbase effects [42]. These are therefore considered negligible in our study.

#### **2.3.4 Inertial sensor-based vehicle maneuver detection**

Previous studies [63, 64] focusing on vehicle maneuver detection typically use different types of sensors and have concentrated on detecting different types of maneuvers, often targeting the identification of parking, stopping, driving conditions, or lane changes. Song and Li [63] provides a comprehensive overview of commonly used maneuver prediction and detection techniques, however, vibration-based approaches for maneuver detection remains relatively underexplored.

Maneuvers can be classified into two main categories based on their characteristics. Normal vehicle maneuvers encompass routine driving actions that occur under typical conditions, such as lane changes or moderate-speed cornering. In contrast, abnormal maneuvers refer to sudden and atypical driving behaviors, including longitudinal maneuvers, such as hard braking, or lateral maneuvers, like sharp turns, which often result from emergency situations or unexpected traffic dynamics. Due to the complexity of vehicle dynamics, the longitudinal, lateral, and vertical axes must also be managed as a coupled system in

maneuver detection. It is not sufficient to detect a maneuver based on a single feature or to assess road pavement quality using only vertical axis data [14, 37, 65].

Several approaches have been proposed for defining aggressive maneuver classes. Some studies have introduced a more detailed classification system, such as Deo et al. [66], who defined six maneuver categories specifically for highway driving. To improve robustness, Bahram et al. [67] advocated for a reduced set of fundamental maneuver classes. Expanding the number of maneuver categories increases model complexity, requiring more advanced feature extraction and processing techniques. This, in turn, leads to higher computational costs and greater hardware demands, particularly in real-time applications. Therefore, a balance must be found between model simplicity and the number of classified maneuvers.

Villanueva et al. [68] utilized data from an integrated sensor unit to classify maneuvers into three broad categories: parking, driving, and stopping. Their approach relies on windowed data segmentation with a breakout detection threshold, using statistical and time-based features for classification. In contrast, Johnson and Trivedi [69] developed a method for detecting and categorizing aggressive driving maneuvers using data from a smartphone's accelerometer, gyroscope, and magnetometer. The start and end points of driving events were identified through a change point detection algorithm, and the extracted time series were compared against a maneuver database using Dynamic Time Warping (DTW). This method effectively measures the similarity between signals of varying lengths. However, a notable limitation of DTW is its high memory consumption, necessitating data recording in five-minute intervals. As a result, the software implementation is restricted by the significant power demands of the DTW algorithm, which surpass those of alternative classification methods.

Mokammel Haque et al. [70] introduced a rule based approach for maneuver detection based on IMU data from smartphones. In this study, the objective was to classify left and right lane changes in addition to separating aggressive longitudinal and lateral maneuvers. Compared to the previously presented methods, the major advantages of this solution are low computational power requirements and reduced runtime. On the other hand, the accuracy of the model was generally reduced, with an overall accuracy above 91% only for the lane change classes and an average of 85% for the others.

It is also important to point out that, although the so-called aggressive maneuvers have



been distinguished above, the level of threshold for these maneuvers is usually determined on the basis of safety considerations and not in terms of the likelihood of error signals in terms of road quality measurement.

### 3 Concept of the novel system

This section presents the objectives of my research, outlining the key requirements and providing a holistic overview of the goals and perspectives that guided the development of the proposed system. Following this, the foundations of the novel system are detailed, including its architecture, structural components, and data flow mechanisms. The implemented measurement methods are also introduced, highlighting their role in ensuring accuracy and reliability. These elements serve as the basis for the subsequent sections, where the system's design, implementation, and validation are discussed in greater depth.

#### 3.1 Aim of this work and research objectives

According to the literature review above, the widespread adoption and availability of traditional multi-sensor road quality measurement systems is hindered by their high cost of ownership, however, road management companies are emerging in demand for a more accurate and reliable system than the visual inspection method that is still widely used today. Based on the results of previous studies, road quality measurement based on vehicle vibrations may be a promising approach, although these previous implementations have had several limitations.

Compared to traditional approaches, an optimized sensor placement strategy, effective correction of potential error sources, and a more flexible training process adaptable to specific requirements can significantly enhance the cost-efficiency and applicability of vehicle-based vibration monitoring systems. By providing higher spatial resolution and improved adaptability, such systems could serve as a viable alternative to the current road monitoring methods.

One of the key advantages of this approach is its scalability. With a high degree of measurement automation, the dependency on human resources is greatly reduced, making large-scale deployment more feasible. This would enable more frequent monitoring of road networks, facilitating the early detection of road defects and the timely identification of sections requiring maintenance. As a result, road repair and maintenance operations can be optimized and predicted more effectively, leading to reduced costs and improved resource allocation. In addition, multi-sensor imaging systems would still be needed, however, these would only be required in the identified poorer quality sections to provide a more accurate and complete picture for final decision-making and reconstruction planning. The primary

stakeholders of these systems are governmental or local road maintenance agencies responsible for assessing and maintaining road quality. By integrating such an automated monitoring solution, these organizations could enhance the efficiency of road maintenance planning, ultimately improving road safety and infrastructure longevity.

This research aims to develop and implement a universal, low-cost end-to-end software sensor-based measurement system that classifies individual road sections into generally interpretable quality categories according to their surface characteristics. The system needs to rank individual sections as they have greater relevance than the separable road defects, according to the above criteria of road reconstruction schedules and health issues. The system should be able to detect road defects affecting surface characteristics, such as potholes, rutting, corrugation, depression, and surface deformations caused by manholes. Finally, the system needs to provide a comprehensive and easy-to-interpret measure of the condition of each road section based on a number of input variables.

Based on the literature review, I have identified three critical aspects of vehicle vibration-based road quality measurement systems that need to be investigated or improved to achieve the above objectives:

1. The first research question is addressed to the sensor placement: Does the vibration measurement on the vehicle suspension provide useful information content compared to the data recorded in the vehicle cabin for the subsequent evaluation process and, if so, can this improve the accuracy of the final classification algorithms?
2. The second research question is related to the distortion effect of sudden vehicle maneuvers: Do abrupt longitudinal and lateral maneuvers with the vehicle have a negative impact on the road quality evaluation process and, if so, what are possible methods to correct for these negative effects?
3. The third and final question addresses the data labeling problem for subsequent machine learning models: What alternative method could be used instead of traditional manual labelling procedures that would also provide flexibility and higher spatial resolution in the evaluation process?

These challenges are critical to the development of an accurate and industry-ready standalone measurement system. Therefore, to address these research questions this work aims to

- design and validate a measurement system for automated and comprehensive road condition monitoring using inertial and location sensors mounted on vehicles,
- develop techniques for the automatic collection of high-quality and sufficient datasets,
- implement a processing workflow to preprocess road data based on measurements of inertial sensors,
- develop an evaluation method for pavement quality estimation from a range of input variables,
- and lastly, incorporate the developed algorithms into a graphical user interface to evaluate datasets effectively.

### 3.2 Novel methods for road condition measurement

This subsection introduces the conceptual foundation and structure of the proposed road quality monitoring system, which combines low-cost sensor technologies with automated software processing to deliver a robust, scalable, and practical solution for real-world applications. Following this conceptual overview, the detailed hardware setup and software implementation strategies will be addressed in subsequent sections.

In particular, the use of vehicle-mounted inertial sensors presents a promising direction for road condition estimation. While previous works have explored similar approaches, as discussed in Section 2, recent studies have highlighted notable gaps in data cleaning, processing pipelines, and final evaluation methods. The present work seeks to address these shortcomings through an integrated, end-to-end approach.

To develop a system that is viable in operational environments, a holistic systems-level perspective is necessary. A holistic design approach considers not just the isolated task of road quality classification, but also how each subsystem interacts with the others.

When deployed in practice, a road condition monitoring system must meet several criteria to ensure practical applicability. First, the system must operate continuously and without the need for manual input during measurements. This is also true for the processing phase, the data should be evaluated automatically with minimal human intervention. Furthermore, affordability is critical for widespread deployment, especially across diverse vehicle fleets. Based on these requirements, the structure and flow of the proposed system are outlined in Fig. 7, which provides a comprehensive overview of the concept for road

infrastructure monitoring.

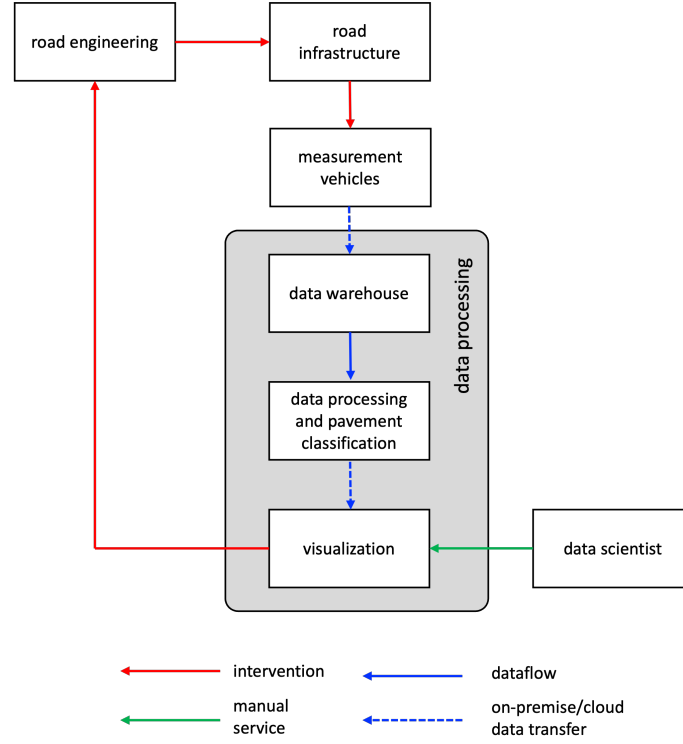


Figure 7: The structure and flow of the proposed data processing system

The diagram illustrates the data flows and processing stages. The entry point of the system is the local road management authority, which oversees the road network and operates the measuring vehicles. These vehicles must be able to interface directly with the data processing framework. The raw measurement data is transmitted and stored in a structured and queryable database environment. Subsequently, data processing takes place in a dedicated environment, where sensor noise correction, synchronization, and classification are performed. This processing is modular and customizable to accommodate local infrastructure demands and road classification standards. The final stage of the pipeline focuses on result visualization. Interactive map-based visualization tools allow the results to be explored geographically, with filtering options that help prioritize intervention. These visual dashboards and automatically generated reports form the basis for decision-making and planning. Lastly, road quality can change quickly due to weather conditions, heavy use, or seasonal cycles. Therefore, the system must support periodic re-measurements and automatic data integration to maintain up-to-date records.

Using a systems-based engineering approach, the main components were defined, which concept is shown in Fig. 8, designed in accordance with the earlier specified requirements. The core of the system is a compact data logger, which stores synchronized data streams

from multiple sensors during measurement. The primary sensor is a tri-axial Inertial Measurement Unit placed in the vehicle cabin, as close as possible to the center of gravity. To supplement this and capture additional vibration information at unsprung mass, two single-axis accelerometers are mounted directly onto the left and right rear suspension control arms. These are directly exposed to road-induced forces, making them ideal for collecting high signal-to-noise ratio acceleration data that correlates closely with road surface quality. Finally, a GPS receiver installed within the vehicle cabin provides spatial coordinates and velocity data for route tracking, spatial segmentation, and later map-matching.

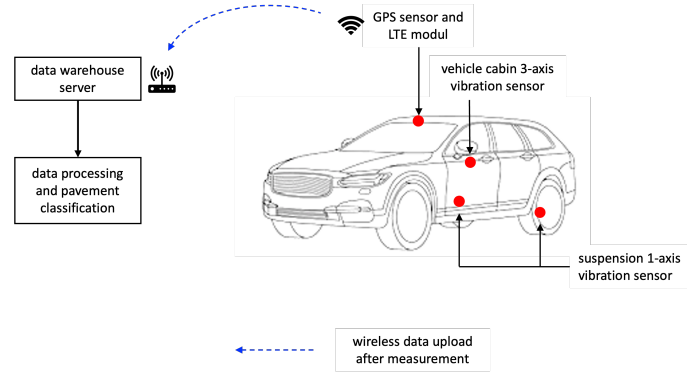


Figure 8: The concept of the measurement system with the applied sensors and the overview of the data flow [71]

To ensure smooth analysis, raw measurement data must be made rapidly available for further processing. Therefore, the system is designed to transmit all collected data wirelessly to a centralized storage system. This is achieved through an integrated LTE communication module that maintains an automated data stream between the data logger and a remote data warehouse. As a result, the structured and pre-formatted data can be processed automatically in a centralized environment, where the road segments are classified and visualized with minimal human intervention.

The selection of sensor locations was driven by practical engineering constraints rather than an explicit optimization procedure. Specifically, sensors have to be fixed to mechanically stable, flat surfaces to ensure secure attachment and minimize external vibrations unrelated to road inputs. For this, the control arms were chosen for their direct exposure to road-induced excitations, while the IMU was placed centrally to approximate whole-vehicle motion without extensive modifications to the vehicle body. Fig. 9 illustrates the orientation of the primary measurement axes as well as the notation applied to the sensor-derived variables used in later processing steps. Although no formal CoG estimation or suspension dynamics optimization was performed, the chosen locations provide a balance between

practical sensor installation and effective signal acquisition for road quality assessment.

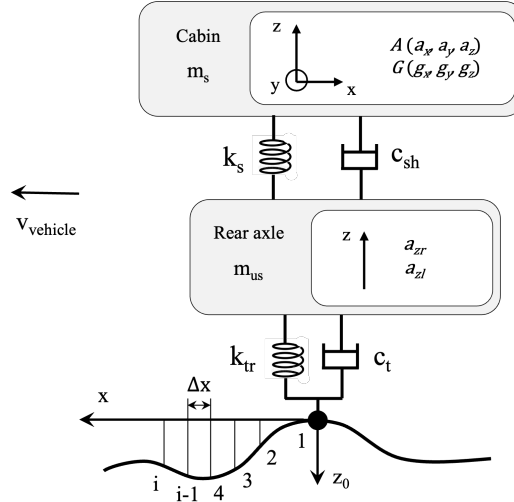


Figure 9: The orientation of the measurement axes with the notations applied to the sensor-derived variables

The single-axis accelerometer sensors were mounted on the upper surface of the front suspension control arms (rocker arms) of the vehicle. In their initial orientation, these surfaces are approximately vertical, ensuring that the accelerometers primarily capture vertical acceleration components induced by road irregularities. To provide mechanical protection and ensure a stable attachment, each sensor was enclosed in a custom-designed 3D-printed housing. The complete sensor assembly was securely mounted to the control arms using pre-existing threaded holes in the structure, with screws providing a stable and repeatable installation without requiring modifications to the vehicle's suspension components.

A six-degree-of-freedom Inertial Measurement Unit sensor was installed inside the passenger compartment to monitor the overall vehicle dynamics. The IMU was positioned in the rear transmission tunnel, located behind the front armrest of the vehicle, where it was securely bolted in place using existing factory-drilled holes. This location was chosen due to its proximity to the estimated center of gravity and its structural integration with the vehicle chassis.

### 3.3 Hardware and embedded software structure of the system

In this section, the physical implementation of the measurement system is introduced. The hardware was developed based on the previously defined functional and system-level requirements. The process began with selecting appropriate sensor types for the specific

measurement roles, followed by the design of the central data acquisition unit, and finally the specification of the communication and data transmission processes.

The unit is responsible for handling all incoming sensor signals, performing analog-to-digital conversion when needed, organizing sensor data into structured tables, temporarily buffering the data, and then transmitting it to a cloud-based database for further processing. After analog-to-digital conversion, all sensor signals are represented in binary form and transmitted internally via serial digital communication interfaces, such as I<sup>2</sup>C. Furthermore, the unit also provides an interface for sensor configuration and allows for controlling the measurements.

Different sensor types were required depending on their location: one set installed on the suspension system, and another in the vehicle cabin. For the cabin, a multi-axis integrated sensor unit was preferred, with a capability of measuring across several degrees of freedom, while directly connectable to the central data logger.

In contrast, the acceleration sensors mounted on the rear suspension required a more rugged and compact solution. The environment near the suspension exposes the sensors to frequent shocks and environmental stress. To meet these challenges, simplicity and mechanical durability were prioritized. Additionally, it was deemed sufficient to measure only the vertical acceleration at these locations, since the geometry of the suspension system physically constrains movements along other axes. This is unlike the cabin environment, where movement is less mechanically restricted and three-dimensional vibration measurement is necessary.

**Cabin-Mounted Inertial Sensor** To measure the motion of the vehicle chassis, a Microelectromechanical Systems (MEMS) based Inertial Measurement Unit was selected. MEMS sensors provide a favorable combination of small size, low power consumption, and high sensitivity. They are particularly well-suited for embedded vehicular applications. The chosen unit was the MPU6050 IMU, which integrates a 3-axis accelerometer, a 3-axis gyroscope, and an optional magnetometer in a single compact package. It supports digital output over the standard I<sup>2</sup>C protocol, making it directly connectable to microcontrollers.

**Suspension-Mounted Acceleration Sensors** To capture the vibrations of the unsprung mass, a piezoelectric accelerometer was selected for each of the rear suspension control arms. The selected model is the 805M1-0020 from TE Connectivity. This sensor meets all the requirements, as it operates with just three wires, has a cylindrical threaded



housing for simple mounting, and supports an appropriate range of vibration amplitudes commonly observed in road surface analysis. The analog signal output corresponds directly to the direction and magnitude of the vibration, centered around a 2.5 V DC bias. To align the signal with the zero reference for digital conversion, a bias-shifting circuit was designed to subtract the 2.5 V offset and re-center the waveform at 0 V baseline.

The key specifications of the two primary sensor types used in the system are summarized in Table 1.

Table 1: Description of the applied vibration sensors' specifications [72, 73]

| Specification      | MPU6050 (Cabin)                         | 805M1-0020 (Suspension)      |
|--------------------|-----------------------------------------|------------------------------|
| Sensor type        | MEMS IMU                                | Piezoelectric accelerometer  |
| Measurement axes   | 3 (acc + gyro)                          | 1 (vertical)                 |
| Acceleration range | $\pm 2g$ to $\pm 16g$                   | $\pm 20g$                    |
| Gyroscope range    | $\pm 250^\circ/s$ to $\pm 2000^\circ/s$ | –                            |
| Frequency response | Up to 1 kHz                             | 2 Hz – 5 kHz                 |
| Sensitivity        | 16-bit ADC                              | 100 mV/g                     |
| Signal output      | Digital (I <sup>2</sup> C)              | Analog (0–5 V)               |
| Output bias        | –                                       | 2.5 V                        |
| Power supply       | 3.3–5 V                                 | 5 V                          |
| Enclosure type     | PCB module                              | Cylindrical, threaded (IP65) |

**GPS Module** To complement the motion data with spatial information, a GPS sensor was also integrated into the system. The sensor type is the Waveshare SIM7600, which has the advantage of being compatible and easy to connect to a microcomputer. The sampling frequency of the unit is 10 Hz [74], which is much lower than previous vibration sensors, therefore these have to be synchronised. The unit was installed in the cabin with a direct connection to the data logger. The selected unit outputs position, altitude, speed, and UTC timestamp.

**Data Transmission and Connectivity** After introducing the sensor configuration, I describe the data acquisition unit below, which schematic layout is illustrated in Fig. 10.

The unit is responsible for handling all incoming sensor signals, performing analog-to-digital conversion when needed, organizing sensor data into structured tables, temporarily buffering the data, and then transmitting it to a cloud-based database for further processing. Furthermore, the unit also provides an interface for sensor configuration and allows for controlling the measurements. Considering the complexity of data management, multi-sensor integration, and the requirement for user interaction via touchscreen Graphical User Interface (GUI), a single-board microcomputer has been selected for the main controller.

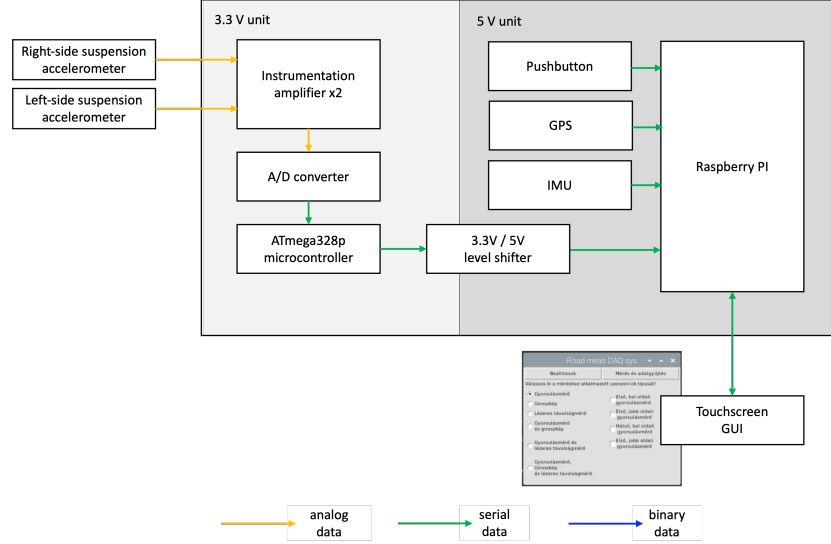


Figure 10: Schematic layout of the hardware of data acquisition unit

The system is built around a Raspberry Pi 4 Model B, which serves as the central control unit for hardware interfacing and external communication. This device is commonly used in Internet of Things or IoT applications due to its cost-effectiveness and robust support ecosystem. The Raspberry unit provides multiple digital interfaces, including I<sup>2</sup>C, SPI, and UART, which enable easy communication with external peripherals. The board operates on a stable 5 V input, while the actual system data is stored on a microSD card.

The data acquisition system defines two independent power domains: a 3.3 V line for the analog piezoelectric sensors, and a 5 V line for the digital sensors and the Raspberry Pi itself. To ensure interoperability between these components, a 4-channel bi-directional voltage level shifter is used for signal transmission. Analog outputs from the piezoelectric accelerometers are connected through custom-built signal conditioning circuits. These circuits handle gain adjustment and DC offset correction. The conditioned signals are then digitized using a MAX186 12-bit ADC, which provides a resolution of 1.22 mV over a 0–5 V input range.

To offload the processing burden from the Raspberry Pi and enable real-time buffering, an ATmega328 microcontroller was employed. It manages data retrieval from the ADC, formats the data, and communicates with the Raspberry Pi over an I<sup>2</sup>C master-slave protocol. Since the ATmega operates at 3.3 V and interfaces with 5 V peripherals, dual-channel voltage shifters were inserted.

To support real-time access and continuous updates, the measurement system transmits data wirelessly to a centralized data warehouse. This is facilitated via a 4G LTE com-

munication module capable of automatic data synchronization. Furthermore, a dedicated power supply was developed for the system, of which the input voltage range of the unit is between 7.5 V and 35 V, allowing it to draw power directly from a car’s cigarette lighter socket. Internal regulation and filtering circuits provide the necessary 3.3 V and 5 V rails for the sensors and Raspberry Pi, ensuring stable operation even under fluctuating automotive power conditions [75].

Finally, for local interaction, the Raspberry Pi is connected to a capacitive touchscreen display via HDMI and USB interfaces. A full-featured Graphical User Interface was developed and deployed, enabling system configuration, sensor monitoring, and measurement control. In addition to the touchscreen, a physical pushbutton was integrated into the hardware system. This button allows for starting or stopping a measurement session without the need for visual attention to the screen.

**Software Architecture** Following the hardware design, I move on to the software architecture, which is illustrated in Fig. 11.

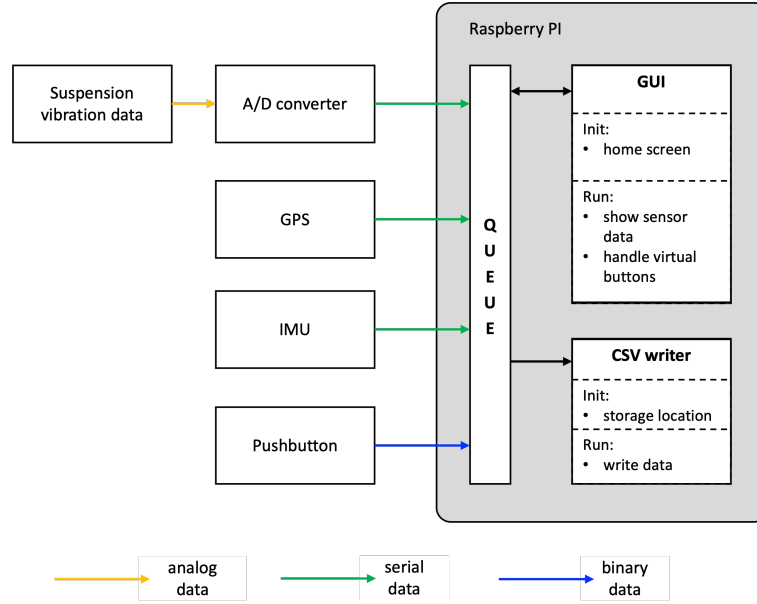


Figure 11: Illustration of the software architecture of the data acquisition system

The software system of the proposed monitoring platform is designed to acquire, process, and store sensor data in a structured format. It is built around a modular, queue-based architecture, where incoming sensor data is received, buffered, and processed sequentially to ensure time-consistent logging and robust data integrity. This architecture allows multiple tasks to operate in parallel without data loss.

