HÖGSKOLAN
DALARNA

Dalarna Doctoral Dissertations 29

Objective Assessment of Loose Gravel Condition using Machine Learning with Audio-visual Observation

NAUSHEEN SAEED

Microdata Analysis
School of Information and Engineering
Dalarna University, Borlänge, Sweden
2024

Dissertation presented at Dalarna University to be publicly examined in room Clas Ohlson, campus Borlänge, Thursday, 28 March 2024 at 13:00 for the Degree of Doctor of Philosophy. The examination will be conducted in English. Opponent: Professor Niklas Lavesson (Blekinge Institute of Technology).

Abstract
Saeed, N. 2024. Objective Assessment of Loose Gravel Condition using Machine Learning with Audio-visual Observation. *Dalarna Doctoral Dissertations* 29. Borlänge: Högskolan Dalarna. ISBN 978-91-88679-59-8.

A well-maintained road network is essential for sustainable economic development, providing vital transportation routes for goods and services while connecting communities. Sweden's public road network includes a significant portion of gravel roads, particularly cost-effective for less populated areas with lower traffic volumes. However, gravel roads deteriorate quickly, leading to accidents, environmental pollution, and vehicle tire wear when not adequately maintained. The Swedish Road Administration Authority (Trafikverket) assesses gravel road conditions using subjective methods, analysing images taken during snow-free periods. Due to cost constraints, this labour-intensive process is prone to errors and lacks advanced techniques like road profilometers.

This thesis explores the field of assessing gravel road conditions. It commences with a comprehensive review of manual gravel road assessment methods employed globally and existing data-driven smart methods. Subsequently, it harnesses machine hearing and machine vision techniques, primarily focusing on enhancing road condition classification by integrating sound and image data.

The research examines sound data collected from gravel roads, exploring machine learning algorithms for loose gravel conditions classification with potential road maintenance and monitoring implications. Another crucial aspect involves applying machine vision to categorise image data from gravel roads. The study introduces an innovative approach using publicly available resources like Google Street View for image data collection, demonstrating machine vision's adaptability in assessing road conditions.

The research also compares machine learning methods with manual human classification, specifically regarding sound data. Automated approaches consistently outperform manual methods, providing more reliable results. Furthermore, the thesis investigates combining audio and image data to classify road conditions, particularly loose gravel scenarios. Early feature fusion using pre-trained models significantly improves classifier accuracy.

The research proposes using cost-effective devices like mobile phones with AI applications attached to car windshields to collect audio and visual data on gravel road conditions. This approach can provide more accurate and efficient data collection, resulting in real-time mapping of road conditions over considerable distances. Such information can benefit drivers, travellers, and road maintenance agencies by identifying problematic areas with loose gravel, enabling targeted and efficient maintenance efforts, and minimising disruptions to traffic flow during maintenance operations.

Keywords: Gravel road condition assessment, Loose gravel, Sound analysis, Machine learning, Image analysis, Audio analysis, Image and audio data fusion

Nausheen Saeed, Microdata Analysis

ISBN 978-91-88679-59-8
urn:nbn:se:du-47915 (http://urn.kb.se/resolve?urn=urn:nbn:se:du-47915)

Acknowledgement

I want to express my heartfelt gratitude to my cherished family, with special mention to my husband, Farrukh Nasim, and my children, Mahad Farrukh and Mohid Farrukh. Their unwavering support, encouragement, and love have been of great value throughout my PhD journey. Their belief in me and my dreams has been the driving force behind my success.

I am thankful to my parents for instilling in me the enduring value of education and wholeheartedly supporting my pursuit of my dreams. Without their unwavering support and invaluable guidance, I would not have reached this significant milestone in my academic journey.

I extend my gratitude to my research advisors, Moudud Alam and Roger G. Nyberg, who have played a vital role in shaping my research expedition. Their expertise, guidance, and firm technical support have been pivotal in my growth as a researcher. I would also like to acknowledge my former supervisor, Mark Dougherty, for his support.

I am immensely grateful to my colleagues, who have been a constant source of inspiration and knowledge sharing. Their feedback, support, and insightful discussions during seminars, teaching, and informal interactions have kept me motivated and focused, especially during challenging research phases and tight deadlines.

The collective support of my family, parents, research advisors, and colleagues has been instrumental in helping me achieve this significant academic milestone.

List of papers

I. **Saeed, Nausheen**., Dougherty, M., Nyberg, R G., Rebreyend, P., Jomaa, D." A Review of Intelligent Methods for Unpaved Roads Condition Assessment", November 2020,15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Kristiansand, Norway.

II. **N Saeed**, RG Nyberg, M Alam, M Dougherty, D Jooma, P Rebreyend Classification of the Acoustics of Loose Gravel, Sensors 21 (14), 4944.

III. **N Saeed**, Alam, M., RG Nyberg, Automatic detection of loose gravel condition using acoustic observations. Under review in the Journal of Road Materials and Pavement Design.

IV. **N Saeed**, RG Nyberg, M Alam, Gravel Road classification based on loose gravel using transfer learning. International Journal of Pavement Engineering, 2022, 1-8

V. **N Saeed**, Alam, M., RG Nyberg A Multimodal Deep Learning Approach for Gravel Road Condition Evaluation through Image, and Audio Integration. Transportation Engineering, Volume 16, 2024, 100228, ISSN 2666-691X.

Reprints were made with permission from the respective publisher.

List of Papers Not Included in the Thesis.

I. **Saeed, N.,** Alam, Moudud., Nyberg, R G., Rebreyend, P., Jomaa, D., "Comparison of Pattern Recognition Techniques for Classification of the Acoustics of Loose Gravel", 7th IEEE International Conference on Soft Computing & Machine Intelligence (ISCMI), 2020.

II. Zhang, F., **Saeed, N**., & Sadeghian, P. (2023). Deep Learning in Fault Detection and Diagnosis of building HVAC Systems: A Systematic Review with Meta Analysis. Energy and AI, 100235.

Contents

1. Introduction

1.1. Background

Gravel roads, also called unsurfaced or unpaved roads, do not have cement, asphalt, concrete, or similar materials in their construction. The low traffic volume on these secondary roads makes surface treatment uneconomical. The annual average daily traffic on gravel roads is around 100 vehicles per day at an average speed. Maintenance activities are required three to four times a year. In Sweden, every year, road maintenance agencies plan and schedule road maintenance activities (Edvardsson 2009)

In addition to the common paved roads, gravel roads are important to road monitoring agencies and society. Gravel roads are a cost-effective choice for areas with low traffic volume and are used for transportation and the movement of forest goods and agricultural produce (Mbiyana et al., 2022). Gravel roads require maintenance, which is distinct from paved roads. It requires periodic grading, treatment of road surface with crushed limestone or gravel, ditch maintenance, removal of vegetation that may cause impairment to visibility, and salt and water treatment for dust control. Road condition assessment must be performed on time for safe and efficient transportation. Keeping track of the condition deterioration of gravel roads is a challenging task. Therefore, this causes delays in the timely rectification of defects (Huntington and Ksaibati 2011; Saeed et al. 2020).

In Sweden, 21% of public roads and 74,000 km of private sector-owned roads are gravel roads (Kans et al., 2020) . Gravel roads consist of one or more layers of soil and aggregates, with a designed shape to prevent water accumulation (see Figure 1). The most important aspect of a gravel road is the ditches that allow water drainage. The primary material used in the construction of gravel roads is aggregate or crushed stone, which provides drive comfort and reduces stress on the subgrade. However, aggregate deteriorates over time due to weather and traffic, leading to loose gravel and rut formation. Gravel road performance depends on roadway surface composition, sub-surface soil conditions, weather conditions, traffic volumes, drainage, stabilisation practices, and maintenance practices.

Loose gravel is a vital assessment factor, as too much loose gravel can be dangerous. Loose gravel can slip from under the tires and cause drivers to lose control, resulting in accidents. In 2017, around 600 people died in accidents on gravel roads in the United States. The causing defects were found to be loose gravel, dust impairing visibility, and lack of traffic signs (Albatayneh et al. 2020). To prevent fatalities, it is critical to promptly address the issue of loose gravel. This thesis is dedicated to the detection of loose gravel. Several solutions can improve gravel road safety, such as a well-compacted road surface, a surface seal that creates a hard water resistance surface, and well-planned maintenance activities. Gravel road maintenance requires regular blading, treating the surface with crushed limestone or gravel, ditching, removal of vegetation, and dust control treatment (Trafikverket 2014).

There are various approaches to assessing the condition of gravel roads worldwide. Visual assessment is the most common but has limitations, such as being time-consuming and subjective. Quantitative measurements are needed to eliminate the human factor. Specialised vehicles equipped with a laser profilometer can help achieve a quantitative assessment. However, this method is not always feasible due to its high cost and operational complexity. The Swedish Road Administration Authority (Trafikverket) rates the gravel road condition during summertime when the roads are free of snow, based on the severity of irregularities such as corrugation and potholes, dust, and loose and gravel cross-section.

1.1.1. Construction of gravel road

The gravel roads are crown-shaped, meaning the gravel road's centre is slightly higher than the edges (Figure 1). This helps water drainage from the road into the ditches along the gravel roads. This prevents moisture and erosion from gravel roads. The structure of gravel roads consists of several layers of material, including a subgrade layer and a surface later. The subgrade layer consists of natural soil on which the road is laid. The base layer comprises layers of stones to provide stability and strength to the road. Figure 1 shows the shape and structure of a gravel road that needs to be maintained. Finally, the surface layer comprises smaller stones to provide a smooth driving surface. Proper construction of gravel roads according to standards provides stability and safety on a gravel road.

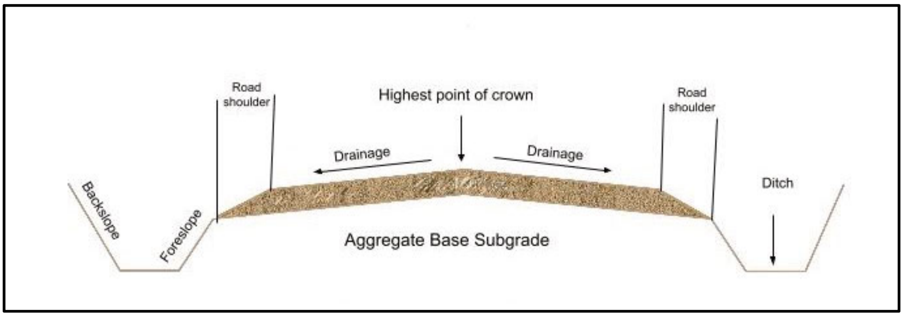

Figure 1 Cross-section structure of a gravel road section (Saeed et al., 2022)

The research on sound and image recognition systems has gained momentum and has been used in a wide range of applied fields, including road texture classification. These systems can detect the amount of loose gravel on the road through sound. They offer a cost-effective and reliable alternative to other assessment methods, such as using specialised vehicles equipped with a laser profilometer. Sound recognition systems can help improve gravel road maintenance by detecting and correcting defects before they become hazardous. Overall, the challenges of maintaining gravel roads can be overcome with the right techniques and technologies, ensuring safe and reliable transportation.

1.1.2. Traditional gravel road assessment methods

Various visual methods are used worldwide to assess the condition of gravel roads, which are closely related to the weather and landscape of each country. Some of the well-known techniques include the U.S. Army Corps of Engineering (USACE) assessment system, Pavement Condition Index (PCI), and Pavement Surface Evaluation and Rating (PASER)(Gregory D. Cline 2003; Robert and Ronald 1992; Walker, Entine, and Kummer 1987) Municipalities rely on visual assessments and statistical data from specialised instruments such as laser profilers, ARAN, and ROMDAS in Sweden, Canada, and New Zealand (Sodikov, Tsunokawa, and Ul-Islam 2005).

The Swedish Transport Administration (Trafikverket) employs the "Bedömning av grusväglag" (Assessment of Gravel Road Conditions) regulation for evaluating the state of gravel roads (Trafikverket, 2014). The regulation stipulates subjective ratings and is used to gather information about road conditions to help decision-makers decide whether maintenance is necessary. Photos of the gravel roads are taken while driving, and experts examine them along with a textual description that assesses the severity of loose gravel in a

4-point grading scale shown in Figure 2 and Table 1. While driving, photographs of gravel roads are taken. These photos, along with a written assessment, are analysed by experts. The assessment focuses on determining the extent of loose gravel using a four-tier grading system, as depicted in Figure 2 and Table 1. Following this assessment, the road sections under observation are categorised in accordance with the standards of Trafikverket.

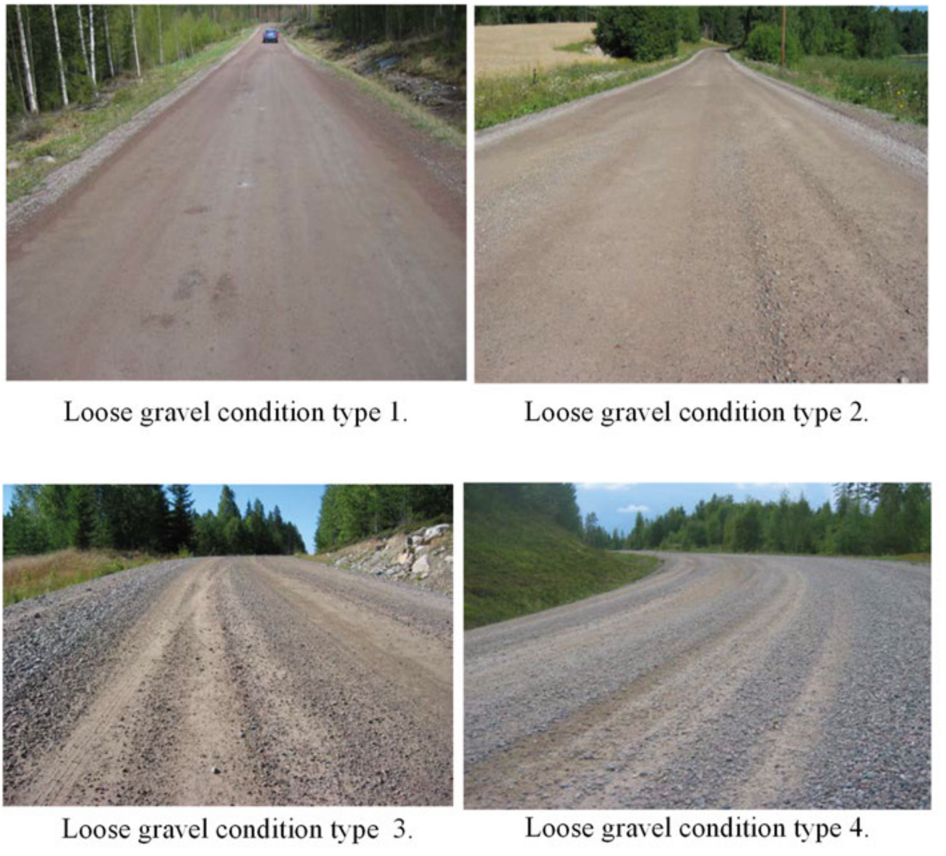

Figure 2 Sample pictures of loose gravel conditions and corresponding grade. Type 1 shows a good road, and Type 4 shows a gravel road with the worst condition (Trafikverket 2014)

Table 1 Textual description of grading of loose gravel conditions described by the Swedish Transport Administration (Trafikverket 2014)

Gravel Road Evaluation Grade	
Grade	**Description**
1	Loose gravel does not exist on the road or only exists to a small extent along the route.
2	Loose gravel exists to a smaller extent on the road and smaller embankments along the edges of the road, not affecting driving comfort or safety noticeably.
3	Loose gravel exists on the road and in smaller embankments along the road edges, affecting driving comfort and safety.
4	Loose gravel exists to a large extent over the road and in pronounced embankments along the road edges, affecting driving comfort and safety to a significant degree.

1.2. The aim and research questions

The study aims to explore machine learning solutions for objectively assessing loose gravel conditions on gravel roads by employing innovative data collection techniques involving sound and image data. This research seeks to enhance the efficiency and precision of gravel road maintenance, leading to improved road safety and socio-economic benefits.

Applications like Roadroid and Roadsense, available on smartphones, offer users information about overall road roughness (Allouch et al. 2017; Forslöf and Jones 2015). However, it's essential to point out that these apps do not conform to the guidelines set by road maintenance agencies, such as Trafikverket. Information about distress types is critical for road maintenance agencies to plan maintenance activities. The thesis aims to address the issue of automatic classification of gravel roads based on loose gravel. As publicly available data on this topic was limited, the thesis includes the creation of datasets for each study. Two important human senses are hearing and vision, allowing us to perceive our environment and make decisions accordingly. Humans can readily classify different sounds: music playing, truck engines running, babies crying, people talking, etc. Machine hearing is the research area in which machines recognise and interpret sound (Lyon 2010). Sound classification has proven helpful in many research areas, such as bird sound classification for early fire warning, detection of lung disease, heart sound signal classification, prevention of illegal deforestation, basement car crash detection and other research areas (Effendy et al. 2022; He et al. 2021; Nguyen and Pernkopf 2022; Permana et al. 2022; Sammarco and Detyniecki 2018).

Similarly, image classification has many applications, such as food inspection, lung and breast cancer detection, COVID-19 detection through lung image classification, dust on gravel roads detection, etc. (Albatayneh et al., 2019; Das & Chandra, 2022; Gumaei et al., 2022; Hameed et al., 2020). This thesis proposes a novel approach for assessing loose gravel conditions on gravel roads by investigating the combination of signal processing, image analysis and machine learning techniques (Purwins et al. 2019). Data was gathered utilising unconventional methods such as capturing the sound of gravel hitting the bottom of the car as well as recording videos while driving on the gravel road in Dalarna, Sweden.

The research questions addressed in this thesis are as follows.

1. How are gravel roads evaluated in Sweden and globally, including the measurement techniques and methods used for assessment? What is the current state of automated methods and research for detecting defects on gravel roads?
2. What is the potential of machine learning algorithms in classifying loose gravel based on sound and image data, and what level of accuracy can be achieved?
3. How effective are pre-trained deep learning-based models, such as VGG-16 with transfer learning, in accurately rating the condition of gravel roads based on images? What other data sources could be utilised?
4. Is it feasible to employ audio data and machine learning algorithms to classify gravel roads based on loose gravel conditions into four classes following the standard by the Swedish Transport Administration Trafikverket?
5. How can the fusion of audio and image data be leveraged to enhance the detection of loose gravel conditions?

1.3. General scientific contribution

This thesis has contributed to microdata analysis by introducing innovative data collection techniques for defect detection on gravel roads. The research in the thesus unique as it is the first to utilise sound and image data to detect loose gravel conditions. While extensive work has been done on objective detection defects on paved roads, there has been little research on automated assessment of gravel road conditions, making this study contribute to the research area.

The thesis includes a literature review, which was previously missing in this field, providing readers with an understanding of traditional and data-driven methods available worldwide. As the research project progressed, the studies discussing gravel roads were presented in the subsequent thesis Papers, keeping readers up to date with the research in the same domain. This thesis also emphasises the importance of sound data and highlights its benefits in solving many research problems.

The thesis also suggests using easy-to-access tools like Google Street View to increase data sets. By leveraging this idea, researchers could tackle multiple research problems.

This thesis contributes to the existing literature by exploring machine learning methods for objectively assessing loose gravel for gravel road condition assessment, which have not been addressed previously. Moreover, it compares the effectiveness of various machine learning algorithms in detecting loose gravel and classifying gravel road conditions, offering valuable insights into the advantages and limitations of each approach.

Another contribution of this thesis is the development of a labelled dataset for gravel road classification based on sound and video recordings, which provides a valuable resource for future studies and research in this field. Upon request, we can share the data set. The most important contribution of this thesis is that road maintenance agencies can utilise the results to create applications that can improve decision-making processes in gravel road maintenance and management, facilitating targeted interventions and more efficient allocation of resources.

This thesis contributes to sustainable infrastructure development by providing a more efficient and targeted approach to gravel road maintenance. There are several social benefits of timely road condition detection, such as improved road safety through timely rectification of gravel road defects, which can help prevent accidents and fatalities on the road and reduce vehicle maintenance costs. Well-maintained roads can improve the transportation of goods and positively impact the economy and environment.

2. Methods

The research project began with a review of previous studies, with Paper I providing insights into global gravel road condition assessment methods, highlighting the prevalence of visual assessment, which relies on field experts and has inherent subjectivity. The paper also discusses available smart assessments methods for gravel road evaluation. Building on this, Paper II emphasized the importance of sound data in detecting gravel road conditions. It explored machine learning algorithms to classify gravel vs. non-gravel audio, demonstrating the potential of sound data utilization, particularly with GoogLeNet's standout performance.

Paper III addressed the feasibility of classifying loose gravel conditions by combining image and sound data. It introduced innovative data extraction from Google Street View, showcasing the effectiveness of pre-trained VGG-16 models with transfer learning. Paper IV focused on assessing loose gravel conditions using audio data, revealing the strong performance of machine learning algorithms, with Multi layer perceptron (MLP)as the top performer. Lastly, Paper V presented a data fusion approach combining audio and image data. This approach can enhance prediction robustness and reliability, even when one modality lacks.

The following sections present detailed discussions on data collection, processing, reliability assessment, machine learning algorithms, and the employed evaluation metrics used in the thesis.

2.1. Data collection, data processing and reliability assessment

In the absence of an established open dataset comprising images and audio for gravel roads, it became essential to initiate a data collection process. The data collection process involved strategically placing two HERO7 GoPro cameras from GoPro Inc., one inside the car and the other on the car's bonnet. This setup aimed to capture audio and video data during the summer seasons of 2020 and 2022 along gravel roads in Dalarna, Sweden. Recordings were conducted at a constant speed of 50 km/h under dry and sunny weather conditions.

The choice of using both internal and external cameras was intentional. The internal camera was suitable for capturing distinct gravel sounds hitting the car's bottom, while the external camera provided a comprehensive visual assessment of the gravel road conditions. The decision to utilise video recording, as opposed to relying solely on acoustic equipment, was motivated by the aim of observing both auditory and visual road conditions. This choice contributed valuable audio data and provided essential image data for the research project.

Segments found irrelevant to the analysis of gravel road conditions, such as travel to the road, turning the car, varying speeds, and conversations, were excluded. The dataset, comprising 15 videos with a combined duration of 1 hour, 13 minutes, and 54 seconds, was used to investigate the state of gravel roads by analysing the audio and video recordings it contained. This dataset played a crucial role in the research conducted for papers 2 through 5 of this thesis.