The graphical user interface component serves as the operator-facing front end of the software. It is responsible for initializing the measurement system, managing user interactions, and displaying real-time feedback from all connected sensors. During operation, the GUI provides continuous updates about the measurement status, including recording progress, elapsed time, and potential sensor or communication errors. As such, it enables complete control of the system in an intuitive and user-friendly manner. Simultaneously, a dedicated CSV writer module manages long-term data storage. This component initializes the file system and saves the processed sensor streams into well-structured, timestamped log files.

The complete unit with the PCB panels and the software interface is shown in Fig. 12.

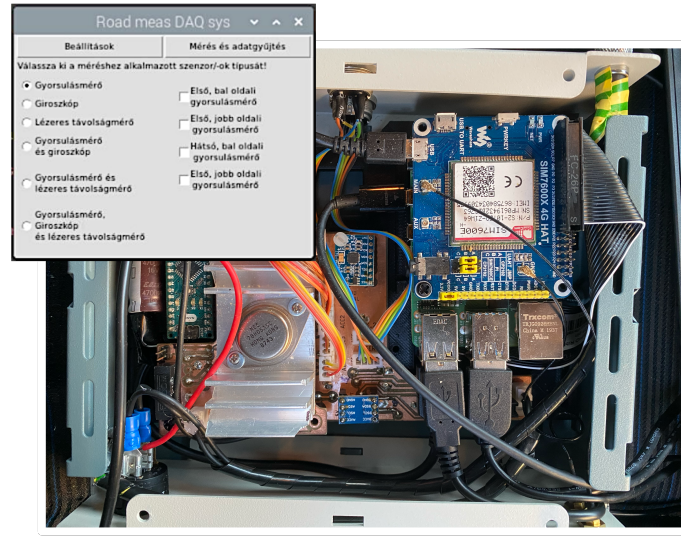


Figure 12: Picture of the central unit of the developed data acquisition system, in addition to the software interface

**Data Flow In Detail** The data flow within the proposed road quality monitoring system follows a structured process, as shown in Figure 13. The analog signals from the two piezoelectric accelerometers, denoted as  $a_{zl}$  and  $a_{zr}$ , are transmitted through shielded wiring to a vehicle-mounted data logger, where they are digitized and synchronized with additional sensor inputs, including GPS and IMU data. The Raspberry PI-based data logger records the vehicle's geospatial position ( $lat, lon$ ), velocity ( $v$ ), and the accelerations along the three measurement axes as  $a_x, a_y, a_z$  and angular velocities as  $\omega_x, \omega_y, \omega_z$ . Once collected, the recorded sensor data is wirelessly transmitted via a 4G LTE network to a central cloud data server. Here, the data is securely stored as CSV files for further analysis.

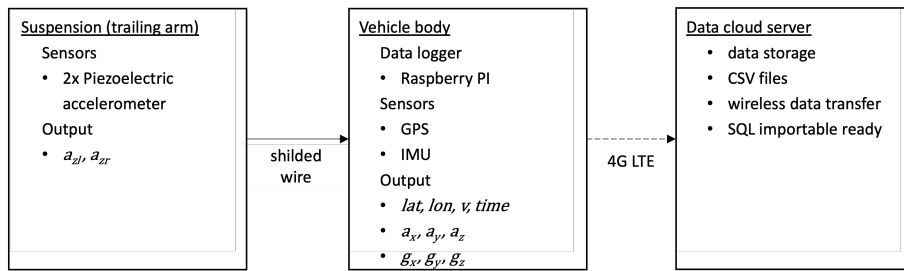


Figure 13: The data flow within the proposed measurement system

## 4 Implementation of the software framework

This chapter presents the architecture of the software framework, which forms the foundation of the entire measurement and evaluation system developed during the research. The design and implementation of this software-based data processing and analysis unit represent the core contribution of my work.

Based on the defined system requirements and critical evaluation of existing approaches, the primary objective was to address and overcome the limitations observed in earlier, similar projects. The developed solution introduces an integrated, modular approach that facilitates robust, scalable, and automated data processing from raw measurement to interpretable output. The presented software framework encompasses all key stages of computational analysis: data cleaning, preprocessing, evaluation, and final visualization. These modules collectively constitute a complete informatics solution tailored to the needs of the measurement system. The architecture of the framework is structured in multiple layers, each responsible for a specific stage of the data analysis pipeline. The subsequent sections of this chapter follow the same hierarchical structure, discussing each component and its internal logic in detail.

The first step of the project is data preparation and preprocessing (Section 4.2) focusing on handling data errors and data cleaning, followed by feature extraction and feature engineering (Section 4.3), which focuses on finding and creating informative and useful features. To train the model, a database with reference labels is needed, so prior to this, the raw data needs to be labelled; the real road quality class needs to be assigned to each measurement data point (Section 4.4). After labelling, a stand-alone classification model can be created, which can assign new, previously unknown data points to the appropriate road quality category. The models tested and the process of training and testing are described in Section 4.5. Finally, in Section 4.7, I detail solutions related to visualization, such as the assignment of individual road segments to more expressive location addresses based on raw location data using geocoding and map matching.

### 4.1 Structure of the algorithm

This subsection provides a comprehensive overview of the entire data processing algorithm, offering a holistic perspective to facilitate the understanding of the data flow and the core concepts applied throughout the framework.

The primary aim here is to outline the major subcomponents of the algorithm and identify the key challenges associated with each. These challenges include, but are not limited to, handling noisy or incomplete data, performing consistent preprocessing across diverse signal types, reducing dimensionality without information loss, and ensuring that the extracted features are both interpretable and relevant for the intended analytical tasks. In addition to summarizing the system’s logic, this subsection also serves as a structural guide for the remainder of the current chapter. Each module introduced here will be discussed in detail in the following subsections.

#### 4.1.1 Detailed approach and setup

The exact steps of data mining pipeline and statistical methods, which are applied in this work, are shown in Fig. 14. The raw data from the sensors is provided in a structured and organised format. Subsequently, the raw measurement data is fed into the data processing pipeline, where the first step is data preprocessing, such as handling data errors, synchronising sampling from different sensors and signal filtering. After this initial step, features are extracted from time series acquired from sensors to reduce the dimension and highlight information. In the last feature engineering step of the pipeline, new variables are created by changing, transforming and combining the original variables, again in order to obtain information relevant to the application.

As indicated in the figure, the data stream can then be separated into two parts, the development phase and the in-operation phase, as the latter no longer requires prior reference labeling, human supervision and model training. First considering the development section on the left, the dataset without ground truth labels for road surface quality has to be labeled. These ground-truth labels plays a central role in supervised learning tasks, as the performance of any subsequent model is highly dependent on the quality and consistency of these reference labels. Based on insights from the literature and findings from preliminary development phases, the manual labeling was not a feasible option in this context. The big volume of data points causes manual annotation impractical. Furthermore, manual labeling introduces uncertainty and variability due to human error and subjective interpretation.

Alternative approaches involving real-time labeling during data collection were also evaluated in previous studies [1, 6, 32]. However, these methods suffer from their own limitations, including inflexibility and similarly high subjectivity. However, when measuring

and evaluating road quality, there are many variables to which the system needs to adapt. These may include varying road quality category classes from country to country, or varying measurement needs with a varying number of target variables and pavement classes. Manual annotations, once completed, are typically fixed and difficult to update retrospectively. As a result, the system required a more flexible and robust solution to labeling that could accommodate changes in measurement criteria or categorization schemes without a complete re-labeling of the dataset. Therefore, a central design requirement of the framework was the development of a dynamic and adaptable labeling strategy. To address these challenges, the final system employs a Machine-In-the-Loop (MIL) labeling technique (Subsection 4.5.1). This approach uses unsupervised learning models to identify statistical patterns within the data and assign categories based on inherent structures, without requiring predefined reference labels.

The pre-labeled database can now be used for training a final and production-ready standalone classification model, which needs a training phase to find the relationship between the dependent variable and the explanatory variables (Subsection 4.5.3). From the supervised learning's testing and training phases, performance measures can be calculated to estimate the generalization (Subsection 4.6.2).

The two paths are linked together through the finalised supervised model. The trained classification model can then be independently applied to make new estimates on the processed but unlabeled data. It is important to note that the input for this is provided by data from the same data processing pipeline as in the development phase. Finally, the raw coordinate values can be transformed into more expressive geographic designations for easier interpretation and visualization using reverse geocoding and map matching (Subsection 4.7).

All computations and data processing steps presented were implemented in the Python 3.12.2 programming language, the most widely used programming language in the industry today, due to its many available add-on libraries and open source code, as well as its free access [76]. The implementation relies on widely used open-source scientific libraries to ensure reproducibility and robustness. Numerical computations and data manipulation were performed using NumPy and Pandas, while machine learning methods—including classifiers, clusterers, and Principal Component Analysis were implemented using the scikit-learn framework. Signal processing and statistical calculations were supported by the SciPy library. Geospatial processing tasks, such as map matching and geocoding, were carried out



using OpenStreetMap-based tools, including OSRM and the mappymatch library. Custom-developed code was used for data preprocessing, feature extraction, workflow integration, and application-specific processing steps.

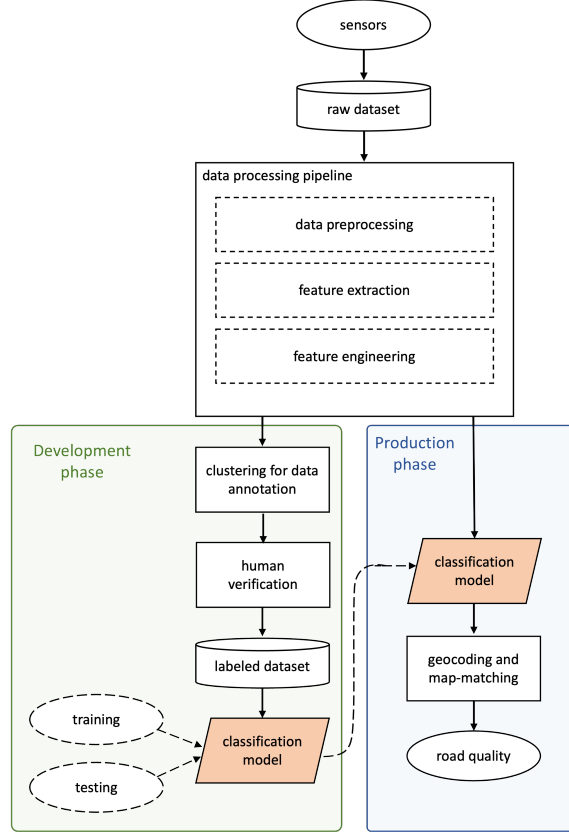


Figure 14: The detailed workflow of data mining pipeline and evaluation methods, with the separation of the machine learning phase into two parts: in green the development and training phase and in blue the final production phase with the evaluation method

#### 4.1.2 Notation scheme

The following gives a brief description of the notation system that is used in the following sections.

In general,  $n$  represents the number of distinct data points, or observations in the sample. In addition,  $p$  denotes the number of variables from measurements that are available for subsequent analysis and model training. For the whole database, two main cases are distinguished, where  $x_{raw}$  denotes the raw measurement data and  $x$  the processed database, which can be the input matrix of the clustering or classification problem. Regarding these databases,  $x_{raw,ij}$  and  $x_{ij}$  represents the value of the  $j$ th variable for the  $i$ th measurement sample or observation, where  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, p$ . Throughout the work,  $i$  will be used to index the samples or observations (from 1 to  $n$ ), and  $j$  will be used to index the variables (from 1 to  $p$ ).

The  $x_{raw}$  raw data matrix and  $x$  input matrix for the subsequent machine learning model training denotes an  $n \times p$  matrix whose  $(i, j)$ th element is  $x_{raw,ij}$  and  $x_{ij}$ , and defined as

$$x_{raw}, x = \begin{bmatrix} x_{raw, x11} & x_{raw, x12} & \dots & x_{raw, x1p} \\ x_{raw, x21} & x_{raw, x22} & \dots & x_{raw, x2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{raw, xn1} & x_{raw, xn2} & \dots & x_{raw, xnp} \end{bmatrix} = \begin{bmatrix} x_{raw, x1}^T \\ x_{raw, x2}^T \\ \vdots \\ x_{raw, xn}^T \end{bmatrix},$$

and the output vector  $\mathbf{c}$  for the unsupervised machine learning-based clustering with  $\mathbf{c} \in \{1, 2, \dots, K\}^n$  as

$$\mathbf{c} = (c_1, c_2, \dots, c_n)^T,$$

and the output vector  $\mathbf{y}$  for the supervised classification model with  $\mathbf{y} \in \{1, 2, \dots, L\}^n$  as

$$\mathbf{y} = (y_1, y_2, \dots, y_n)^T.$$

For clustering,  $c_i \in \{1, 2, \dots, K\}$  denotes the cluster label corresponding to  $x_i$ , where  $K$  is the total number of clusters. Similarly, for classification,  $y_i \in \{1, 2, \dots, L\}$  denotes the class label corresponding to  $x_i$ , where  $L$  is the total number of classes. Furthermore, in the next subsections, a general example of the input matrix for each stage is given.

## 4.2 General data preprocessing

A well-performing classification model can only be developed if the input data is pre-processed correctly, transformed into a suitable structure, and enhanced with informative features. The purpose of the general preprocessing phase is to eliminate data inconsistencies, correct sensor-specific issues, and harmonize the sampling rates across different sources. The following preprocessing sequence includes the below steps:

1. GPS data upsampling,
2. trajectory resampling and smoothing,
3. bias and trend removal from accelerometer and gyroscope signals,

4. sensor synchronization verification,
5. spatial domain resampling,
6. and infinity norm normalization.

The input to the preprocessing pipeline is the raw sensor dataset, denoted by  $x_{\text{raw}}$ , which includes GPS coordinates, acceleration, and gyroscope readings recorded during road quality measurement sessions. These data points are retrieved from the measurement database in structured form, as illustrated in Table 2.

Table 2: Example of an input matrix for feature extraction

| Sample   | $a_x$<br>(m s <sup>-2</sup> ) | $a_y$<br>(m s <sup>-2</sup> ) | $a_z$<br>(m s <sup>-2</sup> ) | $\omega_x$<br>(s <sup>-1</sup> ) | ...      | Longitude<br>(°) |
|----------|-------------------------------|-------------------------------|-------------------------------|----------------------------------|----------|------------------|
| 1        | 0.67                          | 0.51                          | 1.18                          | 3.06                             | ...      | 16.833320        |
| 2        | 0.71                          | 0.65                          | 1.21                          | 2.98                             | ...      | 16.833361        |
| 3        | 0.62                          | 0.57                          | 1.16                          | 3.02                             | ...      | 16.833421        |
| $\vdots$ | $\vdots$                      | $\vdots$                      | $\vdots$                      | $\vdots$                         | $\ddots$ | $\vdots$         |
| $n$      | 0.51                          | 0.38                          | 0.91                          | 2.84                             | ...      | 16.873221        |

This data table does not yet have ground truth data labels and segmentation has not yet been performed. The individual features are given by the data measured by the IMU, the accelerometers and the GPS sensor.

**GPS data preparation** The first task in the preprocessing workflow involves preparing the spatial coordinates derived from the GPS sensor. As the GPS unit operates at a much lower, 10 Hz sampling frequency compared to the inertial sensors, which capture vibration-related signals at 400 Hz, the sampling frequencies requires temporal alignment. To achieve this, the original location data is upsampled using one-dimensional piecewise linear interpolation. This interpolation method estimates intermediate values along the time axis, effectively resampling the trajectory to match the IMU time base while preserving the structure of the original GPS path.

Following the temporal resampling, a moving average-based trajectory smoothing procedure was applied to the raw GPS coordinates in order to enhance the spatial consistency of the recorded trajectory. For trajectory smoothing, a one-dimensional moving average filter was employed, applied independently to each coordinate component. The moving average filter smooths the data by replacing each point with the average of its neighboring values within a fixed-length window.

**Corrections on vibration signals** The next step addresses the vibration sensor signals, where the two points to be handled are the zero-point bias and non-stationary trends. Accelerometers are prone to small static offset errors, known as bias, due to calibration drift or environmental conditions. Gyroscope signals, on the other hand, often exhibit long-term drift or low-frequency trends that compromise stationarity. These non-stationary components distort both frequency and amplitude-based analysis.

To correct these issues, each time series signal undergoes a detrending algorithm. This involves fitting a low-degree polynomial with least squares regression to the signal and subtracting it to isolate the high-frequency, informative component of the signal, and removing low-frequency drifts.

Accurate evaluation of pavement quality using vibration-based methods requires careful consideration of vehicle dynamics, particularly speed. Prior research by Ward and Iagnemma [45] has demonstrated that the amplitude of vehicle-induced vibrations increases with speed, even when the pavement condition remains unchanged. This speed dependence introduces a significant source of variability in vibration measurements, which must be mitigated to ensure reliable road surface assessment.

A widely used approach to address this issue is spatial resampling, where vibration data are resampled at constant spatial intervals ( $\Delta x$ ) rather than at constant temporal intervals. This technique eliminates speed-induced distortions by transforming the dataset into a spatially uniform representation. Furthermore, this uniform spacing is essential for applying the subsequent Fourier Transform, which breaks the signal into frequency components. In this work, both IMU and GPS data were resampled using spline interpolation. The process involves computing the cumulative travel distance from raw GPS coordinates and interpolating all sensor measurements at fixed spatial steps.

The resampling interval  $\Delta x$  determines the spatial sampling frequency:

$$f_s = \frac{1}{\Delta x}, \quad (3)$$

expressed in cycles per meter (1/m). According to the Nyquist-Shannon sampling theorem, the maximum spatial frequency that can be resolved from the data—referred to as the Nyquist frequency—is given by

$$f_{\text{Nyquist}} = \frac{f_s}{2} = \frac{1}{2\Delta x}. \quad (4)$$

This frequency corresponds to the smallest distinguishable surface features. For practical implementation, the spatial resolution  $\Delta x$  should be chosen such that it captures relevant terrain variations while remaining consistent with the vehicle’s operating speed [45]. In this study,  $\Delta x = 0.2$  m was selected, with a trade off between sufficient detail for localized terrain analysis and oversampling.

Lastly, to preprocess the vibration data, infinity norm normalization was employed. This technique rescales each sample independently by dividing all feature values by the maximum absolute value within that sample. Formally, for a row vector of covariates

$$\mathbf{x} = [x_1, x_2, \dots, x_p],$$

the normalized vector  $\mathbf{x}'$  is computed as

$$\mathbf{x}' = \frac{\mathbf{x}}{z},$$

where

$$z = \|\mathbf{x}\|_\infty = \max_i |x_i|. \quad (5)$$

This normalization guarantees that the largest absolute value in each row becomes 1, effectively scaling all elements of  $\mathbf{x}'$  to lie within the interval  $[-1, 1]$ . Furthermore, the method preserves the internal structure and relative differences of the features, while reducing the influence of varying magnitudes and units across datasets. By standardizing feature scales in this manner, infinity norm normalization can enhance the robustness of subsequent analyses.

### 4.3 Aggressive maneuver detection and correction

In this subsection, I describe sudden, so-called aggressive vehicle maneuvers, which represent a significant error potential for road quality measurement. As these aggressive maneuvers have not been addressed in previous studies, I will first focus on their impact and then present a method developed to detect and correct them, which can be used to further improve the accuracy of vibration-based road quality measurement systems.

#### 4.3.1 Overview of the importance of maneuver detection

In the previous subsection, the general data preprocessing workflow was described, focusing primarily on correcting common sensor anomalies using standardized methods. Further-

more, as outlined earlier, the vibration-based road quality measurement system employed in this research utilizes multi-axis accelerometers and gyroscope sensors to assess road surface conditions. These sensors are mounted to capture vibrations and angular motion originating from the vehicle’s suspension and chassis, which are in turn affected by road irregularities.

However, sensor readings can also be influenced by dynamic vehicle maneuvers such as sudden braking, acceleration, or sharp turns. These movements, often referred to as aggressive maneuvers, can introduce false signals into the data stream, leading to incorrect amplitude, distorted frequency content, or altered waveform shapes [2]. Since such distortions are not attributable to the road itself, they represent a significant source of error in vibration-based assessments. Aligned with this, our observations from both experimental trials indicate that the accuracy of road quality classification models significantly improves on straight or moderately curved suburban segments. Similar trends have been reported in other studies [6], which suggests a systemic sensitivity of vibration-based systems to vehicle dynamics induced by turning or abrupt longitudinal maneuvers.

In this context, identifying and reducing the impact of aggressive maneuvers becomes of great importance. Unlike standard sensor noise, these errors are not effectively removed through conventional filtering techniques, so, a specialized solution had to be developed to detect and isolate these maneuvers.

While uniform driving behavior is ideal during data collection, real-world traffic conditions often prevent it. Because the system is designed to function with standard, unmarked vehicles, unlike traditional road assessment platforms that require dedicated and visually marked vehicles, drivers are more susceptible to unpredictable traffic influences. As a result, sudden maneuvers are unavoidable, and their effect on the measured data must be addressed.

The effects, detection, and isolation of these so-called aggressive maneuvers have not typically been addressed in previous studies, despite its potential to improve measurement reliability. Furthermore, solving this problem through a software-based approach makes the solution highly accessible and cost-effective, as it does not require changes to the hardware setup. However, previously proposed maneuver detection methods are not applicable to this use case for two main reasons:

- Most automotive-grade algorithms rely on a diverse set of sensor inputs—such as

steering angle, throttle position, or wheel speed sensors—that are unavailable in vibration-only road quality systems.

- IMU-only maneuver detection algorithms that exist in the literature typically focus on different maneuver categories and are computationally intensive, making them impractical for integration into lightweight embedded systems used in road quality measurement [77].

To overcome these limitations, I propose the development of a software module capable of detecting and correcting signal segments distorted by aggressive driving maneuvers. This module acts as a software sensor, designed to identify aggressive maneuvers that would potentially reduce the accuracy of the road quality classification. The proposed algorithm must distinguish between two classes of motion:

- Normal driving behavior, which includes smooth accelerations, decelerations, and cornering that do not interfere with vibration-based measurements.
- Aggressive maneuvers, which generate distortive artifacts in the vibration signal and therefore degrade measurement reliability.

Upon detection, affected segments are flagged and corrected to restore the validity of the dataset before subsequent classification. For the purpose of maneuver detection, only the IMU sensor mounted inside the vehicle cabin was used. Based on both literature review [63, 69, 78] and experimental observations, the vertical acceleration signals measured at the suspension arms were found to be less relevant for distinguishing vehicle maneuvers. Therefore, excluding them allowed for a simpler and more computationally efficient algorithm without compromising detection accuracy.

The technical implementation and performance of this maneuver detection module are discussed in detail in the following subsections.

#### **4.3.2 Maneuver labelling**

The first step in the development of the maneuver detection algorithm involved labeling the raw vibration data from the piezoelectric and the IMU sensors. For the robustness of the classification model, it was important to define a minimal and well-structured set of maneuver categories. Since road vehicles operate within a limited dynamic range, previous studies and our own development experiences indicate that the most relevant maneuvers

affecting vibration-based measurements are abrupt accelerations, decelerations, and sharp turns—each exceeding a specific threshold in either linear acceleration or angular velocity.

There are two reasons for reducing the number of classes. First, it reduces the complexity of the model training process, helping to reduce the impact of class imbalance, which is a significant problem in this study, due to the nature of longitudinal manoeuvres, which are more frequent by nature and thus overrepresented in the data set. Second, it ensures a higher degree of internal consistency within each maneuver class. These homogenous signal patterns allow the algorithm not only to distinguish between maneuver types but also to apply tailored correction strategies during subsequent processing. Even within a single class, signal shape may vary in intensity but maintain a consistent structure, allowing us to handle error signals of different levels and intensities with the same shape, but varying degree of intervention in the subsequent correction phases.

In my research, five maneuver categories were differentiated between normal and abnormal driving conditions. These include:

- Normal driving: smooth motion with no significant dynamic disturbances.
- Aggressive acceleration: sudden increase in speed exceeding a defined forward acceleration threshold.
- Aggressive deceleration (braking): sudden decrease in speed, characterized by rapid negative acceleration.
- Aggressive right turn: angular velocity exceeding a positive threshold on the vertical axis.
- Aggressive left turn: angular velocity exceeding a negative threshold on the same, vertical axis.

These categories were selected based on our research and empirical observations from real-world measurements. The thresholds used for labeling were not based on sustained values but rather on transient peak events, which are more indicative of momentary maneuver-induced disturbances than long-duration behaviors. These transient spikes typically occur during rapid interactions with surrounding traffic, such as braking in response to obstacles or navigating in a curve.

To determine class membership, axis-specific threshold values were applied to the vibration,



and particularly to the IMU data. For accelerations, thresholds were set based on the longitudinal axis of the IMU’s accelerometer, while for turns, thresholds were derived from the angular velocity measured by IMU’s gyroscope. These limits were empirically calibrated using annotated driving sessions across two of the test vehicles and traffic scenarios. The threshold values for each class are summarized in Table 3. This labelling process created the ground truth required to train and validate the subsequent classification model, which will be discussed in the following subsections.