Audio and image data were extracted during pre-processing, and Roboflow's Annotation Tool played a crucial role in creating bounding boxes and isolating road sections in images. This meticulous approach ensured the generation of a focused dataset that retained essential aspects of gravel roads while excluding irrelevant elements such as sky and vegetation.

Paper 2 utilised a total of 237 audio clips, comprising 133 gravel sounds (56%) and 104 non-gravel sounds (44%). Paper 3 used a dataset of 638 images extracted from video data collected from gravel roads in Dalarna and additional data from Google Street View by traversing through gravel roads in Sweden. These images were categorised into two classes: 1&2 and 3&4.

Paper 4 provided details of the dataset and class distributions, where Class 1 had 203 instances, Class 2 had 285 instances, Class 3 had 241 instances, and Class 4 had 140 instances, totalling 869 instances. Each recording is divided into five-second parts, as a five-second length provides enough time to hear the gravel hitting the bottom of the car. Based on the amount of loose gravel observed in the videos, each segment is manually assigned a class or grade, ranging from 1 to 4, as defined by Trafikverket (Trafikverket, 2014). The loose gravel conditions might change throughout each section of the road, as seen in the videos. However, even within five seconds, loose gravel conditions might change. As a result, the most prevalent class identified inside each segment is allocated to that section.

The reliability assessment of the classification involved a meticulous process to ensure the accuracy and consistency of the results. Manual classification by two listeners, using the Gravel Road Assessment document from Trafikverket as a reference, was performed on all four types of conditioned gravel roads (1-

4). BORIS, a free and versatile event-logging software, ensured consistency in the manual classification process (Friard & Gamba, 2016). The listeners labelled each video segment with Grades 1 to 4, representing the different conditions of gravel roads. Cross-checking between the classifications of individual authors and joint review of segments where differences in class assignments occurred contributed to the refinement of the final classification.

The reliability of the manual classification process was assessed by comparing the results with evaluations from two additional individuals using Fleiss' Kappa measurement. Fleiss' Kappa is a statistical measure that gauges the level of agreement between multiple raters beyond chance. This analysis ensures the consistency and reliability of the classification results, enhancing their robustness for subsequent analysis and interpretation (Kyros & Elias Myrén, 2022).

Paper 5 presented a dataset with a total of 885 instances, divided into two classes, 1&2 and 3&4, for both images and audio. Each class had 487 instances in both images and audio. Roboflow (Lin et al., 2022) was used for image annotation. Roboflow was useful in detecting and segmenting gravel roads, excluding vegetation. This step was crucial to ensure that algorithms only focused on learning features related to road conditions during classification.

A conversion process transformed audio files into spectrograms for audio data. Using the Short-Time Fourier Transform (STFT) technique. The data was labelled by referring to images rated from condition 1 to 4 by Trafikverket's manual, where 1 signifies a good road, and 4 indicates the worst condition road in terms of loose gravel.

2.2. Machine Learning Algorithms

Classical and machine learning algorithms such as Support vector machines, Decision trees, ensemble methods, and deep learning methods such as Transfer learning through pre-trained networks such as GoogLeNet, Residual-net(ResNet), and Alexnet were trained and tested in various studies included in this thesis. Details of these algorithms can be found in standard textbooks (Bonaccorso, 2017, 2018; Yang et al., 2020) , and their implementational specifications are in the papers attached (see Section 8).

2.3. Experimental evaluation of the merits of selected machine learning methods

In the conducted studies included in this, various metrics, including F1-score, recall, and precision, were utilised to evaluate classifier performance. Accuracy, defined as the percentage of correctly classified testing data, is a fundamental metric. In this research, the dataset exhibited an imbalance, characterised by a higher occurrence of road conditions falling into categories 3 and 4 and a relatively lower representation of categories 1 and 2. Addressing this data imbalance was essential in evaluating the effectiveness of classification, focusing specifically on precision, recall, and F1-score, as they are key metrics for gaining valuable insights into classification performance.

- **Accuracy:** Evaluates correctness by considering correctly classified images' ratios to the total testing images.
- **Precision:** Quantifies the exactness of the model, representing the ratio of correctly classified positive class images.
- **Recall:** Measures completeness, computing the ratio of correctly classified positive class images out of the total true class images.
- **F1-score:** Represents the harmonic average of precision and recall, commonly used to balance precision and recall optimisation.
- **Error Rate:** Proportion of instances misclassified, complementing accuracy.

Collectively, these metrics provide a comprehensive assessment of the model's correctness, exactness, completeness, and the trade-off between precision and recall in imbalanced data scenarios.

2.4. Outline for the remainder of the thesis.

Paper I addresses research question 1, shedding light on global gravel road assessment methods, particularly the prevalent subjective visual assessment. It underscores the need for cost-effective and objective solutions, laying the groundwork for subsequent studies. Research question 2 is thoroughly examined in Papers II to V, focusing on the capability of machine learning algorithms to classify gravel road conditions. Paper II emphasizes the importance of sound data and showcases GoogLeNet's success in classifying gravel vs. non-gravel audio. Paper III tackles research question 3, exploring the feasibility of classifying loose gravel conditions through image and sound data. Innovative techniques using Google Street View images and pre-trained VGG-16 models with transfer learning provide high accuracy. Paper IV addresses research question 4, evaluating loose gravel conditions using audio data. Lastly, Paper V answers research question 5, presenting a holistic approach that integrates audio and image data to identify loose gravel conditions. This approach is adaptable to various road defects and enhances resource allocation efficiency while reducing environmental impact. The following section examines the included papers and their respective themes.

The subsequent sections of the thesis are organised as follows: Section 3 discusses the included papers and their respective themes. Section 4 discusses the thesis from a microdata perspective, Section 5 encompasses the discussion and conclusion, and Section 6 addresses the limitations of the thesis and outlines potential future work. Section 7 includes references, and finally, in 8 are the Papers included in this thesis.

3. Themes Across the Included Papers

3.1. Review

Paper I thoroughly examines classical road profiling methods and explores objective approaches employing sensor data for assessing road surfaces, with a particular emphasis on gravel road maintenance. The study examines two main methodologies, namely threshold-based approaches, and machine learning algorithms, in terms of their roles in data acquisition, pre-processing, and analysis to strengthen decision-making processes. The literature indicates the use of various data types and automated techniques for road evaluation, resulting in the emergence of two primary categories, namely threshold-based and machine learning-based methods.

3.1.1. Threshold-Based Method

Automated systems commonly employ threshold-based methods to identify road anomalies, utilising data from sensors like a vehicle's accelerometer or a smart device. These methods analyse changes in data patterns, often assessed through statistical values such as standard deviation [36]. The standard deviation, a critical parameter in detecting road anomalies from accelerometer sensor data, plays a key role in these methods (Wang et al., 2015). Commercial smartphone applications, including Roadroid, Road Bounce, and RoadSense, utilise threshold-based approaches to measure road roughness. These applications, compatible with Android or Windows portable devices, enable continuous assessment of road pavement conditions as a vehicle travels at an average speed. Using phone GPS, they calculate the International Roughness Index (IRI) by correlating registered road vibrations and aligning the data based on longitude, latitude, and altitude.

3.1.2. Machine Learning Methods

Numerous studies explore road roughness evaluation through machine learning techniques, encompassing supervised and unsupervised learning applied to data from sensors such as accelerometers, gyro meters, GPS, images, and sound. While most of these studies focus on paved road defects, a few extend their application to unpaved roads, including gravel roads(Cord & Chambon, 2012; Oppong Adu et al., 2023; Zhang & Elaksher, 2012). Methods developed for paved roads, such as overall road roughness calculations, can also be adapted for assessing gravel roads.

This study showed that from the existing literature, it is evident that current methodologies often concentrate on measuring specific distress types in gravel roads rather than considering a comprehensive approach encompassing all possible issues. Future advancements could benefit from a synergistic combination of diverse techniques, such as machine vision and in-car sound analysis, to establish a holistic system capable of accurately detecting and measuring various distress types in gravel roads, including loose gravel, dust, or other defects. This comprehensive approach would enable the development of more precise and efficient maintenance plans by explicitly identifying distress types and optimising cost and time.

The primary contribution of this review lies in providing an overview of ongoing research on cost-effective automated methods for road evaluation, with a particular emphasis on identifying research gaps related to unpaved roads. Incorporating participatory sensing is a promising avenue for gravel road evaluation. Utilising data from sensors like accelerometers and additional information such as images and sound from gravel roads offers potential ways to identify and address gravel road defects. While participatory sensing may not entirely replace traditional methods, it presents an opportunity to reduce the necessity for frequent visits and manual measurements, thereby enhancing efficiency in the evaluation process.

The paper also introduces globally employed classical gravel road rating systems that assess parameters like road cross-section, drainage, gravel gradation, dust, and distress types such as corrugations, erosion, loose gravel, and potholes. Traditional visual assessment methods involving expensive and cumbersome laser profiler-equipped trucks are discussed, highlighting their limitations, especially on busy roads.

Various systems are detailed, including the U.S. Army Corps of Engineering's Unsurfaced Road Condition Index (URCI), the Pavement Condition Index (PCI) utilising vehicle-mounted sensors, Pavement Evaluation and Rating (PASER), Gravel Road Management System (GRMS) in Wyoming using windshield surveys, and the Swedish National Road Transportation agency's (Trafikverket) Gravel Road Assessment (Gregory D. Cline, 2003; Huntington & Ksaibati, 2010; Trafikverket, 2014; Walker et al., 1987).

The Swedish method, known as "Bedömning av grusväglag" (Gravel Road Assessment), is mainly used in Sweden, where gravel roads constitute 21% of public roads. This method employs a distinctive approach, assessing distress types such as loose gravel, roughness, and dust through images taken from the gravel road. An expert subjectively decides to categorise the road in particular road conditions. It classifies roughness collectively, while rutting is measured separately, and loose gravel and dust are considered binding ability indicators.

Data collection occurs in a moving vehicle, selecting two random 100-meter sections over a 10 km stretch during snow-free periods from May to October. The evaluation classifies roads into three surface condition categories (Good, Acceptable, Poor) and three Average Annual Daily Traffic (AADT) classes (A, B, C). The method emphasises maintaining satisfactory conditions, allowing for specified durations of poor conditions.

The Swedish Gravel Road Assessment method considers various factors, offering valuable insights for regions with similar weather conditions.

3.2. Sound-based detection of the loose gravel condition.

The sound of gravel hitting the car base can contribute to the automated detection of loose gravel road conditions, which was investigated in studies presented in papers II and III. Paper II outlines a preliminary study designed to test whether machine learning algorithms could differentiate or detect distinct sounds from gravel roads. The dataset was segregated into binary, gravel, and non-gravel classes to assess the performance of machine-learning classification models. Positive outcomes from paper II affirmed the discernibility of different sounds originating from various sections of gravel roads, each associated with specific loose gravel conditions. In paper III, audio clips collected from gravel roads were categorised into the four conditions defined by Trafikverket, further supporting the hypothesis regarding sound-based detection of loose gravel road conditions. Paper II and Paper III are discussed below.

Artificial intelligence aims to enable computers to emulate human decision-making processes. Human hearing, a vital sense, facilitates environment perception and decision-making based on sound classification, such as music, engines, cries, and conversations. This research investigates machine hearing and its potential application in introducing an innovative approach to evaluate loose gravel on gravel roads.

Loose gravel poses risks like slipping under the tyre, causing loss of control, especially during quick turns. Drivers must exercise caution, avoiding abrupt braking or swerving on gravel roads with abundant loose gravel. Loose aggregate results from factors like heavy traffic and substandard materials, forming linear berms as gravel moves from the wheel path. These berms can extend up to four to six inches along the roadside. Over time, loose gravel accumulates longitudinal depressions, known as ruts or wheel tracks, with widths varying based on wheel size (Robert & Ronald, 1992; Skorseth, 2000). Research on assessing gravel road conditions is less common compared to paved roads. Some studies explore assessment methods using sensors, including those found in smartphones. For example, the Asfault system proposed by (Souza et al., 2018) classifies paved road conditions using smartphone accelerometer sensors and achieves 92.39% accuracy on a four-tier classification scale. Another study by (Cabral, 2019) employs various smartphone sensors to distinguish between paved and unpaved roads, identifying obstacles like potholes, yet lacks specificity on the nature of road defects. This study uniquely contributes by utilising exclusively collected audio data from driving on gravel roads to train machine learning algorithms for loose gravel classification. It follows the Swedish Transport Agency's four-tier grading scale. The hypothesis suggests that humans can discern gravel road conditions by listening to gravel hitting the car bottom, providing a unique dataset for machine learning algorithms. Applications employing this method can assist road maintenance agencies in efficient, cost-effective, and objective road assessments, enabling the identification of gravel roads with maintenance needs (Grade 1 to 4) for timely interventions. The study employs video recordings, including audio, manually classified into 5-second segments based on the Swedish Transport Administration's grading. Three machine learning algorithms—Multi-layer Perceptron, Random Forest, and Support Vector Machine—are tested for classifying road segments using extracted audio features.

While research has extensively examined paved road conditions, the focus on unpaved roads, particularly gravel roads, has been limited. Noteworthy exceptions include studies on corrugation assessment (Abu Daoud et al., 2021) and gravel sound classification (Saeed et al., 2021). However, the former focuses on the amount of corrugation. The latter, our former study in the same research project, was an attempt to evaluate machine learning algorithms for classifying gravel and non-gravel sounds (Saeed et al., 2021). Potholes and rutting distress dominate discussions on distress detection, leaving a gap in objective loose gravel measurement without visible work in this area. Also the results lack alignment with standardised road maintenance classification schemes. Despite potential of the broader use of audio data for road maintenance remains relatively underexplored in the literature (Kyros & Elias Myrén, 2022)

Although not included, a study conducted in relation to this thesis evaluated inter-rater agreements for manually classifying loose gravel conditions using audio and video data. Two human raters achieved a 64% average matching success rate, and additional raters improved reliability. Despite substantial agreement, the subjective nature of human ratings led to the proposal of machine learning algorithms for more objective classification. Using audio data alone for loose gravel condition classification provides an efficient approach to monitoring gravel roads, reducing the dependence on physical observation, manual image classification, and processing data from multiple sensors (Kyros & Elias Myrén, 2022).

Data collection for paper II involved driving at a speed of 50 km/h on gravel roads situated on the outskirts of the town of Borlänge and in the village of Skenshyttan in Sweden. The researchers used two GoPro HERO7 cameras (GoPro Inc, San Mateo, CA, USA) to capture the environment's auditory and visual elements. One of the cameras was securely mounted on the vehicle's bonnet using a robust double-clip suction cup and an unobstructed camera lens to ensure stability. The other camera was fixed to the windshield to capture the in-car audio of gravel hitting the bottom of the car. The deployment of two cameras aimed to assess whether recording audible gravel sounds from inside or outside the vehicle yielded superior results. The data collection involved two trips conducted under dry and sunny conditions, covering a total distance of 13 miles on the first trip and 10 miles on the second trip along gravel roads. This distance calculation excludes the travel time to reach the gravel road. Valuable data was extracted from 36 minutes of audio recording, excluding travel time to and from the gravel road. The researchers intentionally drove both within and outside the rutting-formed tracks on the gravel road to capture varying levels of audibility.

Audio extracting was extracted from mp4 video files and converted into the .wav format, with a sampling frequency of 44,100 Hz and a 16-bit per sample resolution. An audio pre-segmentation step was implemented, segmenting the continuous audio stream into 5-second portions To facilitate subsequent analysis. The labelled audio groups were then categorised into two classes: gravel sound and non-gravel sound. Gravel sound was identified by the audible impact of gravel hitting the car bottom, sourced from gravel roads, while non-gravel sound encompassed instances where gravel impact was either inaudible or infrequently occurred, primarily from specific road types.

For the classification task, a variety of supervised learning algorithms, including Support Vector Machines (SVMs), tree-based models, and ensemble algorithms, were trained. Feature extraction was performed using 79 spectral and frequency domain features, comprising parameters such as spectral centroid, amplitude, harmonics-to-noise ratio, mean frequency, and peak frequency. The signal processing stage involved converting the audio data from the time domain to the frequency domain using the Fast Fourier Transform (FFT). Short-Time Fourier Transform (STFT) was employed to analyse the frequency content over time. A Hamming window with a 50% overlap was applied to mitigate spectral leakages.

Feature selection was conducted through a t-test, aiming for dimensionality reduction and the identification of features that exhibited significant differences between the gravel and non-gravel classes. Principal Component Analysis (PCA) was also explored for dimensionality reduction, but ultimately, the t-test was chosen. Ten-fold cross-validation was utilised to assess the accuracy of each classification model. The study included 237 audio clips, comprising 133 gravel sounds (56%) and 104 non-gravel sounds (44%).

In addition to traditional classification methods, Convolutional Neural Networks (CNNs) were investigated for gravel sound classification. Gravel acoustics were transformed into spectrogram images using FFT. The pre-trained GoogLeNet, a 22-layer deep convolutional neural network, was fine-tuned for the classification task. Data augmentation techniques were applied to expand the dataset, including image resizing, horizontal flipping, and random rotation.

The audio recordings extracted from the study primarily captured the sound of gravel hitting the vehicle's bottom. These audio datasets were categorised into gravel and non-gravel sounds. Features were extracted, saved in a .csv file, and subjected to t-test analysis.

The identification of features demonstrating significant differences between classes guided the selection for training classifier models, encompassing various machine learning approaches such as Decision Trees, Support Vector Machine (SVM), and ensemble methods like Boosted, bagged RUS boosted tree, and the pre-trained GoogLeNet CNN. The resultant classification outcomes highlighted Ensemble Bagged Trees (EBT) as the standout performer, achieving an impressive 97% accuracy. EBT excelled in classifying both positive (gravel) and negative (non-gravel) classes, with 99% accuracy for gravel and 94% for non-gravel sounds. Despite a higher misclassification rate for non-gravel compared to gravel audio, the overall classification rate of EBT was exceptional. Among the SVM group, Quadratic SVM demonstrated the most favourable performance.

Convolution layers in Convolutional Neural Networks (CNNs) prove advantageous by eliminating the need for the feature extraction process in classical machine learning (ML) algorithms. Despite the demand for ample data to prevent overfitting in CNNs, the study addressed this limitation through transfer learning, utilising the pre-trained GoogLeNet architecture. Spectrograms of audio data for both gravel and non-gravel sound classes underwent data augmentation to enrich the dataset, resulting in an impressive 97.91% accuracy for the pre-trained CNN. Modifying the last fully connected layer facilitated binary classification, with stability achieved after 100 training epochs at a learning rate of 1×10^{-3}. This underscores that CNNs' ability to extract relevant features offers efficient and accurate classification compared to traditional supervised learning algorithms.

Regarding computational efficiency, classical methods underwent training on an Intel Core i5® CPU, with approximately 40 seconds needed to train and test each algorithm. A more powerful machine could notably reduce the feature extraction process, which takes 30 minutes. In contrast, GoogLeNet was trained on Google Colab using a GPU Nvidia K80/T4, expediting the process to around 30 minutes for training and validation results. Google Colab, a cloud-based machine learning and research service, facilitates access to GPUs/TPUs through cloud services, incorporating configurations for A.I. libraries and integration with Google Drive. While CNNs impose higher computational demands than methods like EBT, their capability to manage feature extraction and selection processes presents an advantage. The decision between EBT and CNN is contingent on the available computational resources. The research highlighted a comparative analysis between traditional machine learning methods and CNNs, emphasising the advantage of leveraging the knowledge from pre-trained networks for improved results in scenarios with limited data. The study provided a comprehensive approach to audio signal processing, feature extraction, and advanced machine learning techniques for the classification of gravel and non-gravel sounds based on recordings from gravel roads.

Paper II addresses the gap in road condition monitoring, focusing on gravel roads through acoustic data analysis. The research achieves effective gravel sound classification by leveraging machine learning, including supervised learning and Convolutional Neural Networks (CNN). Ensemble Bagged Tree classifiers excel in classical algorithms, reducing misclassifications, while CNN exhibits a high accuracy. Practical applications include visualising results on road maps for monitoring agencies and providing real-time data for trip planning. The study acknowledges limitations, such as data recorded from a specific vehicle, suggesting the need for diverse sources. Future research may explore integrating machine vision and fusing video and sound data for enhanced gravel road condition assessment in regions with similar terrain to Sweden's.

The data used in Paper III was collected following the same protocol as for Paper II. In the data labelling process, each recording is segmented into five-second parts, aligning with the time needed to capture the sound of gravel hitting the car's bottom. Manual classification assigns a class or grade from 1 to 4 to each segment based on loose gravel observed, as defined by Trafikverket. Two individuals classified video data using sample images and text descriptions, cross-checking their assignments for agreement. Irrelevant sections, such as talking or car manoeuvres and those with significant road disturbances, are omitted. The data labelling revealed an imbalanced distribution of classes, with Grades 2 and 3 being predominant. The dataset comprised audio segments amounting to a total of 869 audio segments labelled into four distinct classes, each denoting specific conditions.

Audio features are extracted from each segment using Short-Time-Fourier-Transform (STFT), which shifts sound signals from the time to the frequency domain. This process, conducted in Python using Librosa, prepares the audio signals for classification models. STFT divides the sound signal into windows and performs a Fourier Transform on each window, facilitating further analysis.

Various sound features, including zero crossing rate (ZCR), spectral centroid, spectral bandwidth, root mean square energy (RMS), Onset Strength, and Mel-Frequency-Cepstral-Coefficients (MFCCs), were extracted from audio files. Support vector machines, Random Forest, and multi-layer perceptron (MLP) machine learning algorithms were trained and tested using cross-validation. All the algorithms display good overall performance. Among all the algorithms, MLP demonstrated the highest accuracy, recall, precision, and an F1 score of 0.97. The f1-score is a more useful overall performance indicator as it is the harmonic mean of precision and recall. The results confirm that machine learning algorithms trained on audio data can help objectively classify gravel roads according to loose gravel conditions. MLP algorithm displayed superior performance across all evaluation metrics for grading the road segments according to the four-tier scale, although SVM exhibited comparable performance with MLP.