Table 3: Absolute threshold values applied for the determination of maneuver classes along the main axes

| Maneuver type           | Threshold limit                  |
|-------------------------|----------------------------------|
| Normal condition        |                                  |
| Aggressive acceleration | $\min( a_x ) = 7 \text{ m/s}^2$  |
| Braking (decelerations) | $\min( a_x ) = 10 \text{ m/s}^2$ |
| Right turn              | $\min( a_y ) = 5 \text{ m/s}^2$  |
| Left turn               | $\min( a_y ) = 5 \text{ m/s}^2$  |

#### 4.3.3 Data processing

The initial data preprocessing steps applied in the maneuver detection task were aligned with those previously described for road surface quality classification. This is justified by the use of the same sensor signals. However, while the preprocessing pipeline remains similar, the subsequent feature extraction and engineering steps differs significantly. This difference comes from the fundamental dissimilarity between the two target tasks. In maneuver detection, the objective is not to identify surface-induced vibrations, but rather to detect and isolate large-scale vehicle movements, typically characterized by low-frequency but high-amplitude signal components. Accordingly, the feature extraction process was tailored to emphasize longitudinal and lateral displacements, and to retain low-frequency components that would normally be filtered out in the pavement classification workflow. The following paragraphs discuss the specific methods used for aggressive manoeuvre detection.

**Dimensionality Reduction** To identify the most informative variables and suppress noise, Principal Component Analysis (PCA) was applied to the cleaned data. PCA transforms the dataset into a new coordinate system where the axes — called principal components — maximize variance in decreasing order. This transformation is computed through eigenvalue decomposition of the data covariance matrix. The principal component score matrix  $S$  is obtained as:

$$S = V\Lambda V^T \quad (6)$$

Here,  $V \in \mathbb{R}^{m \times m}$  is the matrix of orthonormal eigenvectors (loading vectors), and  $\Lambda$  is a diagonal matrix containing the non-negative eigenvalues ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ ) on its main diagonal [79]. The first few principal components, which account for the majority of the signal variance, were retained for further analysis. This step reduced data dimensionality while preserving relevant structural characteristics, thereby improving the signal-to-noise ratio.

**Segmentation Based on Change Point Detection** Following PCA, the reduced dataset, composed of the principal component score matrix  $S$  was segmented into coherent intervals, each representing a single maneuver. The segmentation task was carried out using a Change Point (CP) detection algorithm, designed to identify abrupt shifts in signal structure indicative of the start or end of a maneuver.

To ensure computational efficiency and real-time applicability, a lightweight condition-based method was implemented instead of model-heavy statistical change point detection algorithms. The following logical rule was used to detect change points:

$$\left( S_{m+1}^{\circ 2} \circledast \vec{1} \right) > 0.5 \Rightarrow CP_{m+1} = \text{true} \quad (7)$$

In this expression, the signal  $S$  is squared element-wise and convolved with an all-one vector  $\vec{1}$  of length 50. The squaring operation highlights changes in magnitude, while the convolution acts as a smoothing filter, attenuating short-term fluctuations. Subsequently, the algorithm identifies as segment boundaries the signals that exceed a unit of 0.5, and then the data location ID of the detected change point is saved in the  $CP$  vector. In this way, the  $CP$  vector marks the boundaries of the detected segments, which will be classified later into the proper maneuver class.

To evaluate the performance of the change point detection algorithm, the predicted segmentation boundaries were compared with manually labeled ground truth values. Since exact time alignment is challenging, quantitative assessment was performed using standard distance-based error metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Mean Absolute Error (MAE) provides a direct measure of the average absolute deviation between predicted and actual change points. It is robust to outliers and offers an interpretable measure of average performance across all segments:

$$MAE = \frac{1}{\#CP} \sum_{i=1}^{\#CP} |Predicted(CP_i) - Actual(CP_i)| \quad (8)$$

where  $Predicted(CP_i)$  and  $Actual(CP_i)$  correspond to the predicted and actual values at the  $i$ -th change point, respectively, and  $CP$  to the number of change points.

Mean Squared Error (MSE) squares the individual errors before averaging, which penalizes larger deviations more severely. This property makes MSE more sensitive to outliers and highlights performance degradation in scenarios with significant misalignment:

$$MSE = \frac{1}{\#CP} \sum_{i=1}^{\#CP} (Predicted(CP_i) - Actual(CP_i))^2 \quad (9)$$

Root Mean Squared Error (RMSE) is the square root of MSE, which restores the unit of measurement to the original scale. This makes RMSE especially useful for practical interpretation and direct comparison with the time domain resolution of the dataset:

$$RMSE = \sqrt{\frac{1}{\#CP} \sum_{i=1}^{\#CP} (Predicted(CP_i) - Actual(CP_i))^2} \quad (10)$$

In this study, all three metrics were applied to assess the quality of the segmentation algorithm.

**Segment-Wise Feature Generation** Following successful segmentation, a series of descriptive statistical features were computed for each segment. The aim was to extract meaningful characteristics that differentiate among the predefined maneuver classes. By selecting these appropriately, the number of input variables and the computational cost of the classification model can be further reduced [80].

The features were derived from time-domain characteristics of each segment, such as peak amplitude, variance, duration, and zero-crossing rate. Each of these was evaluated during the development phase for its predictive value and consistency across different driving scenarios. The final list of features tested and compared is summarized in Table 4, where

$S_i$  refers to the signal segment  $i$ .

Table 4: The examined descriptive statistics-based features

| Examined features    | Notation                               |
|----------------------|----------------------------------------|
| mean value           | $(\mu(S_i(1)), \mu(S_i(2)))$           |
| standard deviation   | $(\sigma(S_i(1)), \sigma(S_i(2)))$     |
| variance             | $(\sigma^2(S_i(1)), \sigma^2(S_i(2)))$ |
| peak value           | $(\max(S_i(1)), \max(S_i(2)))$         |
| skewness             | $(s(S_i(1)), s(S_i(2)))$               |
| kurtosis             | $(k(S_i(1)), k(S_i(2)))$               |
| standardized Z-score | $(z(S_i(1)), z(S_i(2)))$               |

The extracted features served as input for the maneuver classification algorithm, the development of which is discussed in the next subsection.

#### 4.3.4 Maneuver classification model

Once the dataset had been segmented using change point detection, each segment was classified into one of five predefined maneuver classes. This classification task forms the final step in the maneuver detection pipeline and enables the system to identify which signal segments contain driving behaviors that may distort road quality measurements.

**Handling Class Imbalance with SMOTE** Due to the nature of typical traffic behavior, the occurrence of linear maneuvers such as acceleration and braking was approximately two to three times more frequent than turning events. This led to an imbalanced dataset, where the minority classes were underrepresented. Imbalanced class distribution can severely affect the performance of supervised learning algorithms, causing them to be biased toward the majority class.

To mitigate this issue, the Synthetic Minority Oversampling Technique (SMOTE) [81] was employed to artificially augment the minority class. Unlike random oversampling, SMOTE generates new, synthetic samples by interpolating between nearby minority class instances in the feature space. This approach introduces more variability into the dataset while preserving the geometric properties of the minority class distribution.

The operation steps of the SMOTE algorithm is shown in Table 5.

By increasing the representation of under-sampled maneuver types, SMOTE improves the ability of the classification algorithm to learn decision boundaries for all classes equally [82].

Table 5: SMOTE algorithm steps

- 
1. Select a random minority class instance  $x^{(min)}$ .
  2. Identify its  $K$  nearest neighbors within the same class, based on Euclidean distance.
  3. Choose  $N$  neighbors randomly from these  $K$  to create the synthetic set  $A_1$ .
  4. Generate synthetic instances using linear interpolation between  $x^{(min)}$  and each neighbor  $x_K^{(min)} \in A_1$ :

$$\tilde{x}^{(min)} = x^{(min)} + |x - x_K^{(min)}| \cdot rand(0, 1) \quad (11)$$

where  $rand(0, 1)$  is a uniformly distributed random number between 0 and 1.

---

**Supervised Classification with Support Vector Machines** Following dataset balancing, the final classification step was carried out using a supervised learning algorithm. Supervised learning refers to a class of machine learning algorithms where the model is trained on input–output pairs. Each training instance consists of a feature vector  $\mathbf{x}_i$  and its associated ground truth label  $y_i$ , with  $p$  number of input feature variables. During training, the algorithm learns a function  $f : \mathbf{x}_i \rightarrow y_i$  that maps features to class labels by minimizing a loss function over the training set. Once trained, this function can be applied to new data to predict the corresponding output categories. In this algorithm, a Support Vector Machine (SVM) decision function has been used.

SVM is trained by the finding of an optimal separation hyperplane between classes in a high-dimensional space. It is effective for classification tasks with small-to-medium feature sets, and it performs well in high-noise environments due to its strong regularization properties. The core idea of SVM is to transform the input feature space into a higher-dimensional space via a kernel function  $k(x_i, x_{i'})$ , and then construct a hyperplane that maximizes the margin between different classes. In this work, a polynomial kernel of degree  $d$  has been used:

$$k(x_i, x_{i'}) = \left( 1 + \sum_{j=1}^p x_{ij} x_{i'j} \right)^d \quad (12)$$

The resulting classification function is then given by:

$$f(x) = \beta_0 + \sum_{i \in V} \alpha_i \cdot k(x, x_i) \quad (13)$$

where  $V$  is the set of indices corresponding to the support vectors, and  $\alpha_i$  are the learned coefficients [83].

Although the basic SVM algorithm is binary, it can be extended to handle multiple classes using a one-versus-all strategy. In this setup, one classifier is trained per class to distinguish that class from all others. During prediction, all classifiers evaluate the same instance, and the final class label is determined by the classifier that outputs the highest confidence score.

While modern machine learning tasks often rely on ensemble models or deep learning methods, SVM remains an optimal solution in environments with limited computational resources. SVM also has a strong robustness in noisy environments and shows resistance to overfitting, making it well-suited for real-world sensor data. Additionally, SVM is effective even with a relatively small number of input features.

#### 4.3.5 Aggressive maneuver correction

After detecting and isolating the segments that correspond to aggressive driving maneuvers, a dedicated correction strategy becomes necessary. These maneuvers introduce distortions into the sensor data, primarily affecting the specific axes along which the maneuver occurs.

Longitudinal maneuvers, such as acceleration and braking, predominantly influence the accelerometer signal along the  $X$ -axis. Conversely, lateral maneuvers such as turning primarily distort the gyroscope data along the  $Z$ -axis, due to increased yaw rotation during cornering. Thereby, the basic idea of the correction method is to exclude these affected axes during the duration of the maneuver. This method is simple, yet effective, as it builds upon prior steps in the pipeline to apply targeted corrections only when and where necessary. It is important to note that these affected features cannot be entirely discarded from the overall framework. Despite their sensitivity to aggressive maneuvers, they still contain valuable information under normal driving conditions. Therefore, the correction algorithm is applied selectively and locally, only to the specific time intervals flagged as problematic, ensuring that the system retains maximum information content while minimizing noise.

To determine which sensor axes should be suppressed for each maneuver class, a variable importance analysis was conducted. For this, the SHAP (SHapley Additive exPlanations) [84] method was employed, to quantify the contribution of each feature to the classification error when aggressive maneuvers occurs. SHAP offers a unified framework for interpreting machine learning models by quantifying the individual impact of each feature on a given

prediction, while accounting for all possible combinations of feature subsets. Specifically, the SHAP value represents the average marginal contribution of a feature across all feature.

One of the key advantages of SHAP is that the entire prediction can be decomposed into the sum of the individual SHAP values of all features, allowing their effects to be interpreted independently and aggregated without overlap. Additionally, the method is model-independent, meaning it does not rely on the internal structure or type of the predictive model. As a result, SHAP values provide consistent and stable interpretations, even if the model architecture or parameters change, as long as the underlying feature effects remain the same.

For each maneuver type, the SHAP analysis indicated which variables reduced classification performance the most. Based on this, a correction logic system was constructed in the form of a rule-based decision tree. The method identifies the maneuver-induced disturbances and applies a smooth and weighted downscaling of the vibration signal in the affected temporal regions. Maneuver events were previously detected, from which the index positions of these detected events is stored, representing the central timestamps of disturbance periods. To account for the onset and decay of the transient dynamics, symmetric temporal padding was applied around each event, using a fixed time window to extend the correction region both before and after each detected maneuver.

A binary correction mask was then constructed over the entire time axis, where padded regions surrounding each maneuver were marked as 1 (indicating affected areas), and all other regions were assigned 0 (unaffected). This initial binary representation defined the regions requiring signal correction.

The preprocessed IMU's acceleration  $a_x$  and  $a_y$  and the gyroscope's  $\omega_z$  and  $\omega_y$  axes were multiplied by the smoothed weighting function, according to the decision tree-based logic shown in Fig. 15. The result was a corrected signal in which the maneuver-affected portions were proportionally downscaled according to their temporal proximity to the detected events. Mathematically, the correction can be expressed as:

$$s^{\text{corrected}}(t) = s(t) \cdot w(t),$$

where  $s(t)$  is the original signal and  $w(t) \in [0, 1]$  is the Gaussian-weighted correction mask.

This rule-based correction method is computationally lightweight and easily integrable into

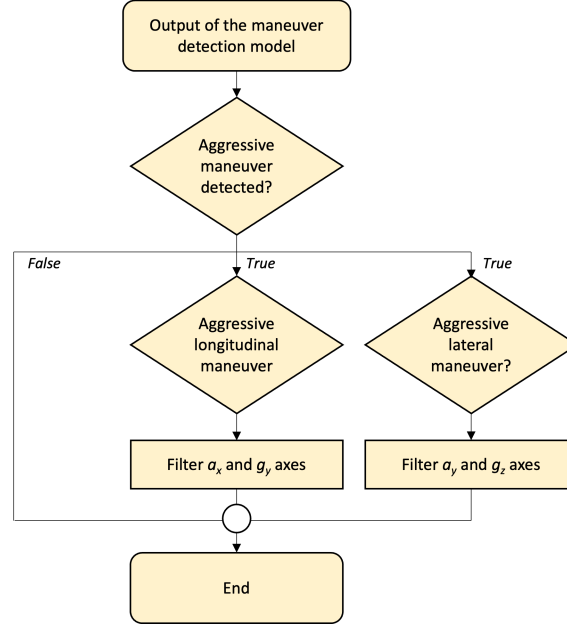


Figure 15: Decision tree of the correction method with variables to drop when an aggressive maneuver is detected

the existing signal processing pipeline. Lastly, the negative effects of aggressive maneuvers are reduced without removing valuable data.

#### 4.4 Feature extraction for pavement classification

In this subsection, the focus shifts from general data preprocessing toward feature extraction, introducing the methods applied to derive meaningful and informative variables from the raw sensor data.

To achieve this, several feature extraction and engineering methods were explored and evaluated. The most effective approaches identified through this process are:

1. frequency-based Power Spectrum Analysis,
2. aggregation of amplitude values along the frequency axis,
3. aggressive maneuver detection and correction,
4. time series segmentation,
5. and generation of new, synthetic features.

The following subsections describe each of these methods in detail, while the complete data preparation pipeline is illustrated in Fig. 16.



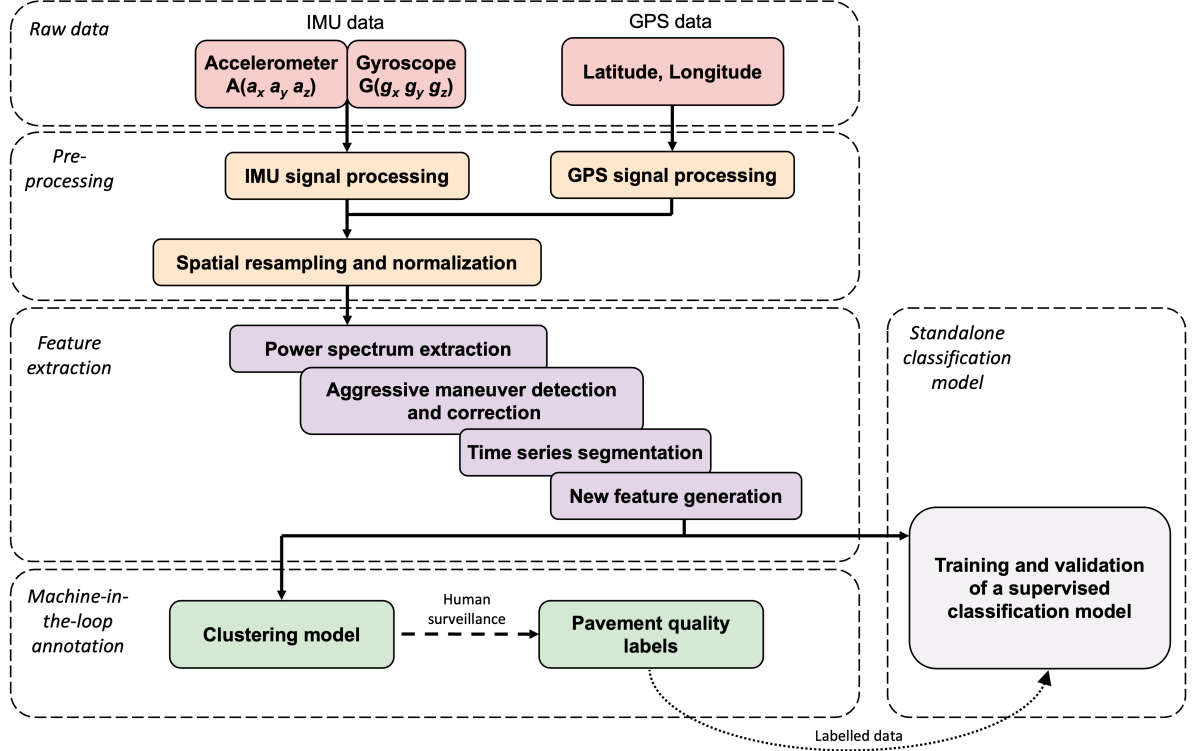


Figure 16: The detailed structure of the applied methodology

#### 4.4.1 Power Spectrum Analysis

To enhance the informativeness of the vibration dataset and support the development of robust machine learning models, the original time-domain signals were transformed into the frequency domain using Power Spectral Analysis. This transformation can be used in the identification of periodic components, dominant frequencies, and energy distribution patterns, which are not easily noticeable in the raw time-series data.

A well-established method for frequency transformation is the Discrete Fourier Transform (DFT), which converts a time-series signal into its constituent frequency components. For a discrete signal  $x = [x_0, x_1, \dots, x_{N-1}]$ , the Discrete Fourier Transform (DFT) is defined as follows [85, 86]:

$$X_k = \sum_{t=0}^{N-1} x_t e^{-j \frac{2\pi kt}{N}}, \quad (14)$$

where  $X_k$  denotes the complex-valued Fourier coefficient corresponding to the frequency bin  $k$ ,  $N$  is the total number of time-domain samples, and  $j$  is the imaginary unit. The magnitude squared of  $X_k$  represents the signal power at the corresponding frequency.

From the DFT, the power spectral density (PSD) function  $T(\omega)$  is derived as:

$$T(\omega) = \frac{1}{2\pi N} |X(\omega)|^2, \quad (15)$$

which quantifies the power contribution of each frequency component within the signal. To obtain the signal power over a given frequency interval  $[f_1, f_2]$ , the PSD is integrated as:

$$P = \int_{f_1}^{f_2} T(\omega) d\omega. \quad (16)$$

This approach facilitates the comparison of energy content across different frequency bands, which is particularly relevant for identifying mechanical or environmental signatures encoded in the vibration patterns.

To analyze how the frequency content evolves over space or time, a spectrogram was computed for each sensor axis. The spectrogram provides a two-dimensional representation of the signal's spectral density, typically as a function of either time or spatial distance  $d_{\text{spatial}}$  and frequency  $f$ . Let the spectrogram be denoted as  $T(d_{\text{spatial}}, f)$ , where  $S$  represents the power at a given spatial location and frequency.

To reduce the dimensionality of the data while retaining meaningful information, spectral values were aggregated along the frequency axis:

$$T_{\text{sum}}(d_{\text{spatial}}) = \sum_f T(d_{\text{spatial}}, f). \quad (17)$$

Here,  $T_{\text{sum}}(d_{\text{spatial}})$  denotes the total spectral energy at each spatial sampling point  $d_{\text{spatial}}$ . This step produces a compressed one-dimensional feature vector per signal segment, expressing the total vibration energy at each location. Aggregation like this is advantageous for subsequent machine learning tasks, as it reduces computational complexity and reduces the risk of overfitting, while still preserving signal characteristics relevant for classification tasks.

The frequency-based analysis is followed by the detection and correction of aggressive maneuvers in the pipeline, which I present in detail in the separate Subsection 4.3 due

to its volume. In the following subsection, I describe the final steps of feature extraction, segmentation and generation of new features.

#### 4.4.2 Time Series Segmentation and Feature Generation

**Time Series Segmentation** Following the correction of maneuver-affected segments, a time series segmentation process was applied to divide the dataset into internally homogeneous intervals. The aim of this step is to isolate connecting sections of the measurement where road surface quality remains statistically consistent. Rather than evaluating the signal at individual data points, the segmentation enables the system to assess road conditions over structured spatial units, improving both interpretability and model performance.

This task is formalized as a change point detection problem, where the objective is to identify points in the signal at which statistical properties, such as the mean or variance, show significant change. When the number of change points is unknown, the problem can be cast as a discrete optimization task defined by:

$$\min_{\mathcal{T}} V(\mathcal{T}) + \text{pen}(\mathcal{T}) \quad (18)$$

Here,  $\mathcal{T}$  denotes the set of change point locations,  $V(\mathcal{T})$  is a cost function representing the fit within segments, and  $\text{pen}(\mathcal{T})$  is a penalty term that constrains the complexity of the segmentation. This balance guarantees that the solution avoids both over-segmentation, what would result in too many short segments, and under-segmentation, which means merging dissimilar regions.

To solve this optimization efficiently, the Pruned Exact Linear Time (PELT) algorithm [87] was employed. PELT is designed to minimize the objective in Eq. 18 under the assumption of a linear penalty function. The algorithm achieves high accuracy in identifying change points by minimizing the within-segment cost while applying a penalty to the number of change points, thus discouraging excessive fragmentation of the signal. Through an intelligent pruning rule, PELT discards suboptimal candidates early in the process, resulting in linear computational complexity relative to the length of the time series, making it well-suited for large-scale datasets typical in road quality monitoring applications.

The result of applying PELT is a set of segments, each representing a spatial unit of the road with consistent vibration characteristics. These segments form the analytical foundation

for subsequent modeling steps and enable more efficient computation by operating at the segment level instead of point-wise resolution.

**New feature generation** Once segmentation is completed, statistical features are extracted from each segment to summarize its information content. This step serves both to reduce data volume and to highlight meaningful variations relevant to pavement condition classification.

For each segment, the median and standard deviation of all relevant sensor channels are calculated. These features were selected based on prior empirical findings, where they were shown to be effective in capturing central tendencies and dispersion, which are the two core indicators of surface uniformity and variability. The median is robust to outliers and better represents the typical vibration intensity in segments affected by sporadic disturbances, while the standard deviation captures the degree of irregularity in the signal.

Overall, segmentation and feature generation transform raw sensor time series into a structured dataset of manageable size, composed of statistically meaningful summaries of road surface sections.

#### 4.5 Machine learning-based clustering and classification

With the dataset cleaned, preprocessed, and segmented, machine learning methods can now be applied to enable the automated classification and categorization of road sections based on their measured vibration characteristics.

This subsection presents the ML pipeline applied for both the annotation of training data and the final supervised classification. In principle, supervised machine learning requires labeled data. However, one of the main challenges encountered during this study was the absence of high-resolution ground truth data. Public road authorities typically do not possess sufficiently detailed pavement condition data to serve as a direct reference for supervised model training.

In the Subsection 4.5.1, a general overview of the principles and key features of the annotation strategy is presented. This is followed by a detailed explanation of the clustering method and model type used in Section 4.5.2. Finally, the supervised learning model trained on the newly labeled dataset is described in Section 4.5.3.

#### 4.5.1 Machine-in-the-Loop annotation

Following the feature extraction process, the first step in the modeling pipeline is the annotation procedure to generate the ground truth labels. This was achieved using a Machine-in-the-Loop approach, which integrates unsupervised machine learning with human expert supervision. The approach is to use data-driven clustering methods to group the dataset into structurally coherent subsets, which are then manually verified and labeled based on domain knowledge.

The need for such an annotation strategy arises from the limitations of supervised learning. Although supervised models offer powerful classification capabilities, they require a big amount of high-quality labeled data, however, what is often unavailable in sufficient resolution in road surface monitoring. Traditional data sources, such as those maintained by road authorities, tend to provide only coarse or incomplete ground truth data, making them unsuitable for training fine-grained classifiers [30].

Manual labeling presents several practical challenges. First, it is highly labor-intensive and time-consuming due to the large volume of sensor data collected. Second, the labeling process introduces human subjectivity and inconsistency, especially when expert interpretation is required. Third, once a manually labeled dataset is created, it is typically inflexible and difficult to adapt to changing classification schemes, measurement standards, or expanded category sets. These limitations motivated the development of a more scalable and adaptive approach.

Machine-in-the-Loop annotation offers a solution to these challenges. MIL combines the exploratory capability of unsupervised learning with the judgment of human experts. In this paradigm, clustering algorithms, such as K-Means, DBSCAN, hierarchical clustering, or Gaussian Mixture Models are applied to the feature-transformed dataset to automatically partition the data into groups with similar statistical characteristics [88, 89]. These clusters represent latent structure in the data that may correspond to different road quality levels or surface conditions.