The multi-layer perceptron demonstrated superior performance among the tested models for the multiclass classification problem, indicating a cautiously positive outcome in replicating manual classification using machine learning models in the four-tier classification scale. Applications derived from such models have the potential to provide maintenance agency experts with an automated, efficient, and objective method for gravel road classification, aiding decision-makers in planning specific maintenance procedures based on loose gravel parameters. However, limitations and areas for future research are acknowledged. The manual classification process faced challenges, particularly in instances where loose gravel appeared visually, but no corresponding sound was detected, leading to potential discrepancies in classification. Addressing this, a subsequent review and discussion were conducted to mitigate labelling discrepancies. In the future, involving experts from road management agencies could enhance the manual assessment of road sections.

Additionally, exploring integrating image data alongside audio data for loose gravel classification may improve accuracy. Combining models trained on both types of data could lead to a more robust classifier. Further research into identifying sound features that best characterise road conditions could contribute to developing more effective classifiers in practical settings.

3.3. Image analysis for the loose gravel condition assessment

Previous research extensively employed machine vision to identify road defects on paved surfaces. Techniques like colour-based segmentation, support vector machines, and neural networks were utilised. Specialised methods were explored, such as using AdaBoost for defect identification on paved roads and analysing gravel road ruts using UAV drone imagery (Zhang & Elaksher, 2012). In (Albatayneh et al., 2019), a dust classification algorithm based on smartphone images was introduced, and in (Abu Daoud et al., 2021), a deep learning-based classifier evaluated the severity of corrugations on gravel roads. Research has predominantly addressed overall road roughness, neglecting distress types on gravel roads and the automation of loose aggregate measurement. The literature review underscores this gap, setting the context for paper IV. It utilises a diverse dataset from on-site cameras and Google Street View images of gravel roads in Sweden, emphasising the importance of image analysis in understanding road conditions.

Paper IV utilised gravel road images from two main sources. Firstly, videos recorded at 50 km/h on gravel roads in Dalarna County, Sweden, were extracted using two GoPro HERO7 cameras—one fixed inside the vehicle and the other on the bonnet. The second source included images of gravel roads from Google Street View, focusing on Sweden's gravel roads. Six hundred thirty-eight images were collected from both cameras and Google Street View, enhancing the dataset's size, diversity, and generalisation for model learning. The images were labelled into two classes: ratings 1 & 2 and ratings 3 & 4, representing road condition grades by Trafikverket. Due to the dataset's limited size, the images were merged into two distinct classes. Normalisation using ImageNet's mean and standard deviation was performed, and all images were resized to 224 × 224 pixels, with crucial pre-processing involving centre cropping for road visibility. Data augmentation techniques tripled the dataset size post-augmentation, including random horizontal flip, random rotation, and resize. The dataset was split into 60% training and 40% test sets.

The application of transfer learning is highlighted in this study, particularly in scenarios with limited datasets. Training an entire convolutional network from scratch with randomly initialised weights is impractical for small datasets. Transfer learning leverages pre-trained networks on large datasets like ImageNet, enhancing model performance on smaller datasets. This approach is advantageous, bypassing resource-intensive training from scratch, especially with GPUs.

Additionally, the study incorporates discriminative learning for fine-tuning pre-trained CNN model layers, adjusting learning rates based on the specific information each layer captures. The combination of Transfer and discriminative learning contributes significantly to the model's success, particularly with limited datasets. Three pre-trained networks—Residual Network (ResNet), Alexandria Network (AlexNet), and Densely Connected Convolutional Network (DenseNet)—were trained and tested, achieving successful gravel road classification despite challenges with a small and imbalanced dataset. VGG16, with transfer learning, outperformed other models, demonstrating the highest recall for positive class cases (class 1&2). The study also tested the system's adaptability by converting images to greyscale, yielding positive results. Employing pre-trained CNN models with transfer learning offers automation, eliminating traditional steps and ensuring reproducibility with high accuracy.

In conclusion, the study showcases the effectiveness of pre-trained CNN architectures, particularly VGG16, in accurately classifying loose gravel conditions. The dataset's augmentation, pre-processing, and the strategic use of Transfer and discriminative learning contribute to the model's success. The proposed system proves adaptable to variations in colour intensities, and its application can potentially replace traditional visual inspection methods in gravel road condition assessments. Future work involves experimenting with more training data, exploring interpretability using advanced algorithms, and engaging experts to validate and refine the proposed method.

3.4. Sensor fusion for improved assessment of loose gravel condition

Multimodal fusion techniques are methodologies employed to combine information or data from diverse sources or modalities, aiming to enhance the comprehension or performance of a specific system or application. These methods play a pivotal role in refining models utilised in affect recognition tasks, where data is collected from varied sources such as audio, visual cues, physiology, and more. Particularly beneficial in classification tasks, multimodal fusion provides substantial advantages. The amalgamation of information from multiple modalities enhances the feature space, amplifies the discriminative capabilities of models, and provides a more comprehensive understanding of intricate phenomena. The existing body of literature explores three primary joint fusion strategies: feature-level, decision-level (or score-level), and model-level fusion (Kanhere, 2011). In the concluding paper (Paper V) of the thesis, the combination of image and audio data is employed to detect loose gravel.

In studies conducted by In Paper II, III and IV the feasibility of objectively classifying loose gravel conditions using audio and images independently was explored, yielding promising results. Paper V introduces an innovative approach to detecting loose gravel on gravel roads by fusing spectrograms from audio recordings and images captured from the road surface. This multimodal fusion aims to significantly enhance the accuracy and reliability of loose gravel detection, aligning closely with the standards established by road traffic agencies globally.

The proposed methodology leverages the synergies between audio and image data, recognising that each modality provides unique insights into loose gravel detection. Audio spectrograms capture acoustic signatures such as gravel impacts, surface disturbances, and vehicle-induced vibrations, offering valuable acoustic signals indicative of loose gravel presence. Concurrently, images provide high-resolution visual information about the road surface, enabling the detection of loose gravel patches, displacement, and surface irregularities.

This paper explores two fusion methods specifically tailored for detecting loose gravel: feature-level fusion and decision-level fusion. Feature-level fusion involves combining features from two different sources—in this context, images and audio from gravel roads. On the other hand, decision-level fusion occurs at a later stage, combining decisions from models trained separately on images and audio. Fusion techniques provide a notable advantage in classification by enhancing accuracy and robustness. This advantage stems from their ability to effectively utilise complementary information from different sources or modalities, addressing the limitations of individual methods. Integrating data modalities or decision outputs improves accuracy, reliability, and adaptability, particularly in handling complex classification tasks (Alsaedi & Jaha, 2022; Das & Chandra, 2022).

The proposed framework for loose gravel detection aims to provide an objective method in line with the standards set by the Swedish Road Transportation Agency (Trafikverket). Automating loose gravel assessment would improve safety conditions on gravel roads and reduce maintenance response times.

The methodology employed in the study encompasses key aspects such as multimodal fusion, data collection, pre-processing, and the application of two distinct fusion techniques: feature-level early fusion and decision-level fusion.

Multimodal fusion techniques involve the integration of information from diverse sources or modalities to enhance system or application performance. In the context of affect recognition tasks, where data from audio, visual, and physiological sources contribute to the model analysis, three primary fusion strategies are explored. Feature-level fusion combines features from different modalities into a single vector, akin to human information processing. Decision-level fusion allows each modality to independently make predictions, with scores or results later combined. Model-level fusion combines the strengths of both feature-level and decision-level fusion, offering flexibility to adapt to specific task requirements.

The study utilised two GoPro HERO7 cameras for data collection, capturing audio and video data along gravel roads in Dalarna, Sweden. Recordings were conducted during the summer seasons of 2020 and 2022 at a consistent speed of 50 km/h under dry and sunny conditions. Excluded segments, such as activities unrelated to gravel road conditions, were omitted from the dataset.

For the pre-processing of image data, an annotation tool (Roboflow) was employed to highlight gravel roads and create bounding boxes, focusing specifically on crucial road aspects. This resulted in a dataset containing only road information, excluding irrelevant elements. For audio data, a conversion process using the Short-Time Fourier Transform (STFT) technique transformed audio signals into spectrograms, representing audio data visually for compatibility with image-based processing.

The dataset, comprising 15 videos, was categorised into two classes (1 & 2 and 3 & 4) based on Trafikverket's road condition classification. Both feature-level fusion and decision-level fusion were applied to both the image and audio datasets.

For feature-level fusion, Feature extraction utilised the VGG19 architecture, renowned for its effectiveness in image classification. Extracted features from road images and audio spectrograms were concatenated to create a unified representation. Principal Component Analysis (PCA) was applied for feature reduction, and machine learning algorithms, including Random Forest, Multilayer Perceptron (MLP), and XGBoost classifiers, were trained on the feature set.

Two variations of decision-level fusion were employed: OR Rule (Majority Voting) and AND Rule (Unanimous Voting). Majority Voting determined the final prediction based on the majority decision of individual models, while Unanimous Voting required all models to agree before making a positive prediction.

This study focused on the integration of audio and image data to enhance the accuracy of a classifier for detecting road conditions, particularly loose gravel scenarios. The dataset was categorised into road classes based on Trafikverket's standards, aimed to distinguish well-maintained roads (Classes 1 & 2) from those needing maintenance (Classes 3 & 4).

Early feature fusion involved extracting features from audio spectrograms and images using VGG19 and combining them into a unified feature space. The Random Forest classifier outperformed other models, achieving superior accuracy, precision, recall, and F1 score.

Late fusion, employing decision-level fusion with AND and OR gates, further improved results. Individual classifiers based on DenseNet121 for images and audio demonstrated high accuracies (0.95 and 0.92, respectively). Notably, the OR gate in decision-level fusion exhibited superior performance with an accuracy of 0.97, showcasing the method's adaptability to diverse input conditions.

Late fusion methods proved effective, compensating for noise or incompleteness in one modality with the strengths of the other. This approach enhances the overall system's robustness by leveraging complementary information from images and audio data. These methods excel in scenarios where one modality may be afflicted by noise or incompleteness, as the other modality can effectively compensate for these limitations. Each modality typically possesses unique strengths and weaknesses; for instance, images might excel in capturing visual details, while audio can contribute additional contextual information. Late fusion, as an approach, enables the fusion of data from both modalities (images and audio), thereby augmenting the overall system's robustness through the utilisation of complementary information from distinct sources.

4. Thesis from Microdata Analysis Perspective

MicroData Analysis (MDA) has proven to be applied to various research problems such as gravel road assessment, transport mode detection, building energy, measuring motor states in Parkinson's disease, and understanding market pricing trends. This section explains how this thesis contributes to each part of the Microdata Analysis chain, as shown in Figure 3. This thesis proposes using cost-effective mobile devices with artificial intelligence (A.I.) applications to collect audio and image data on gravel road conditions, improving data accuracy and efficiency through A.I. models. This real-time data facilitates trip planning on gravel roads and aids road maintenance agencies in identifying loose gravel areas, enhancing targeted maintenance efforts. The study aligns with microdata analysis principles by innovatively collecting data with sound analysis and image recognition. It employs machine learning to interpret and model the collected data effectively, enhancing data interpretation within the microdata analysis framework. Integrating sound and image data further advances the multidisciplinary nature of microdata analysis, offering valuable insights for informed road maintenance decisions and improved road conditions.

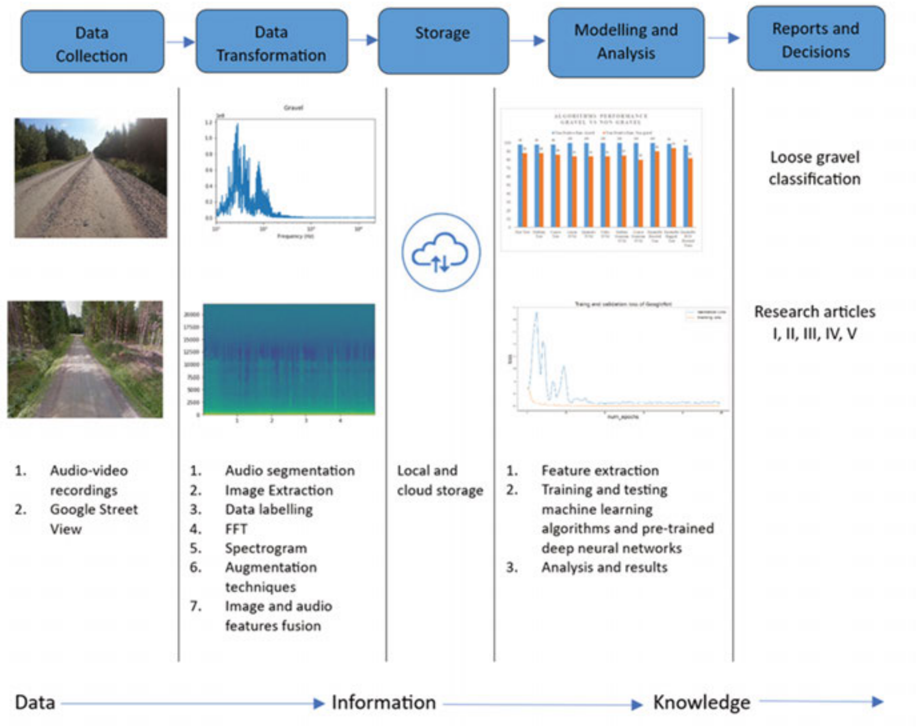

Figure 3 Thesis Workflow in the Context of Microdata Analysis

4.1. Data collection

The gravel road image and audio dataset specific to Sweden did not previously exist publicly. Recognising this gap, a unique dataset was collected using HERO7 GoPro cameras and driving through gravel roads during the summers of 2020 and 2022 in Dalarna, Sweden. Some data was also added from gravel roads of Sweden through Google steet view. The dataset, comprising 15 videos, strategically captures both audio and visual data along gravel roads. Irrelevant segments were excluded, and Roboflow's Annotation Tool ensured focused image data.

Each paper in this thesis features a dedicated phase for data collection. Paper, I perform a literature review to generate knowledge about the gravel roads, defects, and standards laid down and followed by road maintenance agencies around the globe. The paper also focuses on smart data-driven methods and research done on gravel road assessment. Papers II, III and IV collected audio-video data from the gravel roads in Dalarna County, Sweden. These Papers aimed to evaluate the potential usefulness of the sound of gravel hitting the bottom of the car for the classification of loose gravel conditions. Moreover, Paper III incorporated image data from Google Street View to increase the data set.

4.2. Data assessment and transformation

In paper II, audio data was extracted from the videos with 44,100 Hz sampling and 16 bits per sample. Audio segmentation was applied to the audio, and a data set of audio files of 5 seconds was created. The audio set was labelled gravel and non-gravel classes. Audio data were converted from the time domain to the frequency domain by applying the Fast Fourier Transform (FFT) to identify patterns in the data and extract features. Extracted features were used to train and test algorithms. The same audio set was converted to spectrograms in the same study to test traditional machine-learning algorithms against pre-trained convolutional neural networks. Audio files to spectrograms by applying Short Fourier Transform (STFT). This generated a two-dimensional series of matrices representing the frequency spectrum of audio files. These matrices are visualised as images, with the x-axis representing the time, the y-axis representing frequency, and the colour intensity representing the magnitudes of the power of each frequency component at a given point (Ren et al. 2018).

For paper III, images were extracted from videos and Google Street View, and augmentation techniques were applied to increase the data set. Images were labelled into classes according to Trafikverket standards, as seen in Figure 2, combining classes 1& 2 and 3 &4. Algorithms were trained and tested.

The dataset contains images sourced from various origins, resulting in pixel intensity and dimension variations among the images. To ensure stability during model training and reduce bias, we performed normalisation using the mean and standard deviation of ImageNet. This normalisation step is crucial for achieving stable and efficient model training. Additionally, all images were resized to a consistent size of 224 × 224 pixels. A fundamental pre-processing step involved center-cropping the images to ensure that the road is prominently featured in each Image. This step was essential to prevent some images from being resized in a way that makes it difficult to detect the road due to the presence of trees, grass, or the roadside.

Data augmentation techniques were applied to the initial dataset of 638 images. These techniques included random horizontal flipping, random rotation, and resizing. After each augmentation, a new set of images was created, effectively tripling the dataset's size and making it suitable for submission to a Convolutional Neural Network (CNN).

The dataset was then divided into two subsets: a training set comprising 60% of the data and a test set comprising 40%.

In paper IV, more data was added to the previous data set used in paper II. Videos were first labelled into four classes based on the loose gravel condition, and then the audio data was extracted. Algorithms were trained and tested on the features extracted from these audio files. In the data pre-processing phase, each audio recording is divided into five-second segments, chosen for their ability to capture the distinct sound of gravel hitting the car's underside. These segments are then manually assigned a class or grade ranging from 1 to 4 based on the level of loose gravel observed, per Trafikverket's guidelines. Since road conditions may change within each segment, the most prevalent class within that segment is assigned.

Two individuals manually classified the video data by examining sample images and text descriptions for four types of conditioned gravel roads (1-4) from the "Bedömning av grusväglag" document. Behavioral Observation Research Interactive Software (BORIS) was used for this classification (Friard & Gamba, 2016). Discrepancies between individual classifications were resolved through agreement on the final class.

Sections with irrelevant or noisy content (e.g., drivers talking or turning the car) and those with significant potholes, corrugation, or other unrelated sounds (e.g., passing vehicles) were removed.

During the labelling process, it became evident that the class distribution would be imbalanced, with classes 2 and 3 being the most common.

Feature extraction involved extracting audio features from each audio segment using Python's Librosa package. The Short-Time-Fourier-Transform (STFT) was applied to the sound signals, shifting them from the time domain to the frequency domain. STFT divides the sound signal into smaller windows and performs a Fourier Transform on each window, allowing analysis of frequency components in individual segments. The Hann window function was applied to the STFT, serving as a weighting mechanism that focuses on the frequency components within the window while reducing discontinuities in the signal. Overlapping windows, with a 50% overlap in this study, were used to mitigate information loss due to the weighted window.

In Paper V, A comprehensive data pre-processing and transformation procedure was employed on recorded videos to extract and isolate gravel road information from both audio and image datasets. The pivotal tool in this process was the Roboflow Annotation Tool, a specialised platform for visual data tasks. This tool created bounding boxes to highlight and segment the gravel road sections within the images while filtering out unrelated elements, such as the sky and vegetation. This meticulous approach generated a new dataset exclusively containing pertinent road-related data.

Furthermore, the audio data underwent a significant transformation, as audio files were converted into spectrograms using the Short-Time Fourier Transform (STFT) technique. With their time and frequency axes, these spectrogram images effectively converted audio data into a visual format seamlessly integrated with the existing image-based processing pipeline. This extensive data preparation process was undertaken to ensure that subsequent classification algorithms could concentrate on learning distinctive features associated with road conditions.

4.3. Data storage

The audio-video data used in this thesis was collected during summer when gravel roads were free of snow. The data collection period spanned from 2020-2022. The data was stored in external devices for easy access as well as in cloud storage for backup, facilitating sharing between researchers.

4.4. Reports, analysis, and decisions

All the papers in this thesis included the process of analysis and report generation in the form of publication discussing the analysis results. These results can further facilitate decision-making. Paper I explores traditional and smart methods in gravel road defect detection along with their advantages and limitations, and research gaps are identified. Based on this paper, it was found that specific defect detection of gravel road defects is important to enable road maintenance agencies to plan targeted maintenance interventions. As a result of the first study, it was decided to explore further the potential of utilising machine learning algorithms to detect loose gravel by analysing audio and video recordings.

Paper II focused on training and testing machine learning algorithms, including decision trees, support vector machines (Noble 2006), ensemble methods (Dietterich 2000) and the pre-trained network GoogLeNet (Szegedy et al. 2015) to distinguish between gravel and non-gravel sounds. Using pre-trained networks eliminates the need for extensive training data and reduces the time required for training. GoogLeNet performed the best in classifying both classes. This study aimed to determine if algorithms can differentiate the sound of gravel hitting the bottom of a car and explore the potential of mapping the audio to road types based on their condition, according to Trafikverket.

The study of paper III aims to evaluate the performance of deep learning-based pre-trained networks in rating gravel road images according to classical methods by human experts. The dataset consists of images of gravel roads extracted from self-recorded videos and images extracted from Google Street View. The images were labelled manually, referring to the standard images as ground truth defined by the Road Maintenance Agency in Sweden (Trafikverket). Various pre-trained models for computer vision tasks, namely Resnet18, Resnet50, Alexnet, DenseNet121, DenseNet201, and VGG-16 (Dhananjay Theckedath 2020; Li and Lima 2021; Long, Shelhamer, and Darrell 2015), were used in this study. In practice, training an entire convolutional network from scratch with random weight initialisation is uncommon, especially when working with small datasets. Transfer learning, a technique explored in various studies (such as Weiss et al., 2016 and Transfer learning for computer vision tutorial), plays a crucial role in addressing this issue. Transfer learning involves leveraging knowledge gained from pre-trained networks trained on large datasets and applying it to a new, smaller dataset. Pre-trained models serve as a valuable resource for improving the initialisation and convergence of models when dealing with limited data.

These pre-trained models are typically trained on vast datasets like ImageNet, which contains 1.2 million images across 1000 categories. The knowledge acquired by these networks can be effectively transferred to other image classification tasks with smaller datasets. For instance, if a network has learned to recognise cars, it can be repurposed to detect trucks in a different problem, as demonstrated in studies (George Karimpanal, Thommen; Bouffanais, 2019). By utilising transfer learning, the need to train a network entirely from scratch is mitigated, saving considerable computational resources and time, as the training process can be demanding and may require days or even weeks on Graphics Processing Units (GPUs).

The earlier layers of pre-trained networks have learned general features like shapes, edges, and colours, which can be valuable for addressing a variety of computer vision problems. This approach enables the reutilization of these generalised features for different tasks. All the models performed well, with an accuracy of over 92%. The results reveal that the pre-trained VGG-16 with transfer learning exhibited the best performance in terms of accuracy and F1-score compared to other proposed models. This research's primary objective was to evaluate machine vision's capability in identifying loose gravel images. The ultimate aim is to combine sound and image data in future studies, enabling a more precise depiction of gravel roads.