The theoretical foundation of MIL lies in semi-supervised and active learning principles, where machine learning assists the labeling process rather than relying on full manual annotation. MIL significantly reduces the annotation time while preserving label quality. It also allows for iterative refinement, meaning that new data or evolving standards can be incorporated with minimal rework.

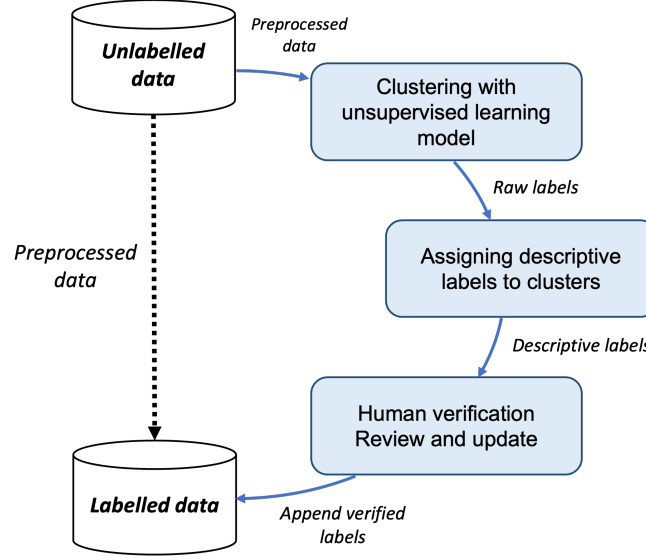


Figure 17: General workflow of the offline MIL labeling

In the present study, the preprocessed features served as the input for clustering. Once clusters were formed, they were reviewed by human supervisors. Since clustering algorithms assign arbitrary categorical identifiers to each group, human annotators interpreted and relabeled these groups based on their physical significance. The complete process is illustrated in Fig. 17, which outlines the key steps of the MIL annotation framework.

This methodology has already proven successful across various domains. In medical imaging, it aids radiologists in identifying patterns in diagnostic scans; in natural language processing, it helps annotate sentiment or topical information in large textual corpora; and in autonomous driving, it supports object recognition and scene classification. However, its application to road quality monitoring has not yet been explored according to our knowledge.

The main advantages of the Machine-in-the-Loop annotation approach are its efficiency and adaptability [88, 89]. By reducing the dependency on fully manual annotation, the method accelerates the labeling process and significantly reduces associated costs. It also enables the system to evolve with new data and revised classification schemes, making it a flexible tool for long-term deployments.

However, the method has some limitations. Clustering outcomes are highly sensitive to the selected features and algorithmic parameters. Poorly chosen features can lead to groupings that do not correspond to meaningful or useful surface conditions. Additionally, while MIL reduces manual effort, human input remains necessary, especially to interpret ambiguous

cases or validate the clustering. This introduces a degree of subjectivity and potential inconsistency that must be managed carefully by the subsequent human supervisors.

Despite these challenges, the MIL framework presents a scalable and practical solution for creating annotated datasets in domains where ground truth labels are scarce or difficult to obtain. In the context of road quality monitoring, it enables the creation of training data at a resolution that would be infeasible to generate through manual labeling alone.

After the MIL-based annotation was completed, the resulting dataset was updated to include the new road quality class labels. The final structure of the annotated input matrix is illustrated in Fig. 6, which shows the full set of features alongside the automatically generated pavement categories.

Table 6: The expected output matrix for Machine-In-the-Loop annotation

| Sample<br>road segment | Features                      |                               |                               |                                  |          | Class    |
|------------------------|-------------------------------|-------------------------------|-------------------------------|----------------------------------|----------|----------|
|                        | $a_x$<br>(m s <sup>-2</sup> ) | $a_y$<br>(m s <sup>-2</sup> ) | $a_z$<br>(m s <sup>-2</sup> ) | $\omega_x$<br>(s <sup>-1</sup> ) | ...      |          |
| 1                      | 0.67                          | 0.51                          | 1.18                          | 3.06                             | ...      | Good     |
| 2                      | 0.71                          | 0.65                          | 1.21                          | 2.98                             | ...      | Good     |
| 3                      | 0.62                          | 0.57                          | 1.16                          | 3.02                             | ...      | Perfect  |
| $\vdots$               | $\vdots$                      | $\vdots$                      | $\vdots$                      | $\vdots$                         | $\ddots$ | $\vdots$ |
| $n$                    | 0.51                          | 0.38                          | 0.91                          | 2.84                             | ...      | Poor     |

#### 4.5.2 Unsupervised machine learning models

In this subsection, the structure and technical implementation of the unsupervised model applied in the annotation process are presented in detail.

Statistical learning problems generally fall into one of two primary categories: supervised and unsupervised learning. In supervised learning, each observation  $x_i$  is associated with a known response variable  $y_i$ , and the goal is to fit a predictive model that captures the relationship between the input and output. By contrast, in unsupervised learning, only the predictors  $x_1, x_2, \dots, x_n$  are available, without any corresponding labels. This makes it impossible to directly optimize for prediction accuracy. Instead, the focus is on uncovering hidden structures and patterns within the data.

A successful clustering approach aims to maximize intra-cluster similarity while minimizing inter-cluster similarity, thereby revealing coherent subpopulations in the data [90]. In the context of road surface classification, unsupervised clustering offers a practical means of generating initial labels in the absence of high-resolution ground truth. Once these clusters

are reviewed and confirmed, the resulting dataset can be used to train a conventional supervised classification model.

Several clustering algorithms are commonly applied in practice, including K-Means, hierarchical clustering, and Gaussian Mixture Models (GMM). Each method relies on different assumptions and distance metrics to define similarity and group structure.

For the purposes of this study, the choice of model had to account for two key factors. First, the vibration-based features extracted from the sensor data follows a lognormal-like distribution, exhibiting significant amplitude variability across categories. Secondly, further normalization or transformation of these features would have risked suppressing important structural differences needed for accurate clustering.

Given these considerations, Gaussian Mixture Models were identified as the most appropriate solution. Gaussian Mixture Models are probabilistic models that assume the data is generated from a mixture of several Gaussian distributions, each representing a different cluster. Unlike K-Means, which assumes spherical clusters of equal variance, GMMs allow for elliptical clusters and variable covariance structures, making them more flexible for real-world data with heterogeneity in scale and shape. Unlike hard clustering methods such as K-Means, which assign each data point exclusively to one cluster, GMMs provide soft assignments, meaning that each observation  $\mathbf{x}_i$  is associated with a probability of belonging to each of the  $K$  clusters.

Formally, Gaussian Mixture Model models the input dataset — here, the segmented and aggregated feature set  $T_{\text{sum, segmented}}$  — as a weighted sum of  $K$  multivariate Gaussian components. Each component  $c \in \{1, \dots, C\}$  is parameterized by a mean vector  $\boldsymbol{\mu}_c$ , a covariance matrix  $\boldsymbol{\Sigma}_c$ , and a mixing coefficient  $\pi_c$ , which represents the prior probability of selecting component  $c$ .

The likelihood of a data point  $\mathbf{x}_i$  under the GMM is given by:

$$p(\mathbf{x}_i) = \sum_{c=1}^K \pi_c \cdot \mathcal{N}(\mathbf{x}_i \mid \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \quad (19)$$

where  $\mathcal{N}(\cdot)$  denotes the multivariate normal distribution. The overall objective is to estimate the parameters  $\{\pi_c, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c\}_{c=1}^K$  that maximize the likelihood of the observed data.

A simplified summary of the GMM clustering process is shown in Table 7. In the present



study, the clusters derived from GMM were interpreted as corresponding to distinct road surface quality categories. This interpretation was feasible because the feature extraction pipeline was explicitly designed to encode information relevant to surface roughness and vibration characteristics. This process continues until convergence, typically defined as a small change in the log-likelihood between iterations.

Table 7: Gaussian Mixture Model clustering algorithm

- 
1. Initialize  $K$  Gaussian components with random mean vectors, covariance matrices, and mixing coefficients.
  2. Iterate until convergence:
    - (a) **Expectation step:** Compute the probability of each data point  $\mathbf{x}_i$  belonging to each Gaussian component using the current model parameters.
    - (b) **Maximization step:** Update the model parameters (means, covariances, and mixing coefficients) by maximizing the likelihood of the data given the computed probabilities.
  3. Assign each data point  $\mathbf{x}_i$  to the Gaussian component with the highest probability, if a hard clustering assignment is set.
- 

The accuracy and reliability of the proposed labeling procedure were ensured through a combination of automated and manual validation steps. While clustering alone can offer a strong foundation for labeling, particularly when the underlying data structure is well-defined, human oversight remains essential in critical applications such as road quality classification. In our work, each stage of the MIL annotation process was manually reviewed to ensure consistency and interpretability. This included the visual inspection of clustering outputs, on-site verification where applicable, and the manual validation of the ordinal relationship among the identified road quality categories.

Such expert involvement allowed for correction of ambiguous clusterings, confirmation of meaningful category boundaries, and the reinforcement of domain-specific knowledge within the labeling process. Although the clustering algorithm provides the initial structure, the human revision ensures that the final labels not only reflect statistical similarity but also correspond to practical and actionable surface quality classes.

#### 4.5.3 Supervised machine learning classification

Following the annotation of the dataset, the final step in the pipeline is the development of a supervised machine learning model for the automated classification of road segments. At this stage, the preprocessed and segmented dataset can be used to train a classifier capable of generalizing the annotation logic and applying it to new, unseen data. In the

previous data preparation stages, continuous measurement signals were segmented into coherent spatial sections based on statistical homogeneity. These segments now serve as the fundamental units for classification, each to be assigned to one of the predefined pavement quality categories derived in the annotation phase.

A variety of supervised machine learning algorithms are available for solving classification problems, each with its own strengths, assumptions, and computational properties. In the development phase, the following supervised classifiers were evaluated during the model selection phase:

- K-Nearest Neighbors: A distance-based algorithm that assigns class labels based on the majority class among the closest training samples.
- Support Vector Classifier: A margin-based method that attempts to find the optimal separating hyperplane between classes in a high-dimensional space.
- Gradient Boosting: An ensemble technique that builds additive predictive models in a sequential manner, optimizing a loss function at each stage.
- Random Forest: A tree-based ensemble method that constructs multiple decision trees and aggregates their outputs for final classification.

After comparative analysis mainly based on classification accuracy and robustness, the Random Forest classifier was selected as the final model. RF showed consistent performance across different folds in cross-validation and demonstrated strong generalization ability on the test set.

Random Forest is an ensemble learning method that combines the predictions of multiple decision trees to improve classification accuracy and control overfitting. It belongs to the class of bagging (bootstrap aggregating) algorithms and is particularly effective when dealing with structured tabular data. The specific training steps of a Random Forest algorithm is shown in Table 8.

This ensemble approach provides several advantages. First, it offers robustness to noise and overfitting, as the use of multiple decision trees trained on different bootstrapped subsets of the data helps reduce the variance of the model without significantly increasing its bias. In addition, Random Forest handles high-dimensional data effectively by selecting a random subset of features at each split, allowing it to manage datasets that include irrelevant or

Table 8: Key steps of the training of a Random Forest algorithm

- 
1. **Initialize the ensemble:** Define the number of trees to be generated in the forest.
  2. **Tree construction** (repeated independently for each tree):
    - (a) Draw a random bootstrap sample from the training dataset.
    - (b) At each decision node, select a random subset of the features.
    - (c) Determine the best split among the selected features based on a splitting criterion (e.g., Gini impurity).
    - (d) Grow the tree to full depth without pruning.
  3. **Classification:** For a new input instance, each tree predicts a class label.
  4. **Aggregation:** The final output is determined by majority voting across all trees.
- 

weakly informative variables. However, Random Forest has certain limitations. It can be less effective on datasets with highly imbalanced classes without resampling or weighting techniques. Additionally, the model can become computationally expensive with a large number of trees or extremely large datasets, though this was not a limiting factor in the present study.

## 4.6 Performance measures

There is a wide range of performance metrics that can be used to evaluate the performance of different machine learning models. In the following, I first review the methods and metrics that has been used to evaluate the results of unsupervised clustering, followed by a discussion of supervised learning classification metrics.

### 4.6.1 Measuring the quality of clustering

As discussed in previous sections, evaluating unsupervised machine learning models presents a fundamental challenge, as there is no absolute ground truth available for validation. Unlike supervised models, clustering methods operate on statistical assumptions, and their quality must be inferred from indirect metrics.

Therefore, while no single objective metric exists to fully determine the correctness of a clustering model, several well-established evaluation criteria can provide meaningful insights into the structure and performance of the clustering. By employing multiple complementary metrics, the robustness, consistency, and plausibility of the model can be analyzed from different perspectives. These metrics also provide a solid basis for comparing competing models and selecting the most appropriate configuration for the annotation task.

A key aspect of clustering performance is the determination of the optimal number of

components, denoted  $K$ , in the Gaussian Mixture Model. Selecting too few clusters may lead to underfitting and the merging of distinct surface quality categories, while too many clusters can result in overfitting and fragmentation of homogeneous data regions. Hence, model evaluation in this context must balance model complexity and data representation fidelity.

Several statistical criteria were used to assess the quality of the clustering models across different values of  $K$ :

- **Bayesian Information Criterion (BIC):** BIC is a penalized likelihood criterion that balances goodness-of-fit with model complexity. It is defined as:

$$BIC = -2 \cdot \log(L) + p \cdot \log(n) \quad (20)$$

where  $L$  is the log-likelihood of the model,  $p$  is the number of estimated parameters, and  $n$  is the number of observations. Lower BIC values indicate a more parsimonious model with better generalization ability.

- **Akaike Information Criterion (AIC):** AIC follows a similar principle but applies a slightly different penalty term:

$$AIC = -2 \cdot \log(L) + 2p \quad (21)$$

Like BIC, lower AIC values suggest a better model. AIC tends to favor more complex models compared to BIC, making it useful for assessing overfitting risk.

- **Log-Likelihood:** The log-likelihood  $\log(L)$  reflects how well the model explains the observed data. Higher values indicate that the data is well represented by the estimated Gaussian components. However, log-likelihood alone cannot be used for model selection due to its monotonic increase with model complexity.
- **Silhouette Score:** The Silhouette score quantifies how well-separated the clusters are. For each sample, it compares the average intra-cluster distance  $a$  with the average nearest-cluster distance  $b$ . The Silhouette coefficient is defined as:

$$s = \frac{b - a}{\max(a, b)} \quad (22)$$

Values close to +1 indicate well-clustered data, values near 0 suggest overlapping

clusters, and negative values imply possible misclassification. The global Silhouette score is the mean of all sample-level coefficients.

The BIC and AIC criteria were applied to evaluate a series of GMM models trained with varying numbers of clusters. The model corresponding to the lowest BIC value was selected as the final configuration, as it offered a strong balance between model accuracy and simplicity. Additionally, Silhouette scores were used to visually assess cluster cohesion and separation using Silhouette charts, which plot the coefficient for each data point sorted by cluster.

**The Role of Human Validation.** In addition to the quantitative metrics discussed above, human validation played an important role in ensuring the quality and reliability of the clustering process. Although statistical measures provide valuable information, they cannot fully capture the practical relevance or correctness of the resulting groups. Furthermore, it is generally expected that good clustering results yield a well-performing subsequent classification model when the cluster assignments are used as labels.

To address these limitations and uncertainties, the clustering outcomes were manually reviewed at multiple stages. This included the visual inspection of cluster assignments, on-site verification of selected road segments, and human validation of the ordinal relationships implied by the cluster structure. In doing so, we confirmed that the data-driven clusters corresponded to meaningful and distinguishable categories of pavement quality.

Although a one-time human review introduces an additional step in the pipeline, it represents a significant improvement over traditional fully manual annotation workflows. The integration of human supervision within the Machine-in-the-Loop framework adds robustness and domain-specific accuracy without compromising scalability. The high performance of the final classifier reflects this joint contribution of unsupervised structure discovery and targeted human validation.

#### 4.6.2 Classification quality metrics

In contrast to unsupervised models, supervised learning allows for the use of explicit and objective metrics to assess model performance. This is possible because supervised classification tasks involve labeled ground truth data, enabling a direct comparison between the model’s predictions and the true class labels.

In the following, I present the key evaluation metrics used in this study to benchmark the

supervised machine learning model developed for road surface classification.

- **Accuracy:** Accuracy is the simplest and most intuitive metric, measuring the overall proportion of correctly classified instances across all classes. It gives a general indication of model performance, but can be misleading in the case of imbalanced class distributions.
- **F1 score:** The F1 score is a single metric that balances both false positives and false negatives. It is especially useful in multi-class classification problems or when class distributions are imbalanced, as it avoids overstating performance. A high F1 score indicates that the model is both accurate and sensitive in identifying the target class, making it a reliable metric for evaluating classification quality under realistic conditions.
- **ROC-AUC:** In multiclass classification settings, the Receiver Operating Characteristic Area Under Curve (ROC-AUC) score is a common measure for evaluating a classifier's ability to distinguish between different classes. ROC-AUC summarizes the trade-off between true positive rate and false positive rate across various decision thresholds. Higher AUC values indicate better discriminatory power. In this study, the one-vs-rest strategy was used to compute the ROC-AUC score for multiclass evaluation.
- **Confusion Matrix:** To gain deeper insight into model performance, a confusion matrix was also generated. Each row represents the actual class, and each column the predicted class.
  - The diagonal entries represent true positives — correct predictions for each class.
  - Off-diagonal elements in column correspond to false negatives for each class — instances that belong to the indicated class but were predicted as another class.
  - Off-diagonal elements in a row correspond to false positives — instances predicted to a class but actually belonging to a different class.

The matrix provides a detailed view of classification performance and helps identify patterns of misclassification, particularly in cases where class imbalance or overlapping features affect accuracy.

## 4.7 Reverse geocoding and map matching

At the end of this section, this subsection presents the implementation of reverse geocoding and map matching techniques. These procedures serve as important components for improving the interpretability, visualization, and manageability of location-based data. While the raw measurement dataset contains only GPS coordinates, such geographic points are difficult to analyze or search in their native form. Therefore, for both human interpretation and database integrity, the coordinates were enriched with spatial semantics using map-based processing.

The purpose of these steps is twofold. First, they assign real-world road network attributes and postal addresses to each measurement segment. Second, they allow for the elimination of redundancy in repeated measurements along identical routes and enhance the clarity of map-based visualizations. Fig. 18 provides a conceptual overview of this geospatial annotation pipeline.

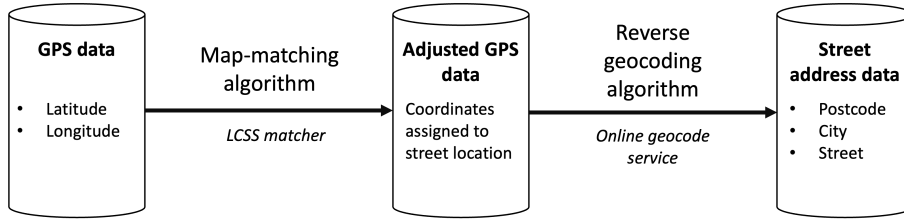


Figure 18: Conceptual diagram of geocoding and map matching processes

### 4.7.1 Theory and implementation of map matching

Map matching refers to the process of aligning a sequence of GPS coordinates with the most likely path on a digital road network. Given the inherent noise and sampling errors of GPS data particularly in urban environments, raw coordinates may not accurately follow the actual road infrastructure. Therefore, a robust map matching algorithm is necessary to reconstruct the true route and represent the trajectory as a continuous linestring object.

In my work, the Longest Common Subsequence (LCSS) algorithm [91] was used for map matching. The process begins by generating an approximate route from GPS points, which is then confined within a buffered geofence region to limit network complexity. Finally, a digital road graph is constructed from OpenStreetMap data.

The LCSS-based matcher is initialized on this constrained road graph and applied to the trace. It efficiently identifies the most plausible path by comparing the GPS trace to the road segments in a non-exact, order-preserving way, which is especially robust to missing or

noisy data points. The resulting matched path is returned as a special linestring geometry representing the true road segment traveled. A linestring object is a geometric data type used to represent a continuous sequence of points connected by straight lines, typically used to model paths, routes, or road segments in geographic information systems. It is defined by an ordered list of coordinate pairs, forming a linear geometry without self-intersections.

For visual and analytic purposes, the full matched linestring was then subdivided into smaller sections at regular, 2 meter intervals using a recursive splitting operation. These micro-segments enable high-resolution road quality visualization and simplify further spatial interpolation tasks. A map-based interactive visualisation can then be generated, with additional colour-coding to illustrate the quality classification of the road section. The input for this is provided by the classification model detailed previously.

#### **4.7.2 Address assignment from coordinates with Reverse Geocoding**

Reverse geocoding is the process of transforming geographic coordinates into a human-readable address or spatial descriptor [92]. While raw GPS points provide absolute position, reverse geocoding allows linking them to meaningful entities such as street names, road types, or postal codes.

The reverse geocoding process in this study used the Nominatim service from the OpenStreetMap database [93]. For each previously segmented linestring, the centroid was computed and passed to the geocoder to retrieve address metadata. This included attributes such as the object class, such as highway or residential way, and type, as for example primary road, and full textual representation.

Despite its advantages, the geospatial post-processing pipeline introduced specific technical challenges. First, discrepancies between GPS traces and road network topology required careful parameter tuning for geofence padding and matching tolerance. Second, duplicate trace detection had to be implemented to ensure database integrity across multiple recordings on the same segment, lastly, reverse geocoding remained dependent on external APIs. The results of the geocoding process were appended to the primary dataset, enriching each spatial segment with contextual address information.

#### **4.7.3 Structure of the final output of the algorithm**

The combination of map matching and reverse geocoding yields a structured geospatial dataset in which each row corresponds to a road segment characterized by both geometric



and semantic properties. The structure of the final output matrix is shown in Table 9.

Table 9: Structure of the geocoded and matched output dataset

| # | LineString Geometry | Raw Coordinates    | Street Type | Address                      | Pavement Quality |
|---|---------------------|--------------------|-------------|------------------------------|------------------|
| 0 | LINestring(...)     | (47.4978, 19.0399) | residential | 12 Göcsej Rd, Zalaegerszeg   | Good             |
| 1 | LINestring(...)     | (47.4941, 19.0401) | primary     | 14 Platán Rd, Zalaegerszeg   | Medium           |
| 2 | LINestring(...)     | (47.4963, 19.0403) | secondary   | 16 Balatoni Rd, Zalaegerszeg | Poor             |

## 5 Results and discussion

Following the detailed presentation of the system development steps, this section presents the key results and insights derived from the proposed road quality monitoring framework. The discussion includes the performance and evaluation of specific subsystems of the data processing pipeline, such as maneuver detection and Machine-In-the-Loop annotation.

### 5.1 Measurement conditions

During the development and testing phase, a series of controlled measurements were conducted on public road sections of varying pavement quality. To ensure robust training and evaluation under diverse input conditions, two vehicles with significantly different suspension architectures and body dynamics were employed. The first test vehicle was a 2019 Nissan Leaf, a modern electric hatchback equipped with a torsion beam rear suspension with trailing arms and separated spring and damper elements. The second was a 1997 BMW 318is, a compact sedan with a fully independent multilink suspension and stiffer structural components. The pronounced difference in suspension stiffness, damping behavior, and chassis response provided an ideal comparative basis for evaluating the robustness of the signal processing and machine learning models.

The measurements were carried out on a predefined urban and suburban route within the public road network, including both newly paved and highly degraded asphalt segments.

For the maneuver detection model, it was critical to collect data from both normal and aggressive driving sections to allow the recognition of corresponding vibration patterns. As the vibrations resulting from road irregularities and vehicle dynamics are intertwined in real conditions, the data must include both influences to reflect reality. To support this, a dedicated training dataset was recorded specifically for the maneuver detection algorithm. In this dataset, aggressive maneuvers were manually annotated with their corresponding locations and types. Special care was taken to include balanced proportions of lateral and longitudinal maneuvers. For this reason, shorter route segments were used to emphasize controlled and diverse motion profiles.

The route defined for the maneuver detection model training dataset covered a total of 31 kilometers, consisting of approximately 70% straight and 30% curved sections. It incorporated a variety of intersections and road features to reflect real-world urban driving, including light-controlled intersections, T-junctions, pedestrian crossings, and railway crossings.

An overview of the measured event types and their occurrence across both test vehicles is summarized in Table 10.

Table 10: Total number of detected events for each vehicle and the maximum values of the main sensor axes per maneuver class

| Vehicle   | Maneuver class | Num of events |
|-----------|----------------|---------------|
| Hatchback | Accelerate     | 100           |
|           | Brake          | 100           |
|           | Right turn     | 60            |
|           | Left turn      | 60            |
| Sedan     | Accelerate     | 70            |
|           | Brake          | 70            |
|           | Right turn     | 40            |
|           | Left turn      | 40            |
| All       |                | 540           |

Following the training of the maneuver detection algorithm, significantly longer and more generalized datasets were used for the training and evaluation of the road surface classification algorithm. Unlike the maneuver detection module, only supervision was necessary from human side, as no initial manual labeling was required once the Machine-In-the-Loop annotation and unsupervised learning pipeline were finalized.