Paper IV used the Swedish Transport Administration's four-tier grading scale to assess gravel road conditions, implying that the classification problem consisted of four distinct classes. The authors manually classified the video recordings into four classes to create a labelled dataset. We created a dataset by extracting audio recordings from these labelled videos. The researchers used the extracted audio features for classification to train and test three machine learning algorithms: Multi-layer Perceptron, Random Forest, and Support Vector Machine (Kolagati, Priyadharshini, and Mary Anita Rajam 2022; Schonlau and Zou 2020). This study shows that the sound from gravel roads varies from road type 1 to road type 4 and thus is helpful to provide information on detecting loose gravel and detecting road type. Among these techniques, the Multi-layer Perceptron (MLP) exhibited the highest classification accuracy at 0.96 and f1 scores, recall, and precision of 0.97. These findings indicate that audio data can be effectively employed to classify loose gravel conditions.

Paper V focuses on developing an innovative approach for identifying loose gravel conditions on gravel roads. This method involves combining audio and image data to enhance the detection process. The research categorises road conditions into predefined classes to align with the standards set by Trafikverket, ranging from Class 1 (good condition) to Class 4 (worst condition). Due to limitations in data volume, Classes 1 and 2 were combined into one category, and Classes 3 and 4 into another.

The methodology involves extracting relevant audio and image data from video recordings to train classifiers. Two fusion techniques were explored: feature-level early fusion and decision-level late fusion. In the late fusion method, both OR and AND gates were tested, with the OR gate demonstrating superior accuracy in the classification process. The initial feature fusion approach, which incorporated a pre-trained VGG19 model and PCA, resulted in improved accuracy for the Random Forest classifier, surpassing other models in terms of accuracy 0.9018, precision 0.9011, recall 0.9018, and F1-score 0.9014. The late fusion technique was used, which involved decision-level processing using logical disjunction and conjunction gates (AND and OR) to improve the results further. When combined with individual classifiers for images and audio based on DenseNet121, the late fusion method exhibited significant performance, particularly when employing the OR gate, achieving an accuracy of 97%. This late fusion approach enhances adaptability by compensating for limitations in one modality with information from the other, making it a robust and versatile method.

In an additional study not included in this thesis, we found that the consistency of humans rating gravel roads is subjective and varies from labeller to labeller. The study also recommended using machine learning algorithms to achieve an objective, efficient and cost-effective method for gravel road assessment (Kyros and Elias Myrén 2022). Figure 4 shows the relation among the papers included in the study.

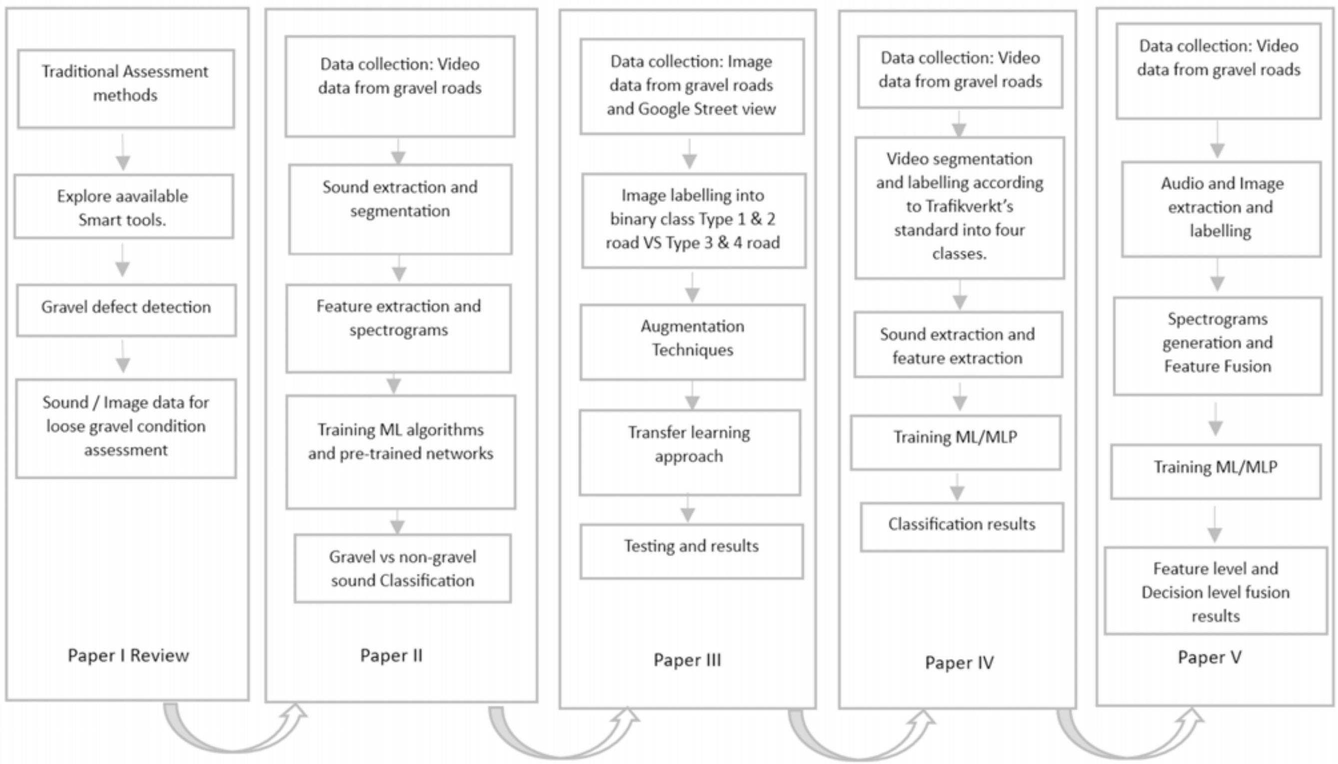

Figure 4 The relation among included studies

5. Concluding discussion

Through the literature review (Paper I), it was evident that there is a need for cost-effective, efficient, and objective methods for gravel road assessment. The sound and image data from gravel roads enabled machine learning algorithms to learn and detect loose gravel conditions according to the Trafikverkets standards. Hence, deploying models discussed in this thesis can enable real-time reporting of loose gravel conditions, effectively easing the manual classification workload on Trafikverket. The proposed method can be used in Sweden and countries with similar terrain and climate. A system based on our proposed methods should be capable of providing precise information regarding the type of defect that exists on the gravel road. This would optimise the utilisation of resources and for the specific maintenance drill required for the identified defect. Additionally, these methods bundled in a suitable application could map the road condition information to real-time maps and make it available for the citizens to schedule their trips and obtain relevant information about the road conditions.

The advantages of using pre-trained CNN models such as Vgg16 and GoogLeNet with Transfer learning for classification include the fully automated classification mechanism, removal of conventional segmentation and feature extraction steps, and the reproducibility of predictions made by the models with high accuracy. The transfer learning approach achieves an overall satisfactory performance regarding accuracy, f1 score, precision, and recall. Moreover, this approach can handle imbalanced or large-scale datasets, providing the necessary flexibility to make informed decisions. The remainder of this section addresses the research questions and provides concluding remarks. It also discusses potential directions for future research.

Paper I contributes to the knowledge around gravel roads and defects that are evaluated for gravel road condition assessment and when and how gravel road assessment is done in Sweden and other parts of the world. Visual assessment is one of the most common methods used for gravel road evaluation worldwide. Road maintenance agency Trafikverket in Sweden follows the method load in the report of *Bedömning av grusväglag* (Trafikverket 2014). This method relies on field experts who visually inspect that road for various parameters such as roughness as one entity comprising potholes, cross fall, and corrugation, whereas rutting is measured separately. Loose gravel and dust are measured separately but are considered together as binding ability.

Experts grade gravel roads into four categories, grade 1 being the best condition road and grade 4 as the worst condition road. To grade gravel roads, the expert requires both images from the road under observation and supplementary text describing the road. However, this method is highly subjective and vulnerable to the error of judgment. However, road profiling trucks can give advanced information on defects present but are not used for gravel road assessment to ensure cost-effectiveness. Paper I also identifies the need for having a cost-effective and objective solution using data-driven methods to determine the presence of specific defects that could help maintenance agencies formulate a maintenance scheme targeting the defect. Paper I proved to be motivation for further studies.

Paper II highlighted the importance of sound data in various research areas and its use to detect events such as loose gravel conditions. The study also tested several machine learning algorithms to see the potential of machine learning algorithms to classify binary class problems of gravel vs non-gravel audio. The study also utilised converting sound data to spectrograms to leverage the capability of Transfer learning through pre-trained convolutional neural networks. Although with the limited dataset, GoogLeNet outperformed other algorithms in correctly classifying the two classes. This study demonstrated the potential of using sound data for identifying loose gravel conditions according to road maintenance agency standards.

Research question 2 encompasses the evaluation of machine learning algorithms' capacity to classify gravel road conditions and seeks to understand the extent to which they can effectively do so. This inquiry is addressed in depth from the second to the fifth paper, focusing on assessing the capabilities of machine learning algorithms in classifying gravel road conditions.

Paper III answers question 3—the feasibility of classifying loose gravel conditions. The idea is to add image data to sound data for a more accurate picture of the road condition. This study uses a new method of extracting image data from Google Street View. The algorithm was developed to always centre crop the images to focus on the road to ensure that the algorithm's learned features are from the road itself and not from the surrounding environment of the gravel road. The results showed that the pre-trained VGG-16 with transfer learning exhibited the best performance in terms of accuracy and F1-score compared to other proposed models. An algorithm was created to assess the proposed system's performance in classifying images with varying colour intensities. VGG16 exhibited consistent and impressive results in this experiment, demonstrating its ability to handle different colour intensities effectively.

Paper IV used audio data to answer the fourth question of the thesis. The paper evaluated the performance of loose gravel conditions using audio data. The results showed that all algorithms had good overall performance, with MLP performing the best among the tested algorithms in accuracy, recall, precision, and F1 score.

Finally, paper V answers research question 5 of the thesis. This study demonstrates significant promise in identifying loose gravel conditions for gravel road maintenance. By leveraging audio and image data fusion, it provides a holistic approach to identifying loose gravel conditions on gravel roads.

Furthermore, the study's methodology is not limited to the detection of loose gravel conditions alone; it is adaptable to other road defects, such as potholes, dust, and corrugations. This adaptability makes it a versatile tool for assessing various aspects of road infrastructure, contributing to a more comprehensive system for monitoring and maintaining roads. Road maintenance agencies benefit significantly from this data-driven approach, enabling them to prioritise resources more effectively and allocate maintenance efforts based on identified road conditions. Efficient resource allocation streamlining brings cost savings and lessens the environmental impact of extensive road repair activities. Moreover, it plays a role in reducing accidents and mitigating the wear and tear on tires.

The fusion techniques employed in this study, utilising multiple modalities such as sound and image data, offer a key advantage by enabling predictions to be generated even in scenarios where data from one modality is unavailable. Incorporating both Image and sound is important in the comprehensive and complementary information they provide. Sound data may capture nuances and details not visually evident in images and vice versa. This synergistic approach enhances the robustness and reliability of predictions, particularly in diverse and complex environments. In addition to the comprehensive insights provided by incorporating both image and sound data, another notable advantage is the ability of these fusion techniques to compensate for the absence of information from one sensor modality. In situations where data from one sensor is missing or incomplete, the complementary nature of the modalities allows for a compensatory effect. For example, if image data are deficient, the information from sound sensors can help fill the gaps, and vice versa. This inherent redundancy enhances the overall robustness of the system, ensuring a more reliable and adaptable predictive model, even when faced with incomplete or partial data from one sensor source.