The total length of road segments recorded for training the unsupervised annotation algorithm was 82 kilometers. After the model was finalized, an additional 267 kilometers of road network were used exclusively for testing and validating the model’s generalization capability on unseen data. To better assess the system’s reliability and robustness, the validation routes differed entirely from the training sets. This separation allowed for an accurate evaluation of the model’s behavior in unfamiliar environments and under varying measurement conditions.

All measurements were conducted in Hungary, primarily within the city of Zalaegerszeg, nearby towns, and rural areas across western Hungary. The measurement datasets presented in this study and used to train and evaluate the algorithms were recorded between June 2024 and October 2024.

Fig. 19 presents an example of raw sensor output recorded during a short section of measurement. The plot shows the vibration signals obtained from the IMU sensor mounted inside the cabin, as well as the two piezoelectric accelerometers installed on the left and right front suspension arms. A clear correlation is visible between the IMU and axle-mounted sensor outputs, but also several distinctions are evident. The cabin accelerometer generally

exhibits a dampened response, with lower amplitude vibrations. In contrast, the piezoelectric sensors on the suspension arms detect more pronounced spikes and higher-frequency oscillations.

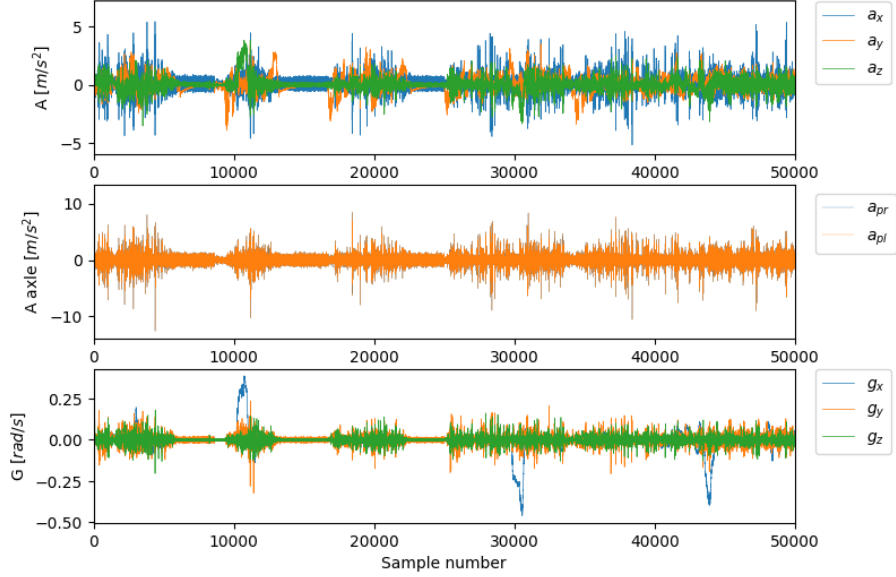


Figure 19: Example of raw sensor output from a short road segment

## 5.2 Data preprocessing in practice

The first step of the preprocessing involved the cleaning and preparation of the raw sensor data, as described in Section 4. To generalize the learning process and prevent overfitting to vehicle-specific characteristics, the vibration datasets from both test vehicles were merged and used jointly for the subsequent clustering and MIL-based annotation process. The preprocessing pipeline started with the removal of sensor errors, followed by synchronization of the sampling rates across all sensors. After temporal alignment, spatial resampling was applied.

As the first step in feature extraction, Power Spectral Density (PSD) was calculated for all acceleration signals. PSD was computed using a fixed segment size of 50 samples with 30 sample overlap to account for the continuous nature of the signal. Fig. 20 shows the spatial frequency spectrogram for the vertical acceleration signal of the IMU sensor. Higher amplitude values within the spectrum indicate areas of degraded pavement. The spectrogram clearly distinguishes high-energy, defect-prone sections from low-energy, smooth segments.

To reduce data dimensionality and simplify clustering, the amplitude values of the power spectrum were summed per data point. Fig. 21 illustrates the summed amplitude values for each sample during a short measurement section. The plot shows clear correspon-

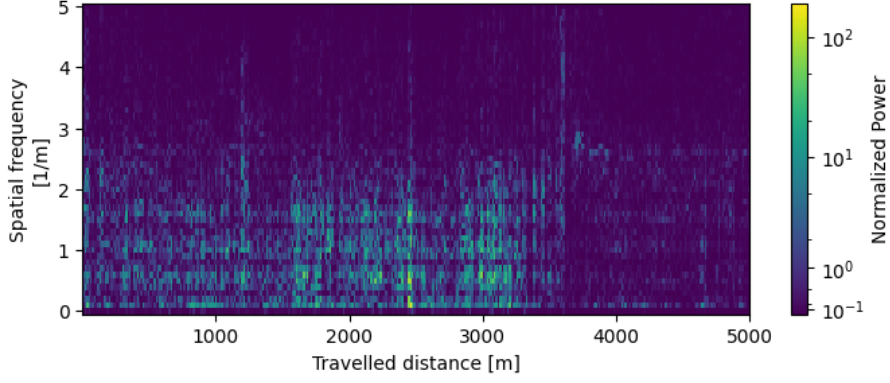


Figure 20: Power spectrum of the spatial resampled data for the IMU's Z-axis accelerometer

dence between the sensor output and road surface variations. It also highlights that each accelerometer axis contributes different information.

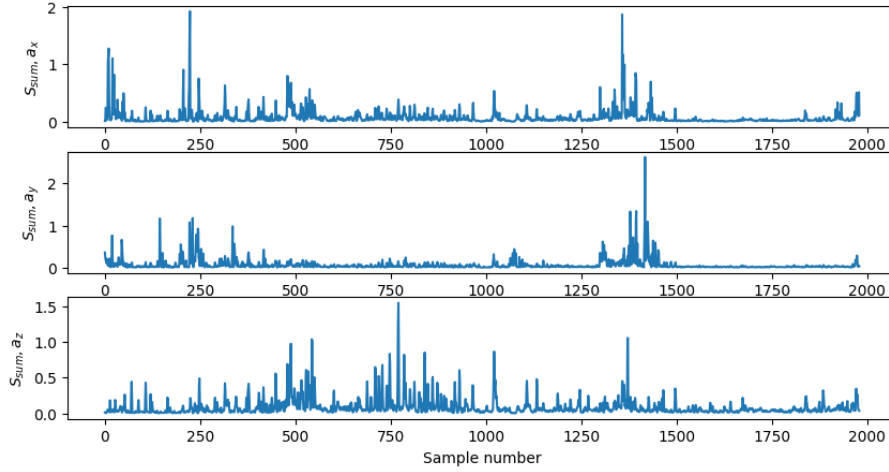


Figure 21: Summed amplitude values of the power spectrum at each sample point for the IMU's accelerometer sensor

The next step involved identifying internally homogeneous road segments based on the vibration characteristics, using the PELT changepoint detection algorithm. The segmentation was performed across all relevant sensor variables using a fixed penalty value of 0.01, with a balance between oversensitive segmentation and coarse grouping. The resulting average was approximately 32.9 segments per kilometer of road, although actual segment length varied based on surface condition.

The penalty value, and thus the defined segment properties, can be easily modified in the future according to customer needs, allowing for a more comprehensive and general segmentation or an even more detailed segmentation focusing on minor changes. Segment boundary accuracy was verified manually through random visual inspections and on-site assessments. The final preprocessing step consisted of extracting the median and standard

deviation of each vibration feature summary statistics for each segment.

### 5.3 Maneuver detection and correction

As previously discussed, the training of the maneuver detection model required a separate and carefully constructed dataset that explicitly included aggressive driving maneuvers. After recording the measurements, the dataset was manually annotated to identify and label these maneuvers based on predefined threshold values and the maneuver categories described in earlier sections.

#### 5.3.1 Importance of maneuvers

After summarizing the measurement conditions, I now examine the role and impact of aggressive maneuvers on the accuracy of vibration-based road quality assessment. While some studies [63, 64] have proposed vehicle-mounted inertial sensor frameworks, the specific effects of sudden driving maneuvers on signal integrity and classification reliability have not been extensively addressed, according to our knowledge, and based on the careful literature review process. Therefore, this subsection highlights these limitations.

It is important to emphasize that aggressive maneuvers can occur not only at curves but also on straight road segments, driven by unpredictable changes in surrounding traffic conditions. Since the proposed measurement system is designed to be integrated into any ordinary passenger vehicle, including unmarked, non-specialized vehicles, it does not enjoy priority treatment in traffic. Consequently, sudden accelerations, braking, and steering maneuvers are sometimes inevitable.

Related to this, Fig. 26 shows a visual representation of detected aggressive maneuvers using the developed algorithm. The left panel illustrates a city route with medium-intensity traffic, while the right panel depicts a measurement outside a residential area under similar traffic conditions. Colored markers represent different maneuver types detected along the route. On the urban route, turning maneuvers were primarily identified at intersections and small roundabouts. However, notable turning events were also detected during sudden actions taken to avoid road hazards on straight road sections. The dominant maneuver category, however, remains longitudinal, induced by changing traffic conditions. These could not be fully avoided during the measurements and occurred throughout the test routes. The same trend was observed in the suburban area, where although fewer turns were recorded, unexpected braking and acceleration events still occurred due to general

traffic dynamics. In summary, while cornering remains the primary source of aggressive maneuver patterns, longitudinal events such as rapid accelerations and decelerations pose equally significant challenges.



Figure 22: Measurement and classification results in urban (a) and suburban (b) environments with medium traffic intensity, for the Sedan vehicle

To investigate the negative influence of aggressive maneuvers, I isolated and compared two distinct types of road sections: (1) straight segments recorded under smooth, uniform vehicle motion, and (2) sharply curved segments where maneuver-induced dynamics were more pronounced. This separation provided a clear and interpretable contrast for analyzing the impact of vehicle behavior on signal quality and model output. For this comparison, variable importance values were computed using the SHAP method, using the Random Forest classification model introduced below. This allowed me to assess how different input variables influenced the predicted pavement quality category across various road geometries.

Fig. 23 presents the mean absolute SHAP values calculated separately for straight and curved segments. Each bar represents the relative importance of a given variable in shaping the model's output for the specified segment type. As can be observed on straight road sections, vertical acceleration values measured at the suspension arms and cabin play a dominant role in the classification process. These components directly reflect the road-surface-induced vibrations and are thus reliable indicators of pavement condition in the absence of external disturbances. In contrast, during curved maneuvers, the most influential feature becomes the angular velocity around the vertical axis, which exhibits significant variation due to the rotational motion of the vehicle. This shift in importance

reduces the influence of features that are normally relevant for road quality assessment, such as vertical acceleration, leading to a distortion of the model’s prediction. This phenomenon illustrates how the introduction of aggressive maneuvers skews the feature space and weakens the model’s interpretability and accuracy.

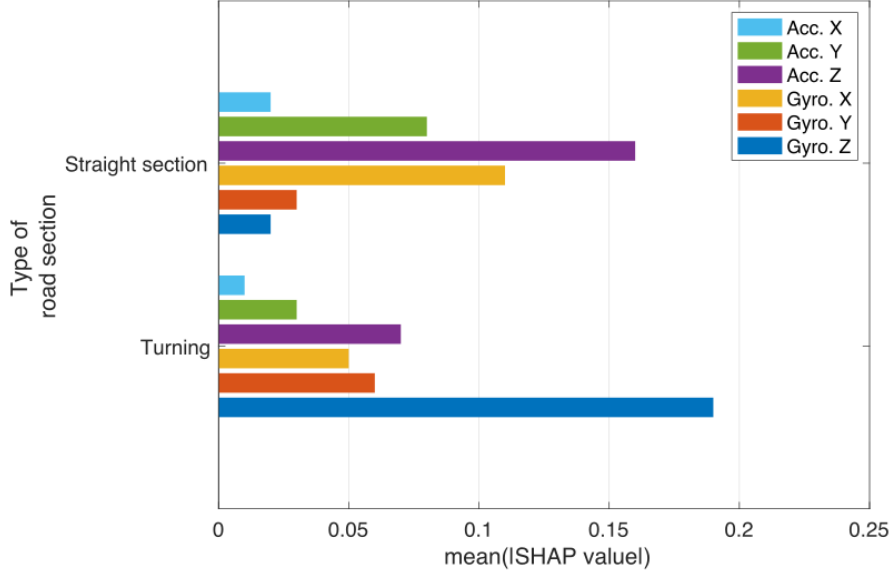


Figure 23: Aggregated absolute average SHAP feature importance for straight and curved sections.

Based on the above observations, the detection and subsequent signal correction are important to ensure the reliability of vibration-based road condition assessment across all traffic and road scenarios, as such disturbances may degrade the measurement quality at any time and in any traffic situation.

### 5.3.2 Evaluation of the detection algorithm

Following the identification of the significance and occurrence patterns of aggressive maneuvers, this subsection presents the results of the developed detection and correction algorithm.

As previously described, the dimensionality of the input dataset was first reduced using Principal Component Analysis. The PCA results indicated that the first two principal components captured 95.2% of the total variance, as summarized in Table 11.

Next, segmentation boundaries were determined using a change point detection algorithm. The algorithm effectively localized transitions between maneuver and non-maneuver sections. The accuracy of the change point detection is summarized in Table 12, showing low



Table 11: Variance explained by the principal components

| Principal component | Variance explained [%] |
|---------------------|------------------------|
| 1 and 2             | 95.2                   |
| 3                   | 2.1                    |
| 4                   | 1.9                    |
| 5                   | 0.5                    |
| 6                   | 0.3                    |

Mean Absolute Error and Root Mean Squared Error across all maneuver categories.

Table 12: Final results of the change point detection algorithm

| Maneuver type | MAE    | RMSE   |
|---------------|--------|--------|
| Acceleration  | 0.45 s | 1.11 s |
| Brake         | 0.36 s | 0.95 s |
| Right turn    | 0.68 s | 1.17 s |
| Left turn     | 0.79 s | 1.09 s |

After segmentation, the dimensionality was further reduced by computing statistical features for each segment. Features such as trimmed mean and peak values were selected to describe the segment characteristics. To mitigate class imbalance, the SMOTE was applied to synthetically augment the minority classes, generating new instances based on nearest neighbors. During the oversampling process, 150% of new observations have been synthesized using a randomly picked of the 5 nearest neighbors and the observation itself. In this case, each observation is used as a basis 1-2 times, which are chosen at random. The distribution of the feature space after SMOTE is shown in Fig. 24 as a function of the most significant features, the trimmed mean and peak value.

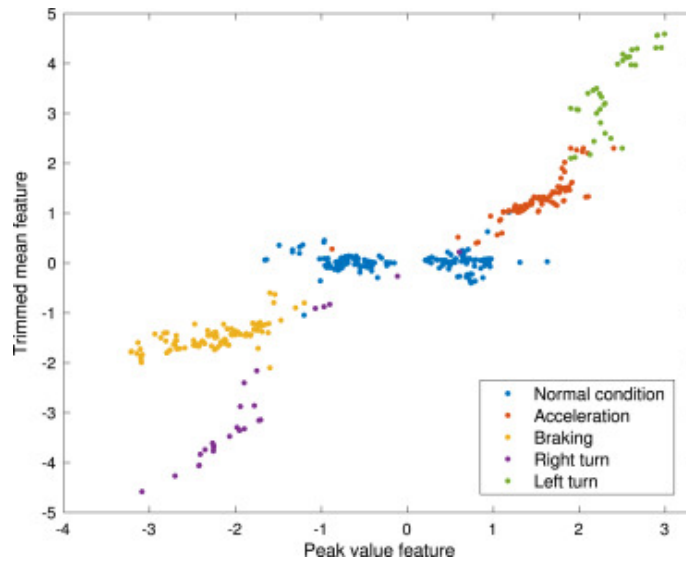


Figure 24: Feature space following the application of SMOTE oversampling method

To ensure generalization, the dataset was split into training and test sets with a 70/30 ratio using holdout validation. The multi-class Support Vector Machine classifier was trained on the training data, and hyperparameters were optimized using grid search with cross-validation. The hyperparameter settings of the SVM model are given in Table 13.

Table 13: Final hyperparameter settings of the multi-class Support Vector Machine classification model

| Hyperparameter           | Final value       |
|--------------------------|-------------------|
| Regularization parameter | 1                 |
| Kernel type              | Polynomial kernel |
| Gamma                    | 300               |
| Kernel degree            | 3                 |

The classification errors per maneuver category and per test vehicle are presented in Fig. 25. The results show higher accuracy for lateral maneuvers, with the right turn class achieving the lowest error. Aggressive longitudinal maneuvers, such as sudden acceleration and braking, yielded slightly higher error rates, especially for the hatchback, what could be due to the softer suspension characteristics. The mean classification error was 7.3% for the Nissan Leaf and 6.7% for the BMW 318is. Overall average classification accuracy across all maneuver types was 93%.

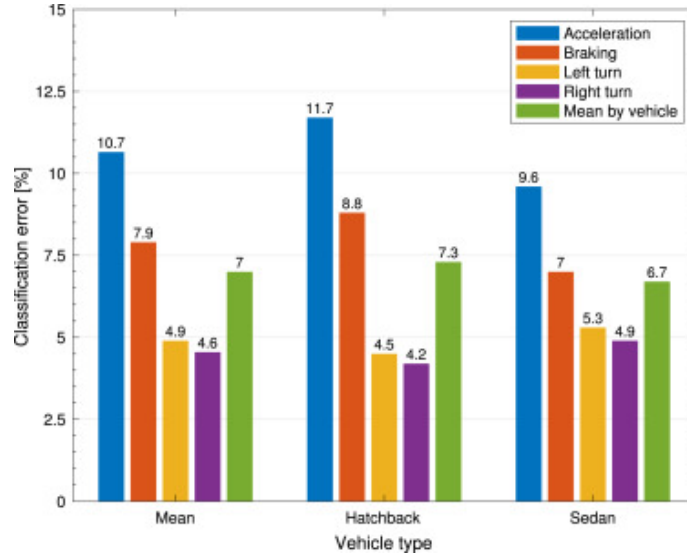


Figure 25: Classification error per test vehicle and by maneuver class

An example output of the maneuver detection algorithm is shown in Fig. 26. Along the Y-axis, the trimmed mean feature is plotted, while red shaded bands denote identified maneuver locations along the X-axis. These maneuvers occurred with varying intensity and duration, predominantly in urban traffic conditions where frequent and unpredictable speed changes are common.

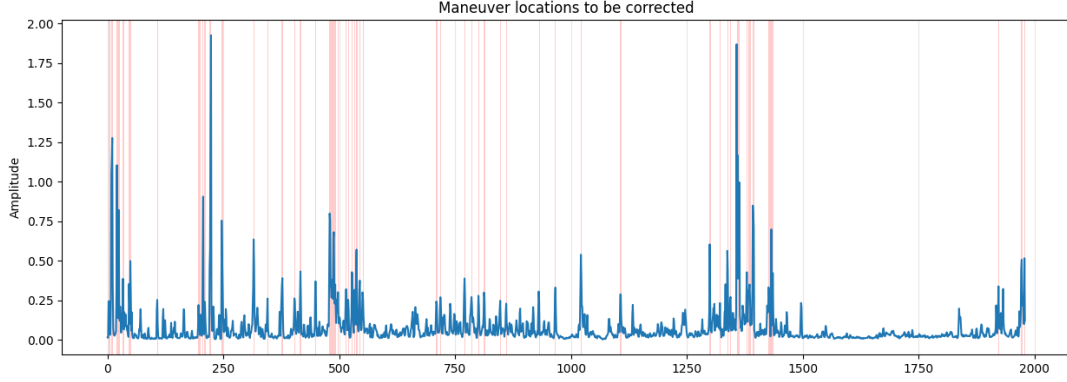


Figure 26: Detected aggressive maneuver sections (red) along with the trimmed mean feature

Vertical acceleration spikes also occurred independently during normal driving, especially on curved sections. Therefore, to avoid misleading the road quality classifier, the labeled maneuver locations were used to selectively correct only the affected variables. The correction algorithm was implemented using the previously described downscaling method on Subsection 4.3.5.

#### 5.4 Data annotation and road unevenness estimation

Following the development and validation of the maneuver detection and correction framework, I now present the results of the core algorithm for road surface quality evaluation. After the preparation phase, to define road surface quality classes without manual labels, the Gaussian Mixture Model clustering algorithm was applied. The number of clusters was selected based on a combination of algorithmic metrics, domain relevance, and practical interpretability.

To determine the optimal number of clusters, I evaluated a range between 2 and 9. From a technical perspective, model evaluation metrics such as the BIC, AIC, Log-Likelihood score, and Silhouette score were considered to ensure that the selected model remained within optimal bounds. These quantitative metrics were complemented by subjective factors reflecting real-world applicability. Specifically, road management companies often segment road quality into 3 to 7 classes [17]. The clusters should not only reflect statistical separability but also be interpretable and actionable by the maintenance authorities. In some cases, overly granular segmentation may lead to difficulties in interpretation and operational use, especially when only a general overview of the road network condition is required. Furthermore, the usefulness and information content of the resulting clusters were also evaluated. The primary aim was to ensure that the formed groups provide a

meaningful differentiation between road conditions.

Before fitting the GMM model, input features were weighted individually based on their prior information content. According to our prior experiments and feature relevance analysis, the most critical indicators of road quality were the Z-axis acceleration and the angular velocity components around the X and Y axes. Secondary indicators included the X and Y axis accelerations and angular velocities. The final feature weights are listed in Table 14. Additional GMM hyperparameters were also fine-tuned, including the covariance type, the initialization method, and the maximum number of iterations.

Table 14: Final hyperparameter settings and weights of the GMM clustering algorithm

| Hyperparameter                           | Final value                                                                              |
|------------------------------------------|------------------------------------------------------------------------------------------|
| Covariance type and tolerance            | 'full' and $10^{-4}$                                                                     |
| Initialization Method                    | 'kmeans'                                                                                 |
| Maximum Number of Iterations             | 300                                                                                      |
| Regularization Parameter                 | $10^{-6}$                                                                                |
| Standard deviation based feature weights | $a_x=4, a_y=4, a_z=9$<br>$a_{pr}=6, a_{pl}=6$<br>$\omega_x=7, \omega_y=7, \omega_z=2$    |
| Median based feature weights             | $a_x=5, a_y=5, a_z=10$<br>$a_{pr}=10, a_{pl}=10$<br>$\omega_x=8, \omega_y=8, \omega_z=2$ |

To assess the optimal number of clusters  $K$ , the BIC, AIC, Silhouette score, and Log-Likelihood metrics were plotted against increasing values of  $K$ . The elbow plot shown in Fig. 27 illustrates that both BIC and AIC decrease as  $K$  increases, suggesting improved model fit. However, this improvement plateaus beyond  $K = 5$  to 6. Simultaneously, the Silhouette score remains the highest between  $K = 3$  and 5, after which it declines sharply. This indicates that cluster quality and interpretability decrease when  $K$  exceeds 5. Ultimately, a balance between model complexity and clustering robustness was achieved in the  $K = 3$  to  $K = 5$  range. The final evaluation metrics for these cluster numbers are summarized in Table 15.

Table 15: Evaluation scores of the GMM unsupervised learning model

| Type of evaluation       | Number of clusters |         |         |
|--------------------------|--------------------|---------|---------|
|                          | $K = 3$            | $K = 4$ | $K = 5$ |
| BIC                      | -19875             | -25273  | -25675  |
| AIC                      | -20080             | -25549  | -26022  |
| Average Silhouette score | 0.53               | 0.49    | 0.50    |
| Log-Likelihood score     | 10069              | 12813   | 13060   |

Once the optimal number of clusters was selected, a final relabeling process was required

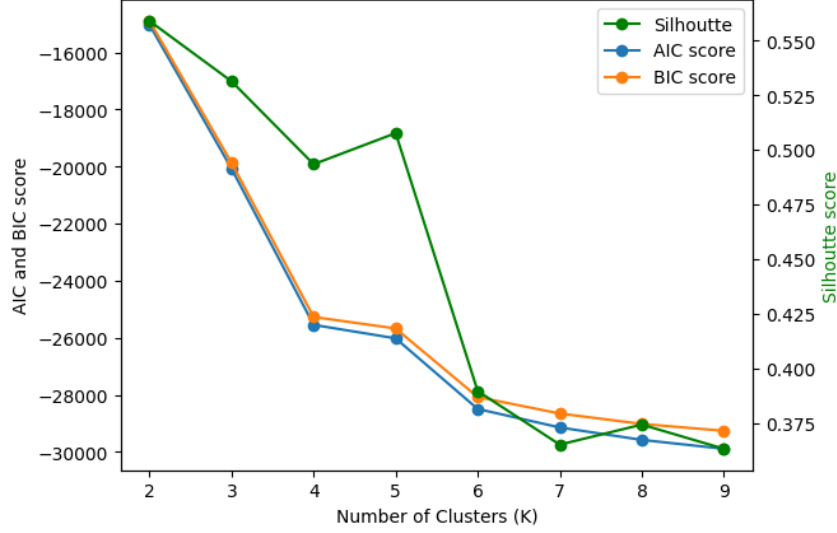


Figure 27: Elbow plot of clustering AIC, BIC and Silhouette metrics as a function of the number of clusters

to sort the randomly assigned cluster labels into an ordinal sequence based on road surface quality. This alignment was achieved by ranking the clusters according to the median Z-axis acceleration of the IMU and suspension sensors, which most directly reflect the vertical dynamics associated with road surface degradation [6, 51, 52]. The category with the highest median Z-acceleration was assigned to label 1 (worst quality), and subsequent labels were ordered accordingly. This alignment was also verified manually using visual observations and camera imagery.

The confidence of cluster assignments was further illustrated using a cumulative probability distribution, shown in Fig. 28. When  $K = 4$  clusters were selected, it was found that approximately 80% of all data points had a confidence score above 90%, confirming that the clusters were generally well-separated. However, uncertainty levels were higher in clusters labeled 2 and 3, which represent intermediate road quality conditions. This distribution reflects the inherent difficulty in categorizing roads that are partially degraded or undergoing transition.

Spatial evaluation of maximum probability assignments across the dataset was visualized on a road segment, as shown in Fig. 29. The figure highlights three distinct regions with varying clustering behaviors. Region 1 contains predominantly light yellow and green sections, indicating high-confidence cluster assignments exceeding 80%. This area corresponds to a nearly continuous stretch of newly paved, uniformly smooth road surface, validating both the clustering accuracy and its associated confidence. In Region 2, the probability

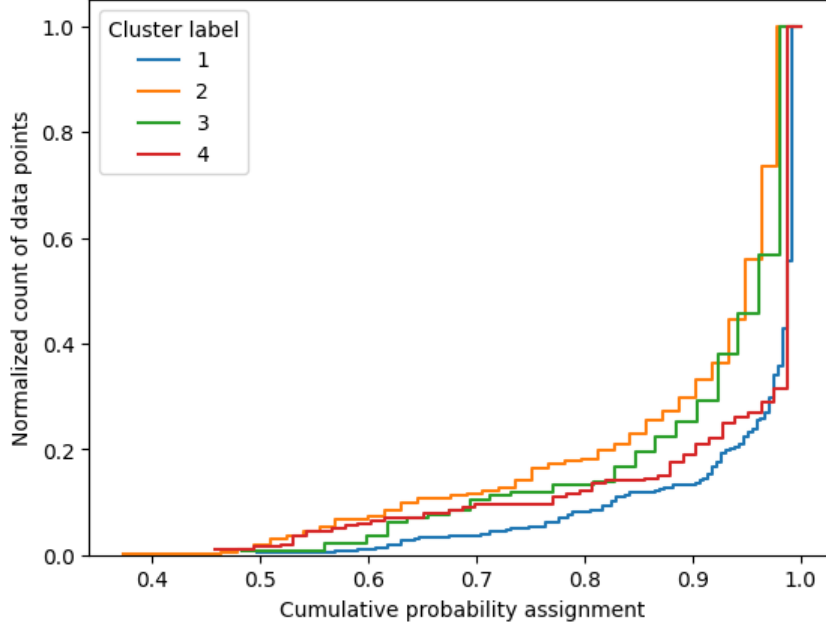


Figure 28: Cumulative probability assignments for each cluster label, if  $K = 4$

assignments fluctuate between 80% and 90%, reflecting a road surface with alternating degraded and recently repaired segments. Despite the surface variability, manual validation confirmed that the clustering remained consistent and reliable in this section. In contrast, Region 3 exhibits significantly lower confidence, with maximum assignment probabilities around 50%. At the time of data collection, this segment was undergoing roadworks, with disrupted pavement and temporary surface materials, likely contributing to the observed local uncertainty in clustering. This specific area was remeasured in subsequent sessions to confirm the initial results.

Finally, to ensure the consistency and credibility of the clustering process, the results were thoroughly reviewed through manual inspection. Visual observations and camera imagery were used to confirm the accuracy of label assignments. In ambiguous cases, in-person site visits were conducted to verify surface conditions. Although a few segments exhibited uncertain cluster labels due to transitional or recently altered pavement states, the majority of segments demonstrated high classification quality. Ultimately, the clustering results were judged accurate and robust, and the labels were preserved unchanged for the subsequent supervised learning phase. The final labeled dataset was saved into a new CSV file to serve as the input for training the supervised road quality estimation model.

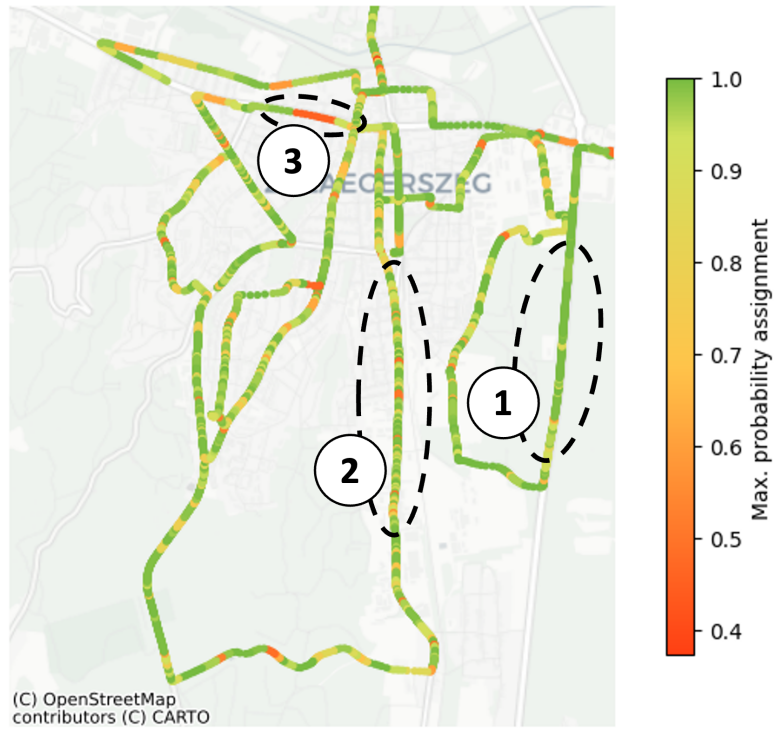


Figure 29: Map visualization of the maximum probability assignment for suburban measurement section

### 5.5 Supervised classification model with labeled data

Following the completion of the MIL-based annotation process, the supervised classification model was trained to evaluate the reliability and usability of the clustering-derived labels. The aim of this phase was to determine whether a standalone, production-ready supervised model could replicate and generalize the patterns identified during the unsupervised clustering. In addition to visual and manual validation, this process provides a quantitative measure of the annotation quality by comparing how well a supervised learner can recapture the clustering-defined class structure.

To carry out this evaluation, the fully preprocessed and labeled dataset was randomly split into training and testing subsets using a 70/30 ratio, such as in the maneuver detection using Support Vector Machine. Furthermore, data from both test vehicles were merged in the dataset to enhance generalization and ensure that the model is not overfitted to a specific vehicle. The train/test split was performed after merging the measurement data of the two vehicles, so both vehicles had an equal chance of being included in the training dataset.

During the model development phase, several candidate classifiers were tested in their

default configurations. These included both simpler linear models and more advanced ensemble methods. The evaluation metrics included overall classification accuracy and macro-averaged F1 score. To ensure comparability with later deployment conditions, the model selection and training were based on the combined cabin and suspension vibration data, with maneuver-corrected features applied throughout. Table 16 summarizes the comparative results of the tested models.

Table 16: Comparison of classification models based on weighted average metrics

| <b>Model</b>              | <b>Accuracy</b> | <b>F1-score</b> |
|---------------------------|-----------------|-----------------|
| Logistic Regression       | 0.79            | 0.79            |
| K-Nearest Neighbors       | 0.82            | 0.82            |
| Support Vector Machine    | 0.84            | 0.84            |
| Decision Tree             | 0.79            | 0.80            |
| Random Forest             | 0.89            | 0.90            |
| Gradient Boosting Machine | 0.87            | 0.88            |

The results demonstrate that simpler models such as Logistic Regression and Decision Tree classifiers provided the lowest accuracy and F1 scores, as expected due to their limited representational capacity. Conversely, all more advanced models, such as Support Vector Machines, Gradient Boosting, and Random Forest achieved significantly better results, all exceeding 80% accuracy. Among these, the Random Forest classifier consistently outperformed the others, achieving the highest accuracy with 89.3%, and macro F1 score with 90.1%.

Given its good performance, along with its robustness and interpretability, the Random Forest model was selected as the final supervised classification model and for the further evaluation and investigation steps. Following the model selection, further hyperparameter optimization and production model development were performed as described in the subsequent sections.

## 5.6 Evaluation of the effects of sensor location and maneuver correction

Following the discussion of clustering and supervised model, this section discusses the evaluation of specific research questions posed in Subsection 3.1, focusing particularly on the effects of sensor placement and maneuver correction on the overall performance and accuracy of the road surface classification model.



### 5.6.1 Approach of the investigation and evaluation

This subsection first introduces the methodology and statistical framework used to conduct the evaluation. The primary basis for comparisons is the set of performance metrics derived from the supervised learning model. For consistency and optimal accuracy, the Random Forest classifier selected previously was used as the baseline model throughout all comparisons, as it consistently yielded the best validation scores across the experiments.

To ensure a controlled evaluation environment, each configuration under examination was trained independently with the same preprocessing pipeline and raw measurement data. Only the examined variables were altered across the configurations, preserving identical training conditions otherwise.

The effects of maneuver correction and sensor placement were first investigated individually, with isolated assessments conducted for each. These were followed by a combined evaluation to examine their joint effects on classification accuracy and feature behavior. For the analysis of these isolated effects, statistical methods were employed to quantify the magnitude of impact, expected differences, and confidence intervals. Once the most optimal combination of maneuver correction and sensor placement was identified, additional model-based diagnostic tools were used to analyze internal changes. For the statistical tests, I applied statistical techniques including bootstrapped confidence intervals and non-parametric effect size estimators, to determine whether the improvements observed were genuine or merely the result of random variation.

The following steps describe the applied statistical methods.

**Bootstrapping for Confidence Intervals** To statistically characterize the model's performance and quantify the uncertainty in classification accuracy, I employed a bootstrap-based confidence interval estimation method. Given a dataset of size  $N$ , the accuracy and macro-F1 metrics can be obtained. Through bootstrapping, the 95% confidence intervals (CIs) for these metrics can be obtained.

To estimate the confidence interval of any *metric*, the bootstrap resampling was performed as follows:

1. For  $b = 1, \dots, B$  (with  $B = 1000$  iterations),  $N$  instances was randomly sampled with replacement from the original dataset index set  $\{1, \dots, N\}$ .

2. The accuracy  $metric^{(b)}$  is computed on each bootstrap sample.

From the resulting distribution of  $metric^{(b)}$  values, the 95% confidence interval can be defined for the accuracy as:

$$CI_{95} = [metric_{(2.5)}^{(b)}, metric_{(97.5)}^{(b)}], \quad (23)$$

where  $metric_{(q)}^{(b)}$  denotes the  $q$ -th percentile of the bootstrap distribution for the given metric.

**Quantifying gain ( $\Delta$ )** To evaluate the influence of maneuver correction or sensor placement modifications on model performance, I computed the difference in accuracy and macro-F1 score between the modified model and the baseline configuration. The improvement metric is expressed as:

$$\Delta = metric_{\text{modified}} - metric_{\text{baseline}}. \quad (24)$$

To obtain a robust estimate of this difference and its uncertainty, I applied bootstrap resampling to both model configurations. For each resample iteration  $b$ , I computed the gain as:

$$\Delta^{(b)} = metric_{\text{modified}}^{(b)} - metric_{\text{baseline}}^{(b)}, \quad b = 1, \dots, B. \quad (25)$$

If the 95% confidence interval does not include zero, I considered the improvement statistically significant, thus confirming that the observed performance difference is unlikely due to random variation.

**Effect size** To complement the confidence interval analysis and assess the magnitude of the observed effects, Cliff's Delta ( $\delta$ ) was used as a non-parametric effect size measure. It quantifies the degree to which one distribution tends to yield higher values than another. Given the bootstrap accuracy and macro-F1 score samples  $x_{metric}$  and  $y_{metric}$  from the modified and baseline models, respectively, the Cliff's Delta is defined as:

$$\delta = \frac{\#(x_{metric} > y_{metric}) - \#(x_{metric} < y_{metric})}{N_x N_y}, \quad (26)$$

where  $\#(x_{metric} > y_{metric})$  and  $\#(x_{metric} < y_{metric})$  represent the number of pairwise comparisons where the accuracy from one model exceeds the other, and  $N_x$ ,  $N_y$  are the

respective sample sizes [94].

**Evaluation of combined best methods** Based on the metrics outlined above, the configuration exhibiting the best performance was identified as the optimal setup. Subsequently, I conducted further in-depth analysis using model-based interpretation tools to get a detailed view about how the model works under this configuration.

Principal Component Analysis was used as a first step in this evaluation as a diagnostic technique. While PCA is not inherently suited to interpreting model decisions, it provides insight into the structure and variance distribution across sensor signals. Specifically, PCA was applied to explore and visualize redundancy or correlation among different sensor channels, and to compare the variance captured by different sensor placements, highlighting whether certain configurations capture more distinctive signal patterns.

However, it is important to note that PCA identifies directions of maximal variance, not directions that maximize prediction performance. Therefore, for quantifying feature importance in the context of model outputs, I relied on complementary approaches better suited to supervised evaluation, such as feature importance and SHAP values.

### 5.6.2 Maneuver detection and correction experiments

The first experimental analysis focuses on evaluating the impact of maneuver correction, while maintaining a fixed sensor configuration. For this experiment, both suspension-mounted and cabin-mounted sensors were used to ensure a comprehensive input feature set. Separate models were trained and evaluated for two scenarios; with and without the application of the maneuver correction algorithm.

The results show a notable improvement in classification performance, with an absolute gain in accuracy of  $\Delta\text{accuracy} = 2.9\%$  and a macro-averaged F1 score improvement of  $\Delta\text{F1} = 2.7\%$ . The 95% confidence intervals confirm the statistical robustness of these gains. In addition, effect size was quantified using Cliff's Delta, yielding a value of  $\delta = 0.887$ . According to standard interpretation thresholds [94], this represents a large effect size, suggesting that maneuver correction has a consistent positive influence on the model's predictive capabilities. The detailed results of the comparison are presented in Table 17.

Table 17: Model metrics related to the maneuver correction

| <b>Metric</b>               | <b>Value</b> | <b>95% CI</b> |
|-----------------------------|--------------|---------------|
| $\Delta$ accuracy (%)       | 2.9          | [2.1, 3.7]    |
| $\Delta$ F1 (macro, %)      | 2.7          | [1.9, 3.5]    |
| Cliff's $\delta$ (accuracy) | 0.887        |               |

### 5.6.3 Sensor placement comparison

The second set of experiments investigated the effect of sensor location on road quality classification accuracy. Specifically, the performance of models trained using only cabin-mounted IMU sensor was compared with models trained on a combined input from both cabin and suspension-mounted sensors. In both cases, the maneuver correction algorithm was applied.

The results show a relative significant increase in model performance when both sensor locations are used. The absolute gain in accuracy was 6.9%, while the macro F1 score improved by 6.6%. Bootstrap-derived confidence intervals confirm the robustness of these gains. The effect size, as measured by Cliff's Delta, was  $\delta = 0.948$ , which is also interpreted as a large effect.

The findings suggest that although additional suspension-mounted sensors require a more complex measurement setup and marginally higher cost, their contribution to classification performance is high. Thus, supplementing the cabin-mounted IMU with suspension sensors is strongly justified, especially for applications requiring high reliability or performance in challenging environments. A summary of the results is provided in Table 18.

Table 18: Sensor placement comparison (cabin-only compared to suspension and cabin; both with maneuver correction)

| <b>Metric</b>               | <b>Value</b> | <b>95% CI</b> |
|-----------------------------|--------------|---------------|
| $\Delta$ accuracy (%)       | 6.9          | [5.5, 8.3]    |
| $\Delta$ F1 (macro, %)      | 6.6          | [5.3, 7.9]    |
| Cliff's $\delta$ (accuracy) | 0.948        |               |

### 5.6.4 Integrated modeling and feature importance analysis

The next step in the analysis was to integrate all preprocessed data and evaluate how the individual features, derived from different sensor types and locations, contribute to the model's classification performance.

First, I applied the PCA to examine the variance structure and interdependencies between

input features. The PCA biplot for the combined model configuration is presented in Fig. 30. The median-based sensor features are represented as vectors. The PCA results of the standard deviation-based variables showed similar results, therefore, these are not shown on the plot for better visibility.

The biplot indicates that the first principal component (PC1) primarily captures a combined vertical and longitudinal dynamic mode of the vehicle. Strong contributions to PC1 are observed from the cabin-mounted vertical acceleration  $a_z$  and the longitudinal gyroscope signal  $\omega_x$ , whose vectors are closely aligned with the PC1 axis and exhibit comparatively large magnitudes. The vertical accelerations measured at the suspension control arms ( $a_{pl}$  and  $a_{pr}$ ) are also oriented in a similar direction, indicating positive correlation with PC1, although their contributions are less dominant in magnitude. This alignment reflects the transmission of vertical road-induced excitations through the suspension system to the vehicle body, coupled with longitudinal rotational dynamics of the chassis. In contrast, the lateral cabin acceleration ( $a_y$ ) is oriented closer to the second principal component (PC2) and is nearly orthogonal to the dominant PC1 direction, suggesting that it captures vehicle dynamics largely independent of the vertical and longitudinal responses. Overall, the PCA results highlight the complementary nature of the cabin- and suspension-mounted sensors, demonstrating that they capture distinct yet partially correlated aspects of vehicle dynamics, thereby providing non-redundant information for the subsequent classification task.

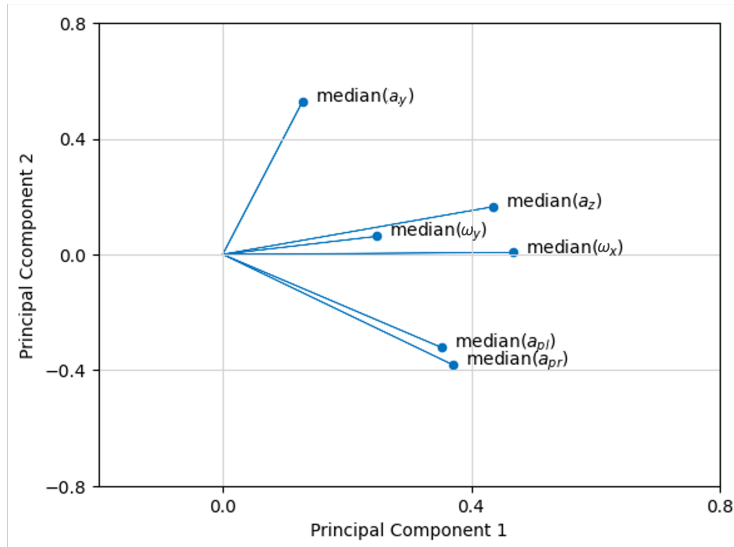


Figure 30: Principal Component Analysis of sensor variables based on median values

Fig. 31 illustrates classification accuracy under various model complexities and input configurations. Three configurations were tested: (1) using cabin and suspension sensors

without maneuver correction, (2) using only cabin sensors, and (3) the full pipeline including both sensor domains and maneuver correction. The accuracy trends show that increasing the number of decision trees in the Random Forest generally improves model performance across all configurations. The lowest average accuracy of 85.1% was observed in the cabin-only configuration, even at high model complexity. Adding suspension sensor data without maneuver correction, increased the accuracy to 87.5%. The full model configuration, which also included maneuver correction, achieved the highest classification accuracy of 90.8%. Notably, the combined model reaches competitive accuracy even with lower model complexity, highlighting the robustness of the feature set. These results reinforce the earlier findings, indicating that while maneuver correction is beneficial, however, the greatest impact stems from proper sensor placement.

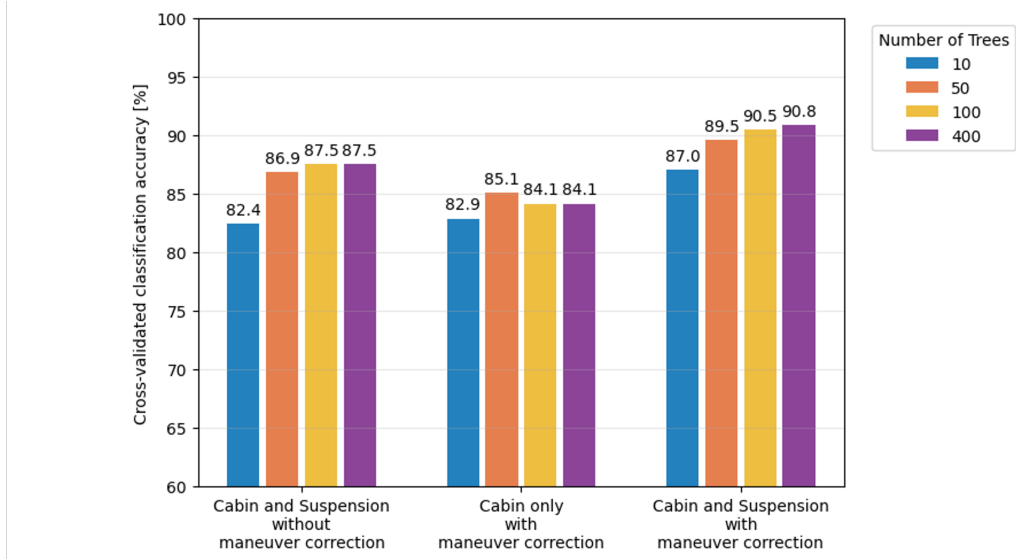


Figure 31: Classification accuracy under different model complexities and feature configurations

In the final phase of this investigation, feature importance was computed for the combined model using the internal scoring mechanism of the Random Forest classifier. Fig. 32 presents the relative importance values for each input feature. The figure shows that the top three features span across different sensor locations and types, underlining their collective value in capturing road surface conditions. The most dominant variable is the Y-axis acceleration from the cabin IMU, followed by the gyroscope's X-axis component, and finally the left rear suspension's vertical acceleration. The rear right sensor, located closer to the road shoulder, frequently encounters worse pavement conditions, which may explain its higher importance compared to its right-side counterpart. Feature importance decreases beyond the top three, with the lowest values assigned to the vertical-axis gyroscope signal.

Similarly, the forward-facing X-axis cabin acceleration also had low importance, which magnitude is often amplified during acceleration and braking, which maneuver correction aims to normalize.

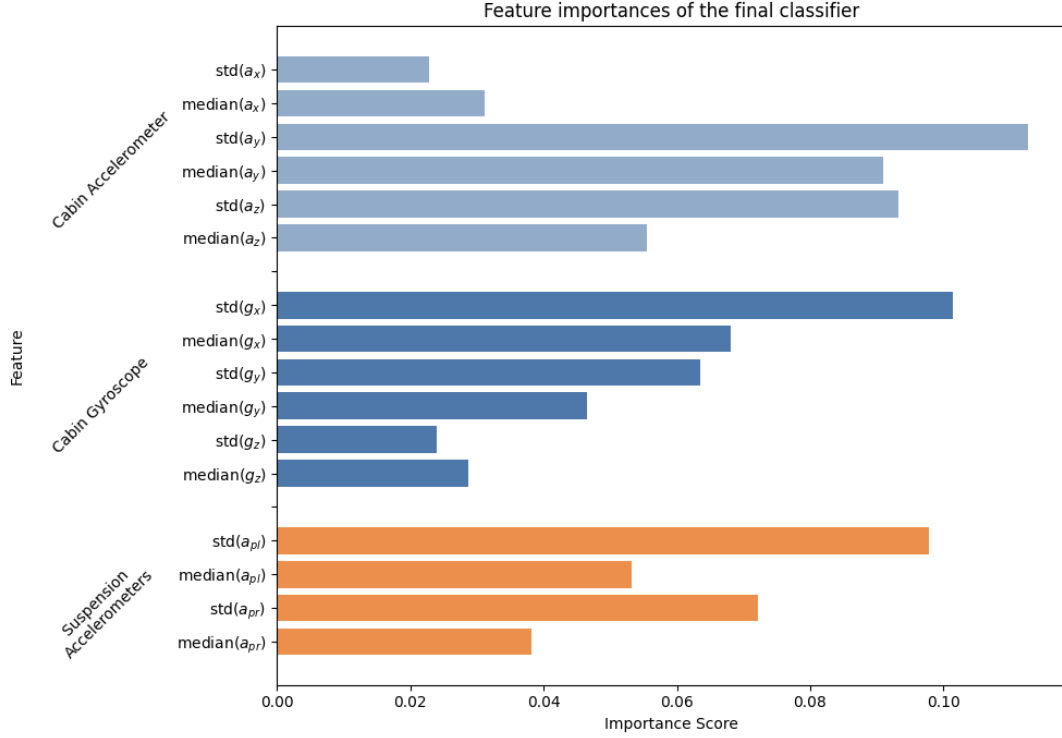


Figure 32: Feature importance in the final Random Forest model

Although the above chart represents average feature importance, it is important to examine the individual-level variation that can be hidden by aggregation. For this, SHAP values were calculated to analyze how individual sensor features influence the prediction for specific data points. Fig. 33 shows how specific sensor variables interact in different pavement categories to the model output. The diagram shows the top 9 features, sorted by the absolute SHAP value, to reduce the overall size of the plot to the most important axes.

Each SHAP beeswarm plot displays the impact of features on the prediction probability for a given class. The horizontal axis represents the SHAP value, which quantifies the marginal contribution of a given feature to the model output for that class. Positive SHAP values indicate that a feature increases the model's confidence in assigning the corresponding class, while negative SHAP values denote a decreasing effect. The color coding reflects the original feature value for each sample, with red indicating high values and blue indicating low values.