While using only one modality, such as image data alone, might still provide valuable insights, it comes with limitations. For instance, relying solely on images might miss important auditory cues that could indicate certain conditions. Likewise, using only sound data might lack the visual context necessary for a more complete understanding. The combination of both modalities not only improves the accuracy of predictions but ensures a more holistic and thorough analysis of the given scenario. Therefore, incorporating both image and sound data is essential for maximising the effectiveness and reliability of the predictive models in this study. The findings presented in the Papers are integral to the attainment of the central objective of the thesis, which is to evaluate the feasibility of machine learning-based automated road condition detection systems that closely emulate Trafikverket's established procedures. Drawing from the insights derived within this thesis, it becomes evident that adopting applications on widely available devices, such as smartphones, for collecting audio-visual data and subsequent result mapping can offer substantial advantages to road maintenance agencies. This approach holds the promise of numerous societal benefits and the delivery of precise measurements and maintains cost-effectiveness. Notably, the study's strength lies in its reliance on easily accessible devices, like smartphones, for data collection. This strategy not only opens avenues for community engagement in data collection efforts but also empowers local residents, fostering a collaborative approach to infrastructure assessment and maintenance, thereby enhancing the active involvement of the community in these endeavours.

6. Limitation and scope for future studies

This thesis presents a machine learning model designed for loose gravel detection, primarily aimed at assisting road maintenance agencies. Nonetheless, it is essential to acknowledge certain limitations in the study. One significant limitation is the exclusive use of recordings solely from one vehicle, namely the Volkswagen Passat GTE. Additionally, data collection was restricted to Swedish gravel roads during the summer season. Recognising the potential variability in engine sound levels among different vehicles, including recordings from a diverse range of vehicles, is imperative. This inclusivity would likely provide a more comprehensive understanding of the system's efficacy in detecting gravel sounds across various car models.

As a recommendation for future work, exploring collaboration with experts from road management agencies for data labelling is proposed. This collaborative effort holds the potential to significantly augment the validation process by incorporating the specialised knowledge of these experts. Their insights can contribute valuable context, aligning the trained models more closely with the practical challenges encountered in road maintenance.

In the specific context of collecting sound and image data from gravel roads, collecting data during dry, sunny days in the summer is advised, ensuring the absence of snow on the roads. Two distinct data collection approaches warrant consideration. One approach involves crowdsourcing, where individuals with smartphones can enthusiastically contribute data via a user-friendly mobile application, concurrently gaining access to real-time gravel road conditions on maps for their journeys. This approach boasts the advantage of yielding extensive geographic coverage and diversifying data perspectives. Alternatively, data collection could be managed exclusively by the road maintenance agency, involving the deployment of specialized sensors during specific weather conditions to guarantee a more dependable and uniform data-gathering process. For the efficient management of data storage and organization, the exploration of cloud-based storage solutions is highly recommended, offering secure storage options fortified with robust access controls and encryption measures. Data sharing can be seamlessly facilitated through cloud-based collaboration platforms, while data analysis can be streamlined by utilising cloud-based machine learning and data analytics tools.

Additionally, future research could be conducted for quantitative profiling aspects to enrich the analysis further. This entails incorporating metrics such as the average amount of loose gravel per unit area and the depth of potholes. Integrating these quantitative measures would enhance the dataset, paving the way for the development of machine-learning models that provide more nuanced and quantitative insights into the overall condition of roads.

In conclusion, as we look ahead in this domain, it becomes evident that future efforts should adopt a dual focus. Firstly, collaboration with road management experts for qualitative data labelling is crucial to effectively refine and validate machine learning methods. Their specialised knowledge adds context and aligns models with the practical challenges encountered in road maintenance. Secondly, there is a need to explore quantitative profiling, including various aspects such as the detection of potholes and the overall assessment of road quality. Incorporating both qualitative and quantitative dimensions can significantly enhance machine learning methods' overall effectiveness and precision in evaluating and addressing road maintenance needs.

Practically implementing this comprehensive approach could involve the development of an innovative solution, perhaps in the form of a user-friendly app similar to "RoadRoid." This app could leverage the widespread use of mobile devices equipped with A.I. applications. Users, including road maintenance agencies and individual drivers, could actively contribute by using their smartphones to record audio and capture images of road conditions during their travels. The app would employ machine learning models capable of analysing diverse data, detecting loose gravel, identifying potholes, and evaluating the overall road quality.

Collaboration with road management experts would remain integral for accurate data labelling, ensuring that the models are trained to recognise and interpret various road conditions accurately. By encouraging user participation and incorporating quantitative profiling metrics, such as the depth of potholes, the app could provide a real-time, detailed assessment of road conditions. This proactive and collaborative approach has the potential to revolutionise road maintenance practices, enabling more efficient and targeted interventions.

In summary, future endeavours in this field should embrace a comprehensive strategy that combines qualitative collaboration, quantitative profiling, and practical implementation through user-friendly applications. This holistic approach promises to amplify the effectiveness and precision of machine learning methods, ultimately contributing to improved evaluations and more informed decisions in road maintenance.

7. References

Abu Daoud, O., Albatayneh, O., Forslof, L., & Ksaibati, K. (2021). Validating the practicality of utilising an image classifier developed using TensorFlow framework in collecting corrugation data from gravel roads. *International Journal of Pavement Engineering*, *May*. https://doi.org/10.1080/10298436.2021.1921773

Albatayneh, O., Forslöf, L., & Ksaibati, K. (2019). Developing and validating an image processing algorithm for evaluating gravel road dust. *International Journal of Pavement Research and Technology*, *12*(3), 288–296. https://doi.org/10.1007/s42947-019-0035-y

Alsaedi, N. H., & Jaha, E. S. (2022). Dynamic audio-visual biometric fusion for person recognition. *Computers, Materials and Continua*, *71*(1), 1283–1311. https://doi.org/10.32604/cmc.2022.021608

Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

Bonaccorso, G. (2018). *Machine Learning Algorithms: Popular algorithms for data science and machine learning*. Packt Publishing Ltd.

Cabral, F. S. (2019). An Automatic Survey System for Paved and Unpaved Road Classification and Road Anomaly Detection using Smartphone Sensor. *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*.

Cord, A., & Chambon, S. (2012). Automatic Road Defect Detection by Textural Pattern Recognition Based on AdaBoost. *Computer-Aided Civil and Infrastructure Engineering*, *27*(4), 244–259. https://doi.org/10.1111/j.1467-8667.2011.00736.x

Das, K. P., & Chandra, J. (2022). Multimodal classification on PET/CT image fusion for lung cancer: a comprehensive survey. *ECS Transactions*, *107*(1), 3649.

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11). https://doi.org/10.1111/2041-210X.12584

George Karimpanal, Thommen; Bouffanais, R. (2019). Self-organizing maps for storage and transfer of knowledge in reinforcement learning. *Adaptive Behavior*, *27*(2), 111–126.

Gregory D. Cline, J. A. B. (2003). *Automated data collection for pavement condition Index Survey*.

Gumaei, A., Ismail, W. N., Rafiul Hassan, M., Hassan, M. M., Mohamed, E., Alelaiwi, A., & Fortino, G. (2022). A Decision-Level Fusion Method for COVID-19 Patient Health Prediction. *Big Data Research*, *27*. https://doi.org/10.1016/j.bdr.2021.100287

Hameed, Z., Zahia, S., Garcia-Zapirain, B., Aguirre, J. J., & Vanegas, A. M. (2020). Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors (Switzerland)*, *20*(16), 1–17. https://doi.org/10.3390/s20164373

Huntington, G., & Ksaibati, K. (2010). *Gravel road management:Programming Guide* (Vol. 3, Issue October).

Kanhere, S. S. (2011). Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. *IEEE 12th International Conference on Mobile Data Management, 06-09 June, Lulea, Sweden*, 3–6. https://doi.org/10.1109/MDM.2011.16

Kans, M., Campos, J., & Håkansson, L. (2020). Condition Monitoring of Gravel Roads-Current Methods and Future Directions. *Advances in Asset Management and Condition Monitoring*, 451–461.

Kyros, G., & Elias Myrén. (2022). *Automatic loose gravel condition detection using acoustic observations, Master Thesis in Microdata Analysis* [Dalarna University]. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1682300&dswid=9166

Lin, Q., Ye, G., Wang, J., & Liu, H. (2022). RoboFlow: a Data-centric Workflow Management System for Developing AI-enhanced Robots. In A. Faust, D. Hsu, & G. Neumann (Eds.), *Proceedings of the 5th Conference on Robot Learning* (Vol. 164, pp. 1789–1794). PMLR. https://proceedings.mlr.press/v164/lin22c.html

Mbiyana, K., Kans, M., Campos, J., & Håkansson, L. (2022). Literature Review on Gravel Road Maintenance: Current State and Directions for Future Research. *Transportation Research Record*, 03611981221133102. https://doi.org/10.1177/03611981221133102

Oppong Adu, W., Dumedah, G., & Adams, A. C. (2023). Surface Condition Assessment of Unpaved Roads Through the Use of Unmanned Aerial Vehicle. *International Journal of Pavement Research and Technology*. https://doi.org/10.1007/s42947-023-00374-z

Robert, E., & Ronald, B. (1992). *Unsurfaced Road Maintenance Management* (Issue December).

Saeed, N., Nyberg, R. G., & Alam, M. (2022). Gravel road classification based on loose gravel using transfer learning. *International Journal of Pavement Engineering*, 1–8. https://doi.org/10.1080/10298436.2022.2138879

Saeed, N., Nyberg, R. G., Alam, M., Dougherty, M., Jooma, D., & Rebreyend, P. (2021). Classification of the acoustics of loose gravel. *Sensors*, *21*(14). https://doi.org/10.3390/s21144944

Skorseth, K. (2000). *Gravel roads: maintenance and design manual*. US Department of Transportation, Federal Highway Administration.

Souza, V. M. A., Giusti, R., & Batista, A. J. L. (2018). Asfault: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning. *Pervasive and Mobile Computing*, *51*, 121–137. https://doi.org/10.1016/j.pmcj.2018.10.008

Trafikverket. (2014). *Bedömning av grusväglag (Assesment of gravel roads),TDOK 2014:0135 Version 1.0 ,Trafikverket*. https://trafikverket.ineko.se/Files/sv-SE/10845/RelatedFiles/2005_060_bedomning_av_grusvaglag.pdf

Walker, D., Entine, L., & Kummer, S. (1987). *Pavement Surface Evaluation and Rating PASER Manual*. https://trid.trb.org/view/260614

Wang, H. W., Chen, C. H., Cheng, D. Y., Lin, C. H., & Lo, C. C. (2015). A real-time pothole detection approach for intelligent transportation system. *Mathematical Problems in Engineering*.

Yang, Q., Zhang, Y., Dai, W., & Pan, S. J. (2020). *Transfer learning*. Cambridge University Press.

Zhang, C., & Elaksher, A. (2012). An unmanned aerial vehicle-based imaging system for 3D measurement of unpaved road surface distresses. *Computer-Aided Civil and Infrastructure Engineering*, *27*(2), 118–129. https://doi.org/10.1111/j.1467-8667.2011.00727.x