For the class 1, the poorest road quality class, higher values of several gyroscopic and accelerometric features, such as  $\text{median}(\omega_x)$ ,  $\text{median}(\omega_y)$ , and  $\text{median}(a_z)$ , tend to push predictions toward class 1. These features exhibit SHAP values with strong rightward spread in red hues, suggesting that high feature magnitudes are associated with the poorest road quality category. For class 2 and class 3, or intermediate quality classes, the SHAP patterns are less sharply polarized but still informative. The most influential features show more mixed color gradients and symmetric SHAP distributions, suggesting that intermediate road classes are not driven by extreme feature values but rather by transitional patterns in vibration characteristics. Notably, features like  $\text{std}(a_y)$  and  $\text{median}(a_y)$  appear frequently as important contributors, indicating that lateral vehicle acceleration variability provides discriminative information in moderate road conditions. Class 4, or the best road quality is in contrast to class 1, predictions for class 4 are primarily driven by low values of the same set of features. SHAP values are right-shifted for blue-colored points, indicating that low  $\text{std}(a_z)$ ,  $\text{median}(\omega_x)$ , and  $\text{median}(a_z)$  values increase the likelihood of a class 4 prediction.

### 5.7 Final production model and long-term measurements

In this subsection, I present the results of the production-ready classification model, which is based on the optimal combination of methods identified in the previous chapters. These include the use of maneuver-corrected input signals, and sensor placement at both the cabin and suspension. The model’s architecture were finalized through a hyperparameter optimization process, followed by evaluation on both held-out test data and previously unseen road segments to validate robustness.

To identify the optimal set of model parameters for the final Random Forest classifier, a grid search algorithm was employed. Table 19 presents the final combination of selected hyperparameters from the optimization procedure.

Following model training and validation, the confusion matrix of the final model is shown in Fig. 34. As observed, the highest classification accuracy with 93% was achieved for class 4, while the lowest was for class 3, with 88%. Importantly, most misclassifications were within one different class, which reflects the model’s ability to distinguish gradual transitions in pavement quality.

The average classification accuracy on the test data was 91.25%, and the overall ROC-AUC score was 0.986, with minimal deviation across classes. These metrics confirm that



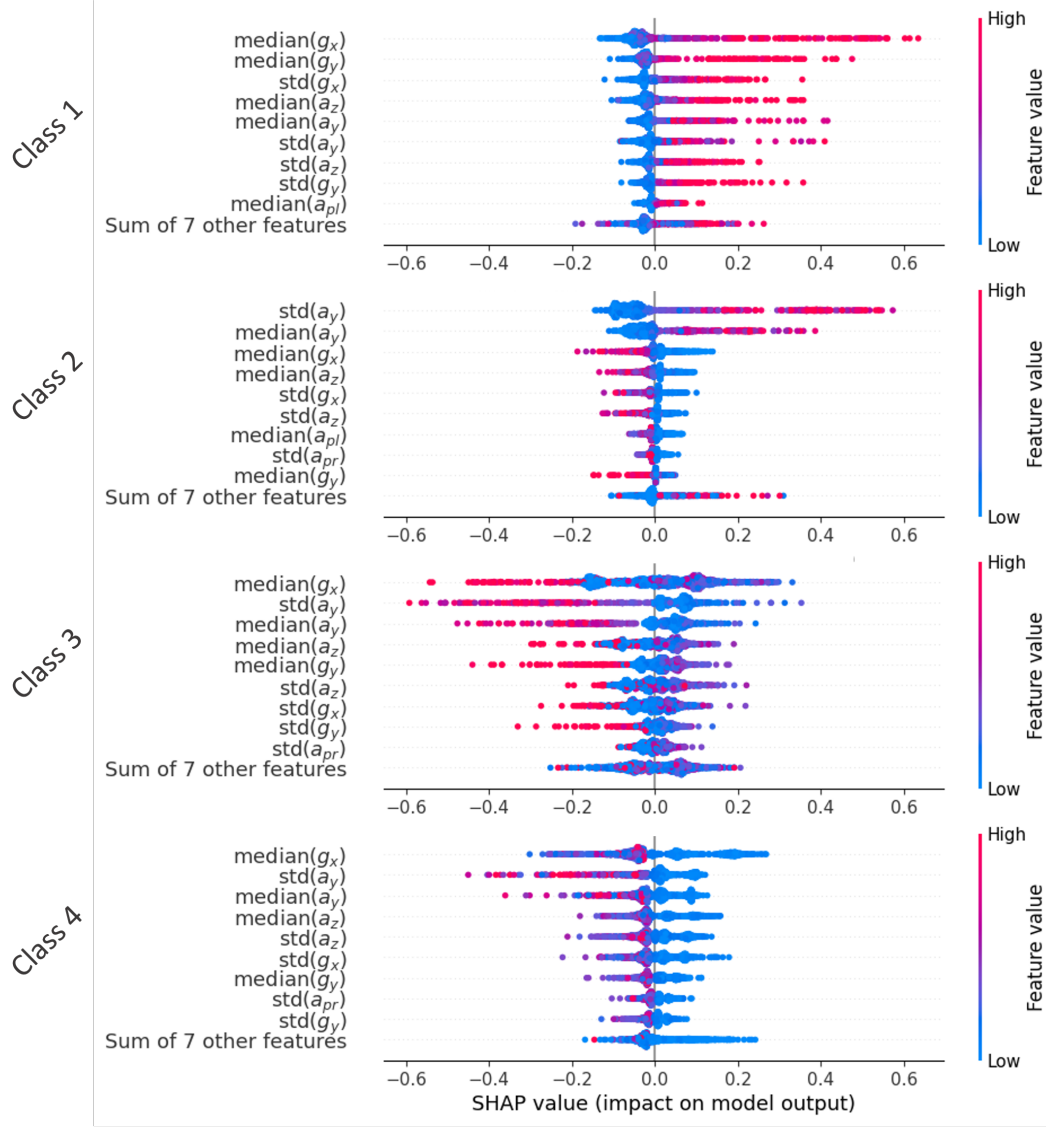


Figure 33: The nine most important variables and its SHAP values for each road quality class

Table 19: Optimized hyperparameters for the Random Forest classifier

| Hyperparameter               | Selected Value   |
|------------------------------|------------------|
| Bootstrap                    | False            |
| Criterion                    | Gini             |
| Max Depth                    | None (unlimited) |
| Min Samples Split            | 2                |
| Min Samples Leaf             | 1                |
| Number of Estimators ( $n$ ) | 300              |
| Max Features                 | sqrt             |
| Out-of-Bag Score             | False            |

the combination of maneuver correction and multi-location sensing, alongside MIL-based annotation with careful human supervision, yields a good performing supervised model

for road surface classification. Moreover, although the absolute gain from hyperparameter optimization was moderate, it still contributed to fine-tuning the accuracy.

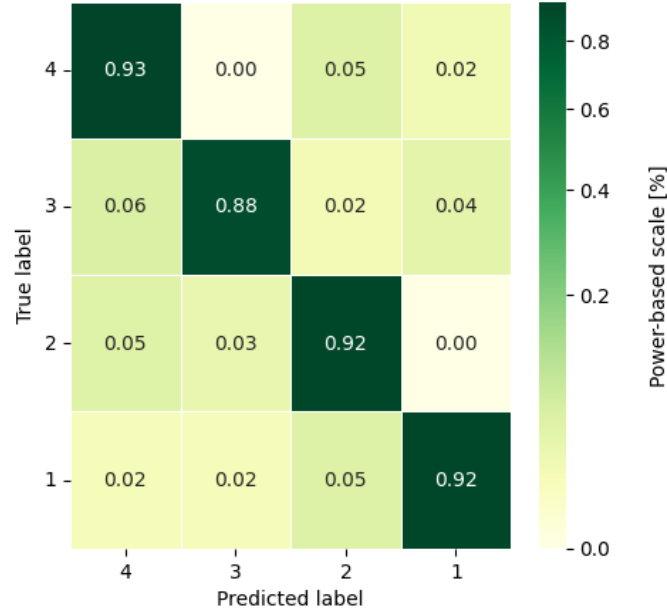


Figure 34: Confusion matrix of the predicted values, based on the Random Forest model

For visualization and interpretation purposes, the raw GPS coordinates obtained during measurements were post-processed through the map-matching algorithm, aligning them with real-world public road networks. Fig. 35 illustrates an example of this map-matching process on a short measurement segment. The orange dots represent the original GPS points recorded during the test drive, while the continuous blue line indicates the matched path generated by the algorithm using a linestring object representation. While the raw GPS points visually follow the approximate road shape, inconsistencies and positional drift are observable, especially in areas with poor satellite visibility or abrupt maneuvers. These inaccuracies can hinder the exact localization of pavement defects or transitions. In contrast, the map-matched linestring offers a spatially coherent and topology-respecting route, with proper alignment with the actual road network.

The map-matching algorithm has several beneficial results. First, map-matching ensures that all measurement points refer to a precise, physical road segment, which is important for aggregating and comparing measurements over time. Second, it addresses the duplicate data problem that arises during repeated drives over the same segment, as raw GPS coordinates, due to their random noise, would differ even when the vehicle follows the same route. With map-matched linestring identifiers, repeated measurements over the same physical location are automatically renewed, eliminating spatial redundancy.

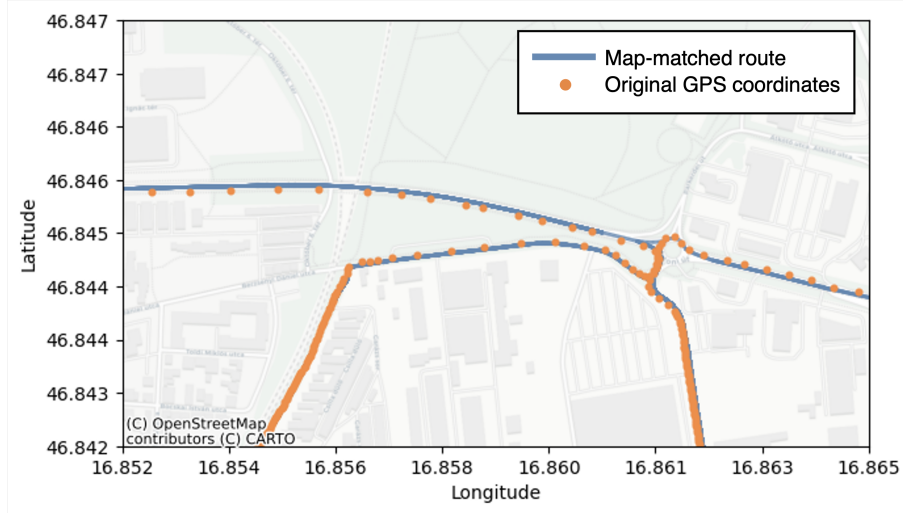


Figure 35: Comparison of raw GPS coordinates (orange) and map-matched linestring (blue) for a sample measurement segment

Finally, I illustrate the output of the entire algorithm and the classification results of the finalized RF model with the two figures below. Fig. 36 shows a larger area with the carried measurements, carried out in a populated area. The road surface quality was highly variable, with a number of coherent longer sections. Labels 1 and 2 indicated sections contained a diverse quality, but mostly degraded condition pavements, with potholes and short ruttings. These areas show that the poorer quality sections are typically composed of shorter, highly variable sections, indicating a continuous occurrence of small individual defects. These observations and experiences are supported by my visual observations. In contrast, Label 3 indicated area had an overall good quality pavement with a few minor cracks and previous repairs, as confirmed by the classification results varying between Good and Perfect.

Fig. 37 shows a measurement section outside a city, mainly on lower-order roads. Label 1 area shows a small, low-traffic road section connecting villages. This is a section that was renovated a long time ago and is in extremely poor condition, with only a few minor surface repairs. The surface was mostly characterized by cracks, rutting, and potholes. The section is therefore classified as Poor in the model, with a few better-rated sections due to the repaired surfaces. The Poor category sections are also relatively longer, which was rare during the measurements. The Label 2 indicated area shows a higher-ranked road section, which in reality has a relatively new or partially renovated road surface. In line with this, the classification of the section is mostly in the Good and Perfect categories, with a few small Medium or Poor sections. When scheduling road maintenance, these small, poor-quality outlier sections can provide useful information so that repairs can be made before

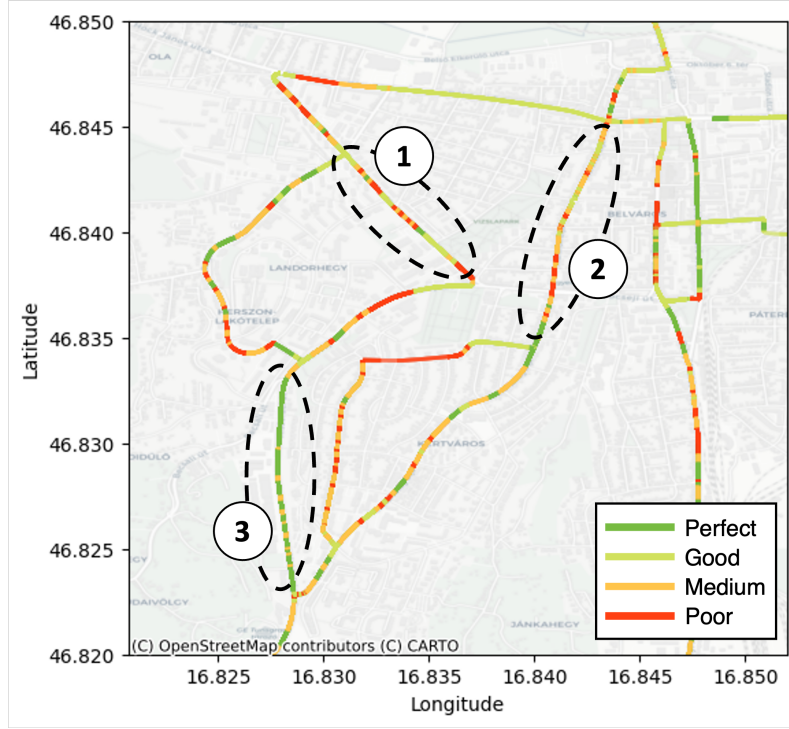


Figure 36: Low-speed measurement and classification results from the final model in an urban residential environment

their extent increases. The rest of the route continues on a road surface of the same grade, with Label 3 indicating a notable section. This section of the road had been renovated previously, but due to a lack of subsequent repairs, numerous surface irregularities and potholes have begun to form, which can be clearly seen in the classification. Due to the exponential spread of road surface deterioration, it would be of priority importance to repair these defects as soon as possible, while they are still isolated and small in extent.

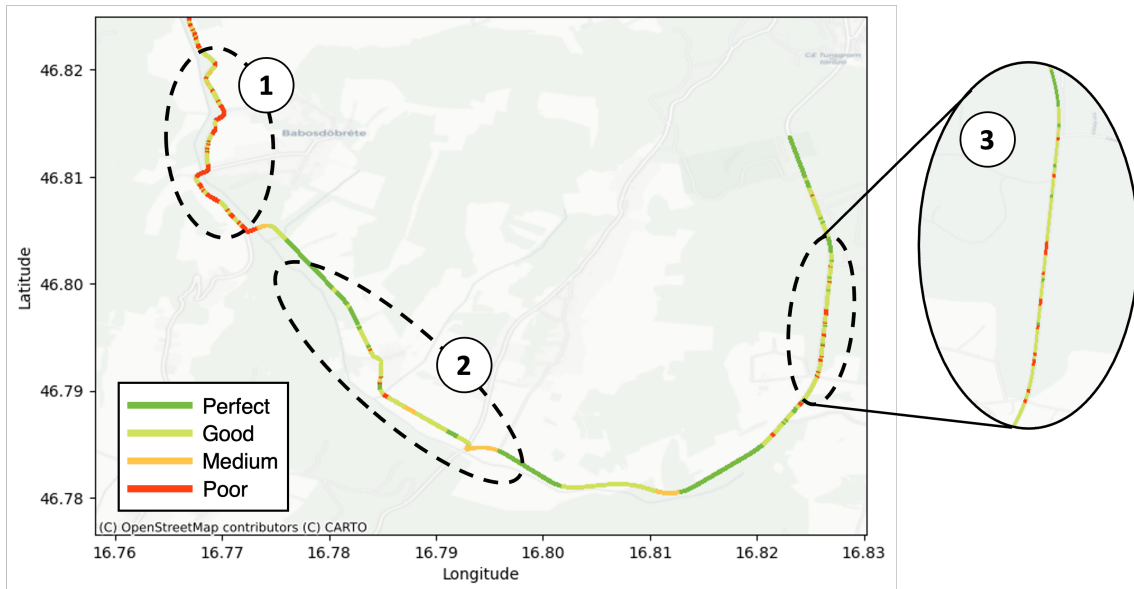


Figure 37: Classification results for variable vehicle speed measurement sections outside urban areas on lower and medium-grade road sections

## 6 Summary

The primary objective of the dissertation was to develop a robust and practical method for automated pavement quality assessment using vehicle-mounted vibration sensors and machine learning algorithms.

My progress in the research was greatly helped by my personal motivation and the new challenges inherent in the solution. The quality of the road surface in Hungary shows a significantly different picture, but on many, mainly lower-level roads, its condition is unsatisfactory, maintenance is often delayed or carried out too late. In my opinion, this significantly affects everyday life, since we spend a relatively significant amount of time in traffic, so daily travel on a poor-quality road surface can also have a negative impact on our mental state, not to mention the specific health risks and technical failures of the vehicle. Therefore, I hope that my work will contribute to the growing research field of indirect road monitoring, where vehicle dynamics and sensor feedback are utilized for road surface condition estimation.

Firstly, I summarized the state of the art knowledge about the topic, which is presented in Section 2. In this, I briefly reviewed the relevant vehicle dynamics knowledge, the currently applied and widespread techniques in the field of road surface quality measurement, and finally summarized the most relevant knowledge related to data mining and machine learning.

After that, I defined the goals and main topics of the research based on the shortcomings of the current literature and the unexplored but potentially beneficial development aspects. The main goal of my research can be divided into three aspects, thus I examined the applicability of labeling based on the Machine-In-the-Loop method in the field of vibration-based road quality evaluation, then the added information content of sensor placement on the suspension and the effect of aggressive maneuver correction on the classification algorithm.

To achieve the defined goals, the research was carried out in several stages. First, an integrated data acquisition platform was developed, using multiple sensor types, specifically, an IMU mounted in the vehicle cabin and two piezoelectric accelerometers located on the rear suspension. The IMU sensor measured linear acceleration and angular velocity, while the piezoelectric accelerometer captured the vertical accelerations of the rear control arms, allowing for detailed reconstruction of the vehicle's dynamic response to surface

irregularities.

Subsequently, a preprocessing workflow has been created for handling raw sensor data, including time alignment, noise filtering, spatial resampling, and power spectral density feature extraction. Then, the unsupervised clustering was applied using a Gaussian Mixture Model, with clustering parameters optimized for real-world interpretability and robustness. Clusters were interpreted as discrete pavement quality classes, validated through manual inspection and continuous supervision, and evaluated using confidence-based assignment metrics.

The labeled dataset served as the basis for the final supervised learning model, where Random Forest classifiers were trained to predict road quality from sensor signals. Then, I investigated how different sensor configurations and data corrections affect classification performance. Specifically, the impact of lateral or longitudinal maneuver-induced noise and sensor placement between cabin exclusive or cabin and suspension were examined in isolation and combined. The resulting metrics confirmed that both factors significantly affect prediction accuracy, with sensor placement having a larger absolute impact. The addition of suspension-mounted sensors increased average classification accuracy by up to 6.9%, while maneuver correction contributed a further 2.9% gain. Feature importance rankings and PCA visualizations confirmed that both cabin and suspension signals carry complementary, non-redundant information.

It is noted that certain road hazards, including localized road surface defects, may not always generate a distinct vibration signature in the measured vertical acceleration signals if they are successfully avoided by the driver through a rapid steering maneuver. In such cases, both the road-induced excitation and the associated vehicle maneuver may be suppressed by the applied maneuver-filtering algorithm, resulting in the event remaining undetected. This behavior represents a limitation of single-vehicle, onboard vibration sensor-based road monitoring approaches. However, since the primary objective of this research is the characterization and classification of overall road quality and road segments rather than the exhaustive detection of individual surface defects, the potential omission of isolated events does not significantly affect the validity of the presented results. The impact of this limitation can be reduced in future implementations by aggregating measurement data from multiple vehicle passes over the same road segment using geospatial alignment, thereby increasing the likelihood of consistent road hazard detection.

In conclusion, the research addressed a novel methodology for indirect road surface classification using in-vehicle sensors and machine learning. The system is practical for real-world deployment, provides high classification accuracy, and supports spatially coherent data visualization. Future work may explore real-time onboard processing, integration with other sensor modalities, for example cameras or GNSS-IMU fusion, and the extension of this framework to predict not only current surface state but also degradation trends over time.



## Theses

**Thesis #1.** I investigated the added information content of vehicle cabin and suspension sensor data for vibration-based road surface quality measurement and classification using a custom-developed multi-sensor measurement system.

- I developed and built a data acquisition system for vehicle vibration measurements. The unit is capable of recording tri-axial acceleration and angular velocity inside the vehicle cabin, as well as vertical acceleration on both sides of the rear suspension control arms. The system also logged GPS data for location tracking. Piezoelectric 805M1-0020 accelerometers were mounted on the left and right rear control arms, while an MPU6050 MEMS-based IMU sensor was installed inside the cabin. The main controller was a Raspberry Pi 4B. Additional electronic circuits, including 12-bit MAX186 ADC module and level-shifters, were implemented on custom-designed PCBs.
- I conducted long-term vehicle vibration measurements on public road networks mainly in Zala and Vas counties, including both urban sections and interurban routes connecting major towns. The database therefore contains road segments of widely varying surface quality, ranging from newly built or resurfaced sections to moderately worn and heavily degraded pavements with continuous cracking and potholes. The total length of road segments recorded for training the subsequent unsupervised annotation algorithm was 82 kilometers. After the model training, an additional 267 kilometers of road network data were used for validation purposes. Measurements were performed using two test vehicles with different suspension architectures, first a 1997 BMW 318is sedan with a fully independent multilink rear suspension, and second a 2019 Nissan Leaf with a torsion beam rear suspension and separated spring–damper elements, supporting the generalization of the findings.
- I processed the raw measurement data through a multi-stage data processing pipeline developed to filter out measurement errors, normalize signals, and correct for sudden maneuvers and vehicle speed variations. The measured routes were segmented into homogeneous sections based on road surface quality using the Pruned Exact Linear Time (PELT) change point detection algorithm, with automatic identification of homogeneous segments without prior definition of segment length or number. Each segment was characterized by the median and standard deviation of the relevant variables to reduce data dimensionality. The resulting dataset was then processed

and evaluated using supervised machine learning models to assess the influence of sensor placement and type on classification performance. Among six tested models, the Random Forest classifier achieved the best performance. After hyperparameter optimization, I trained the final model using the Gini criterion, a minimum split size of 2, a minimum leaf size of 1, and 300 estimators without depth limitation. The model classified road sections into four quality categories, ranging from severely degraded to newly resurfaced surfaces. The final classification accuracy was 91.2%, and the area under the ROC curve (ROC-AUC) was 98.6%.

- Through Principal Component Analysis, feature importance metrics, and SHAP value evaluations of the Random Forest model, it has been established that vibration data measured on the suspension control arms contribute significant additional information compared to cabin-mounted sensors. The incorporation of suspension acceleration data resulted in a 6.9% increase in classification accuracy compared to models trained exclusively on cabin acceleration and angular velocity signals. The results confirmed the necessity of multi-location sensor configurations, as different measurement axes and mounting positions provide critical information depending on specific road surface conditions. I determined that the most influential variables are the vertical and lateral accelerations measured in the cabin, the angular velocity around the longitudinal axis in the cabin, and the vertical acceleration measured on the left rear control arm.

**Related publications: 1, 2, 3, 8**

**Thesis #2.** I introduced and evaluated the applicability of a Machine-In-the-Loop (MIL) annotation method for vibration-based road surface quality classification.

- Due to the absence of ground-truth annotations for the measured road sections, a semi-automated Machine-In-the-Loop annotation approach was developed. The proposed method is based on an unsupervised machine learning algorithm, wherein a Gaussian Mixture Model were employed to cluster vibration segments lacking reference labels. The clustering model utilized all vibration-related variables of the pre-segmented road sections obtained through the PELT change point detection algorithm. During model fitting, the maximum number of iterations was set to 300, the covariance type was defined as “full,” the tolerance threshold was set to  $10^{-4}$ , and the initialization method was “k-means.”
- I demonstrated that the developed MIL-based clustering algorithm was capable of

grouping road sections of varying lengths into coherent clusters that exhibited internal homogeneity in terms of vibration and surface quality characteristics, without spatial resolution limitations. In addition to manual verification, the optimal number of clusters and thus the number of distinguishable road quality classes were determined based on statistical metrics of the clustering model, including the Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), Silhouette score, and Log-Likelihood score. The first three indicators were analyzed using the elbow method for values of  $K$  between 2 and 9, which were selected based on industry standards and previous empirical practice.

- The elbow analysis revealed that both BIC and AIC values decreased with increasing  $K$  cluster number, indicating improved model fit; however, a plateau was observed beyond  $K = 5$  to 6. The Silhouette score reached its maximum between  $K = 3$  and 5, followed by a steep decline, indicating reduced cluster interpretability beyond this range. I established that the balance between model complexity and cluster robustness was achieved for  $\leq 3 \leq K \leq 5$ . These class numbers are consistent with industry practice, providing both interpretable quality categories and sufficient differentiation for maintenance decision-making. I evaluated the confidence of cluster assignments through the cumulative probability distribution, demonstrating that for four identified quality classes, approximately 80% of all data points were assigned to clusters with confidence above 90%. The supervised learning model trained on the MIL-labeled dataset accurately captured the underlying vibration–surface quality relationships and achieved strong classification performance. Consequently, it has been established that the proposed Machine-In-the-Loop annotation approach is suitable for generating reliable training data for supervised models in vehicle vibration-based road surface quality assessment.

#### **Related publications: 4, 9**

**Thesis #3.** I investigated the negative impact of lateral and longitudinal aggressive vehicle maneuvers on vibration-based road surface quality assessment and introduced a detection and correction method to mitigate the signal-distorting effects of such abrupt maneuvers.

- I analyzed the occurrence and effects of aggressive vehicle maneuvers that adversely influence vibration-based road quality evaluation. A machine learning–based algorithm was developed to detect such abrupt driving events, focusing on four primary maneuver types: left turn, right turn, sudden braking, and sudden acceleration. For

model training, a dedicated dataset was collected over a total distance of 31 kilometers, containing approximately 70% straight and 30% curved road sections. In total, 340 longitudinal and 200 lateral maneuvers were labeled across both test vehicles. Due to the underrepresentation of lateral maneuvers, oversampling was performed using the Synthetic Minority Oversampling Technique (SMOTE), generating 150% additional synthetic observations from each sample. The labeled thresholds were defined as  $7 \text{ m/s}^2$  for acceleration,  $10 \text{ m/s}^2$  for braking, and  $5 \text{ m/s}^2$  for turning both directions. I trained a multiclass Support Vector Machine classifier to detect the type of aggressive maneuver. The model employed a polynomial kernel with a gamma value of 300 and a kernel degree of 3. The results demonstrated higher accuracy for lateral maneuvers, with the right-turn class exhibiting the lowest error rate. Longitudinal maneuvers, such as braking and acceleration, yielded slightly higher errors, particularly for the hatchback vehicle, which may be attributed to its softer suspension characteristics. The mean classification error was 7.3% for the Nissan Leaf and 6.7% for the BMW 318is, resulting in an overall average classification accuracy of 93%.

- Based on the detected maneuvers, the occurrence and impact of aggressive maneuvers in vibration-based road surface evaluation were analyzed. I established that such maneuvers can occur on both curved and straight sections, often as a result of sudden traffic changes. It was observed that these abrupt accelerations, braking, or steering events may distort the vibration signals and negatively affect road surface classification accuracy.
- SHAP-based analysis of the previously trained classification model showed that during normal, steady driving on straight roads, the dominant features influencing classification were vertical acceleration and longitudinal angular velocity, both of which are directly related to the pavement condition. However, during sudden turning maneuvers, the dominant signal component shifted to the yaw angular velocity around the vertical axis, which represents vehicle motion rather than road quality. These observations confirmed that maneuver detection and subsequent correction are necessary to maintain the reliability of vibration-based road quality assessment under diverse driving conditions. Longitudinal maneuvers primarily distort the accelerometer signal along the longitudinal axis, while lateral maneuvers mainly affect the gyroscope data along the vertical axis due to increased yaw motion during cornering.
- To mitigate the signal-distorting effects of aggressive maneuvers, I introduced a cor-

rection method based on the SHAP analysis results. A rule-based decision tree logic was constructed to proportionally dampen the signal components most affected by the detected maneuvers. For abrupt lateral maneuvers, the transverse acceleration and yaw angular velocity variables were downscaled, while for longitudinal maneuvers, the longitudinal acceleration and pitch angular velocity components were corrected. The correction was applied proportionally to the temporal proximity of the maneuver events, ensuring smooth signal recovery. As a result of this correction method, the overall classification accuracy of the vibration-based road quality model improved by 2.9%.

**Related publications:** 5, 10, 11

## Publications related to theses

### Articles in international journals

1. R. Nagy, I. Szalai, "Design and Construction of a Vibration Data Acquisition System for Road Vehicles" *Hungarian Journal of Industry and Chemistry*, vol. 48(1), pp. 71-75, 2020, doi: 10.33927/hjic-2020-11.

Scimago Journal Ranking: Q4, Impact factor: 0.5

2. R. Nagy, I. Szalai, "High-Resolution Road Segment Analysis using a Multi-Sensor Road Quality Classification System" *London Journal of Engineering Research*, vol. 23(1), pp. 9-20, 2023, ISSN: 2631-8474.

Scimago Journal Ranking: -, Impact factor: -

3. R. Nagy, A. Kummer, J. Abonyi, et al., "Machine learning-based soft-sensor development for road quality classification" *Journal of Vibration and Control*, vol. 30(11-12), pp. 2672-2684, 2024, doi: 10.1177/10775463231183307.

Scimago Journal Ranking: Q1, Impact factor: 2.4

4. R. Nagy, I. Szalai, "Vehicle maneuver recognition and correction algorithm for road quality measurement system optimization" *Measurement: Sensors*, vol. 38, 101816, 2025, doi: 10.1016/j.measen.2025.101816.

Scimago Journal Ranking: Q2, Impact factor: -

5. R. Nagy, I. Szalai, "Development of an unsupervised learning-based annotation method for road quality assessment" *Transportation Engineering*, vol. 21, 100358, 2025, doi: 10.1016/j.treng.2025.100358.

Scimago Journal Ranking: Q1, Impact factor: -

### Articles in conference publications

6. R. Nagy, I. Szalai, "Gépjárművekben alkalmazható vibrációs adatgyűjtő rendszer tervezése és építése" *XXVII. Nemzetközi Gépészeti Találkozó (OGÉT)*, Nagyvárad, 27, 2019, ISSN: 2068-1267.

7. R. Nagy, I. Szalai, "Személygépjárművekben alkalmazható több szenzoros útfelület

minőségmérő eszköz fejlesztése" *XXX. Nemzetközi Gépészeti Találkozó (OGÉT)*, Székelyudvarhely, 30, 2022, ISSN: 2068-1267.

### Conference abstracts

8. R. Nagy, I. Szalai, "Development of machine learning assisted suspension vibration data-based road quality classification system" *19th IMEKO TC10 Conference "MACRO meets NANO in Measurement for Diagnostics, Optimization and Control"*, Delft, The Netherlands, pp 33-39, 2023, ISBN: 9789299009048.
9. R. Nagy, I. Szalai, "Applying data science methods to road vehicle manoeuvre detection" *The People Events: 2nd World Conference on Data Science, Amsterdam, The Netherlands*, 17-19.06.2024.
10. R. Nagy, I. Szalai, "Unsupervised learning-based MIL annotation method for road quality assessment" *The People Events: 3rd World Conference on Data Science, Amsterdam, The Netherlands*, 16-18.06.2025.
11. R. Nagy, I. Szalai, "Development of an unsupervised learning-based annotation method for road quality assessment." *D3A Danish Digitalization, Data Science and AI 3.0, Nyborg, Denmark*, 26-27.08.2025.

## Publications not related to theses

### Articles in international journals

12. R. Nagy, I. Szalai, "Investigation of the sedimentation characterization of magnetorheological fluids" *Journal of Molecular Liquids*, vol. 390, 123047, 2023, doi: 10.1016/j.molliq.2023.123047.

Scimago Journal Ranking: Q1, Impact factor: 5.2

### Conference abstracts

13. R. Nagy, I. Szalai, T. Fülöp, "Pozíció meghatározása inerciális szenzorokkal" *XXIII. Energetika, Elektrotechnika, Számítástechnika és Oktatás Multi-konferencia, Marosvásárhely*, pp. 41-46, 2022, ISSN: 2734-6757.
14. D. Balogh, S. Mester, R. Nagy, et al., "Gépjármű gyalogossal való ütközése a gyalogos tesztbábu szempontjából" *XXIII. Energetika, Elektrotechnika, Számítástechnika és Oktatás Multi-konferencia, Marosvásárhely*, pp 1-6, 2022, ISSN: 2734-6757.



## References

- [1] Nuno S., Vaibhav S., João S., et al. Road anomalies detection system evaluation. *MDPI Sensors*, 18(7):1–20, 2018.
- [2] Ahlin K., Granlund J., and Lundström R. Whole-body vibration when riding on rough roads—a shocking study.technical report. *Swedish National Road Administration*, 31E, 2000.
- [3] Granlund J. Vehicle and human vibration due to road condition. *The ROADDEX “Implementing Accessibility” Project*, pages 1–10, 2012.
- [4] Marasco G., Rosso M. M., Aiello S., and et al. Ground penetrating radar fourier pre-processing for deep learning tunnel defects’ automated classification. *International Conference on Engineering Applications of Neural Networks*, pages 165–176, 2022.
- [5] Blank M. and Margolis D. The effect of normal force variation on the lateral dynamics of automobiles. *Society of Automotive Engineers*, (960484), 1996.
- [6] Masino J., Thumm J., Levasseur G., et al. Characterization of road condition with data mining based on measured kinematic vehicle parameters. *Wiley, Hindawi - Journal of Advanced Transportation*, pages 1–10, 2018.
- [7] Al-Mamany A., Hussein A. H., and Nihad R. Z. Development of pavement managementsystem on road network using geographic information system method - kirkuk to erbil highway. *Design Engineering*, 6:37–47, 2021.
- [8] Sauerwein P. and Smith B. Investigation of the implementation of a probe-vehicle based pavement roughness estimation system. *Mid-Atlantic Universities Transportation Center*, 6:21–63, 2021.
- [9] DIN EN 13036-6. Road and airfield surface characteristics - test methods - part 6: Measurement of transverse and longitudinal profiles in the evenness and megatexture wavelength ranges. 2008.
- [10] Chengyi Huang and Xin Huang. Effects of pavement texture on pavement friction: a review. *International Journal of Vehicle Design*, 65(2-3), 2014.
- [11] Roos R. Lecture notes in betrieb und erhaltung von straßen. *Institute of Highway and Railroad Engineering, Karlsruhe Institute of Technology, Karlsruhe, Germany*, pages 1–10, 2013.
- [12] E. Straube and K. Krass. Straßenbau und straßenerhaltung - ein hand- buch für studium und praxis, 10th ed. berlin. *Erich Schmidt Verlag*, page 336, 2016.
- [13] US Department of Transportation Federal Highway Administration. Comprehensive truck size and weight study, 1995. Last accessed 6 February 2025.
- [14] Wong J. Y. Theory of ground vehicles. *John Wiley and Sons Inc.*, pages 438–489, 2001.
- [15] Sattar S., Li S., and Chapman M. Developing a near real-time road surface anomaly detection approach for road surface monitoring. *Measurement*, 185, 2021.
- [16] Jakob E., Lewis G., Bret H., et al. The pothole patrol: Using a mobile sensor network

- for road surface monitoring. *MobiSys'08*, 2008.
- [17] Nguyen T., Lechner B., and Wong D. Y. Response-based methods to measure road surface irregularity: a state-of-the-art review. *European Transport Research Review*, 11(43):1–18, 2019.
- [18] Masino J. Road condition estimation with data mining methods using vehicle based sensors. *KIT-Fakultät für Maschinenbau Karlsruher Institut für Technologie*, pages 3–20, 2019.
- [19] Radopoulou S. C. and Brilakis I. Improving road asset condition monitoring. *Transportation Research Procedia*, 14:3004–3012, 2016.
- [20] Ruseruka C., Mwakalonge J., Comert G., et al. Road condition monitoring using vehicle built-in cameras and gps sensors: A deep learning approach. *Vehicles*, 5(3):931–948, 2023.
- [21] Douangphachanh V. and Oneyama H. A study on the use of smartphones for road roughness condition estimation. *J-STAGE Journal of the Eastern Asia Society for Transportation Studies*, 10:1551–1564, 2013.
- [22] Doumiati M., Erhart S., Martinez J., et al. Adaptive control scheme for road profile estimation: Application to vehicle dynamics. *IFAC Proceedings Volumes, 19th IFAC World Congress*, 47(3):8445–8450, 2014.
- [23] Laurent J., Talbot M., and Doucet M. Road surface inspection using laser scanners adapted for the high precision 3d measurements of large flat surfaces. *International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, page 303–310, 1997.
- [24] Hebert J. F. and Laurent J. High performance 3d sensors for the characterization of road surface defects. *IAPR Workshop on Machine Vision Applications, Nara, Japan: Machine Vision Applications*, page 388–391, 2002.
- [25] Laurent J., Hebert J. F., Lefebvre D., et al. Using 3d laser profiling sensors for the automated measurement of road surface conditions. *RILEM Bookseries*, 4:157–167, 2012.
- [26] Chen K., Lu M., Tan G., et al. Crsm: Crowdsourcing based road surface monitoring. *IEEE 10th International Conference on High Performance Computing and Communications and 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCCEUC)*, page 2151–2158, 2013.
- [27] Radopoulou S. C. and Brilakis I. Automated detection of multiple pavement defects. *Journal of Computing in Civil Engineering*, 31(22):2151–2158, 2016.
- [28] Sabillon C. Evaluation of data imputation techniques in pavement texture processing. *Masters Thesis*, 2020.
- [29] Martinez E. A., Bustamante-Bello R. M., and Arce-Sáenz A. L. A review of road surface anomaly detection and classification systems based on vibration-based techniques. *MDPI Applied Sciences*, 12(9413), 2022.
- [30] Hungarian Public Road Authority. On the public road network.

- 
- [31] Rajput P., Chaturvedi M., and Patel V. Road condition monitoring using unsupervised learning based bus trajectory processing. *Multimodal Transportation*, 4(1):100041, 2022.
  - [32] Nuno S., Joao S., Vaibhav S., et al. Anomaly detection in roads with a data mining approach. *Procedia Computer Science*, 121:415–422, 2017.
  - [33] Standard practice for computing international roughness index of roads from longitudinal profile measurements. *ASTM International*, pages 1926–08, 2015.
  - [34] Sayers M. W., Gillespie T. D., and Paterson W. Guidelines for conducting and calibrating road roughness measurements. *World Bank Technical Paper*, 46, 1986.
  - [35] Sayers M. W., Gillespie T. D., and Queiroz V. A. C. The international road roughness experiment: A basis for establishing a standard scale for road roughness measurements. *Transportation Research Board*, 1084, 1986.
  - [36] Mucka P. and Granlund J. Comparison of longitudinal unevenness of old and repaired highway lanes. *American Society of Civil Engineers: Journal Of Transportation Engineering*, 138(3):371–380, 2012.
  - [37] Praznowski K., Mamala J., Smieja M., et al. Assessment of the road surface condition with longitudinal acceleration signal of the car body. *MDPI Sensors*, 20(21):1–19, 2020.
  - [38] Loizos A. and Plati C. An alternative approach to pavement roughness evaluation. *International Journal of Pavement Engineering*, 9(1):69–78, 2008.
  - [39] Sun L. Simulation of pavement roughness and iri based on power spectral density. *Mathematics and Computers in Simulation*, 61:77–88, 2002.
  - [40] Marcondes J. A., Snyder M. B., and Singh S. P. Predicting vertical acceleration in vehicles through road roughness. *Journal of Transportation Engineering*, 118(1):33–49, 1992.
  - [41] Hesami R. and McManus K. J. Signal processing approach to road roughness analysis and measurement. *TENCON 2009 - 2009 IEEE Region 10 Conference Proceedings*, pages 1–6, 2009.
  - [42] Wang G. An investigation of the suitability of smartphone devices for road condition assessment. *University of Birmingham*, pages 109–154, 2018.
  - [43] Moad K., Bruno M., Xavier M., et al. Adaptive robust vehicle motion control for future over-actuated vehicles. *MDPI Machines*, 7(26):1–19, 2019.
  - [44] R. N. Jazar. Vehicle dynamics: Theory and application. *Springer*, 1066, 2014.
  - [45] Ward C. and Iagnemma K. Speed-independent vibration-based terrain classification for passenger vehicles. *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility*, 47(9):1095–1113, 2009.
  - [46] Samaniego E., Anitescu C., Goswami S., et al. An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications. *Computer Methods in Applied Mechanics and Engineering*, 362(112790), 2020.

- 
- [47] Guiggiani M. The science of vehicle dynamics: Handling, braking, and ride of road and race cars. 2014.
- [48] Zheng Z., Zhou M., Chen Y., et al. Qdetect: Time series querying based road anomaly detection. *IEEE Access*, 99, 2020.
- [49] Zheng Z., Zhou M., Chen Y., et al. A fused method of machine learning and dynamic time warping for road anomalies detection. *IEEE Transactions on Intelligent Transportation Systems*, 23(2):827–839, 2020.
- [50] Mednis A., Strazdins G., Zviedris R., et al. Real time pothole detection using android smartphones with accelerometers. *International Conference on Distributed Computing in Sensor Systems and Workshops*, pages 1–6, 2011.
- [51] González A., O’Brien E. J., Li Y. Y., and Cashell K. The use of vehicle acceleration measurements to estimate road roughness. *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility*, 46(6):483–499, 2008.
- [52] Harris N. K., Gonzalez A., OBrien E. J., et al. Characterisation of pavement profile heights using accelerometer readings and a combinatorial optimisation technique. *Journal of Sound and Vibration*, 329(5):497–508, 2010.
- [53] Chao W., Zhen W., Simon H., et al. An automated machine-learning approach for road pothole detection using smartphone sensor data. *MDPI - Sensors*, 20(5564), 2020.
- [54] Mikko P., Oleksiy M., Fengyu C., et al. Distributed road surface condition monitoring using mobile phones. *Springer - Verlag*, (6905):64–78, 2011.
- [55] Harikrishnan P. M. and Varun P. G. Vehicle vibration signal processing for road surface monitoring. *IEEE Sensors Journal*, 17(16):5192 – 5197, 2017.
- [56] Society of Automotive Engineers; Vehicle Dynamics Committee. Ride and vibration data manual. *SAE J6a: Report of Riding Comfort*, 114, 1965.
- [57] Nitsche P., Stutz R., Kammer M., et al. Comparison of machine learning methods for evaluating pavement roughness based on vehicle response. *Journal of Computing in Civil Engineering*, 28(4), 2012.
- [58] Y. c. Tai, C. w. Chan, and J. Y. j. Hsu. Automatic road anomaly detection using smart mobile device. *15th Conference on Artificial Intelligence and Applications (TAAI), Taipei, Taiwan: Taiwanese Association for Artificial Intelligence*, pages 1–8, 2010.
- [59] Seraj F., B. J. van der Zwaag, Dilo A., et al. Roads: A road pavement monitoring system for anomaly detection using smart phones. *Big Data Analytics in the Social and Ubiquitous Context*, Springer, pages 128–146, 2016.
- [60] Gareth J., Daniela W., Trevor H., et al. An introduction to statistical learning with applications in python. *Springer*, 1, 2023.
- [61] Raslan E., Mohammed F. A., Mohammed Y. A., et al. Iot for measuring road network quality index. *Neural Computing and Applications*, (35):2927–2944, 2022.
- [62] Menant F., Martin J.-M., Meignen D., et al. Using probe vehicles for pavement monitoring: Experimental results from tests performed on a road network. *Transportation*

- Research Procedia*, 14(3):3013–3020, 2016.
- [63] Song R. and Li B. Surrounding vehicles’ lane change maneuver prediction and detection for intelligent vehicles: A comprehensive review. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):6046–6062, 2021.
  - [64] Engelbrecht J., Booysen M. J., Rooyen G. J., and et al. Survey of smartphone-based sensing in vehicles for intelligent transportation system applications. *IET Intelligent Transport Systems*, 9(10):924–935, 2015.
  - [65] Wickramaratne T., Garg V., and Bauer P. On the use of 3-d accelerometers for road quality assessment. *IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, Portugal*, pages 1–5, 2018.
  - [66] Deo N. and Trivedi M. M. Convolutional social pooling for vehicle trajectory prediction. *Proc. IEEE/CVF Conference Computer Vision Pattern Recognition Workshops*, pages 1468–1476, 2018.
  - [67] Bahram M., Hubmann C., Lawitzky A., et al. A combined model- and learning-based framework for interaction-aware maneuver prediction. *IEEE Transactions on Intelligent Transportation Systems*, 17(6):1538–1550, 2016.
  - [68] Villanueva C. J., Zapata C. D., Saenz T. F., et al. Vehicle maneuver detection with accelerometer-based classification. *Sensors*, 16(1618):1–23, 2016.
  - [69] Johnson D. A. and Trivedi M. M. Driving style recognition using a smartphone as a sensor platform. *14th Int. Conf. on Intelligent Transportation Systems (ITSC)*, pages 1609–1615, 2011.
  - [70] Mokammel H., Supriya S., and Ali A. D. Driving maneuver classification from time series data: a rule based machine learning approach. *Applied Intelligence*, 52:16900–16915, 2022.
  - [71] ERA Paints. How to find your tesla paint code. <https://www.google.com/url?sa=i&url=https%3A%2F%2Ferapaints.com%2Fhow-to-find-your-paint-code%2Ftesla%2F&psig=AOvVaw2QcPSEsBkVrHQZYrnnv-b60&ust=1759573410313000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhXqFwoTCIiT6PPnh5ADFQAAAAAdAAAAABAW>, n.d. Accessed: 2025-10-03.
  - [72] TE Connectivity. 805m1-0020 : Meas vibration sensors specification.
  - [73] TDK Corporation. Six-axis (gyro + accelerometer) mems motiontracking device specification.
  - [74] Waveshare Electronics. Sim7600e-h 4g hat - waveshare wiki and specification.
  - [75] Raspberry Corporation. Raspberry pi 4 model b specification.
  - [76] Kamingu G. Python vs. r for data scientists: An introduction. 2023.
  - [77] Sivaraman S. and Trivedi M. M. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1773–1795, 2013.
  - [78] Castignani G., Derrmann T., Frank R., and et al. Smartphone-based adaptive driving

- maneuver detection: A large-scale evaluation study. *IEEE Transactions on Intelligent Transportation Systems*, 18(9):2330–2339, 2017.
- [79] Jing C. and Jian H. Svm and pca based fault classification approaches for complicated industrial process. *Neurocomputing*, (167):636–642, 2015.
- [80] Zebari R. R., Abdulazeez A. M., Zeebaree D. Q., and et al. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(2):56–70, 2020.
- [81] Chawla N. V., Bowyer K. W., Hall L. O., and et al. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.
- [82] Fernandez A., Garcia S., Herrera F., and et al. Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, 61:863–905, 2018.
- [83] James G., Witten D., Hastie T., and et al. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *An Introduction to Statistical Learning*, pages 374–393, 2023.
- [84] Yuan M., Nianhua Y., Zhilin Q., and et al. What makes an online review more helpful: An interpretation framework using xgboost and shap values. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(3):466–490, 2021.
- [85] Fei C., Guang C., Wen Z. T., and et al. Quantitative diagnosis of rotor vibration fault using process power spectrum entropy and support vector machine method. *Shock and Vibration 2014*, 2014.
- [86] Anitescu C., Atroshchenko E., Alajlan N., and et al. Artificial neural network methods for the solution of second order boundary value problems. *Computers, Materials and Continua*, 59(1):345–359, 2019.
- [87] Truong C., Oudre L., Vayatis N., and et al. Selective review of offline change point detection methods. *Signal Processing*, 167(107299):345–359, 2020.
- [88] Xingjiao W., Luwei X., Yixuan S., and et al. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, 135(107299):364–381, 2022.
- [89] Smith A. G., Peterson J., Campos C. T., and et al. Rootpainter3d: Interactive-machine-learning enables rapid and accurate contouring for radiotherapy. *Med Phys*, 49(1):461–473, 2022.
- [90] Samreen N., Agib A., Sania A., and et al. An unsupervised machine learning algorithms: Comprehensive review. *IJCDS Journal*, 13:911–921, 2023.
- [91] Zhu L., Jacob R. H., and Jeffrey D. G. Trajectory segmentation map-matching approach for large-scale, high-resolution gps data. *Transportation Research Record: Journal of the Transportation Research Board*, (2645):67–75, 2017.
- [92] Yin Z., Goldberg D., Hammond T., and et al. A probabilistic framework for improving reverse geocoding output. *Transactions in GIS*, 24:656–680, 2020.
- [93] Arsanjani J. J., Zipf A., Mooney P., and et al. Openstreetmap in giscience: experiences, research, and applications. lecture notes in geoinformation and cartography.

*Springer*, 2015.

- [94] Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, New York, 2 edition, 1988.

## Acknowledgement

I would like to thank my parents and my brothers (Parents: Bernadett Szép, Zoltán Nagy; Brothers: Noémi Nagy, Marcell Nagy) for their selfless and honest support during my whole life, without them I am sure I would not write these lines now. Grateful thanks to my sweetheart (Patricia Pintér) for making me happy every day. Without your patience and honesty I would not be here.

There are no words for my gratitude to my supervisor (István Szalai) who has constantly supported my progress and development, guiding and inspiring me with his expertise; therefore, I just would like to thank you everything!

I would like to thank my colleagues in the Mechatronics Research and Education Institute, Zalaegerszeg for their support, and I would like to thank my co-authors their valuable time and work.