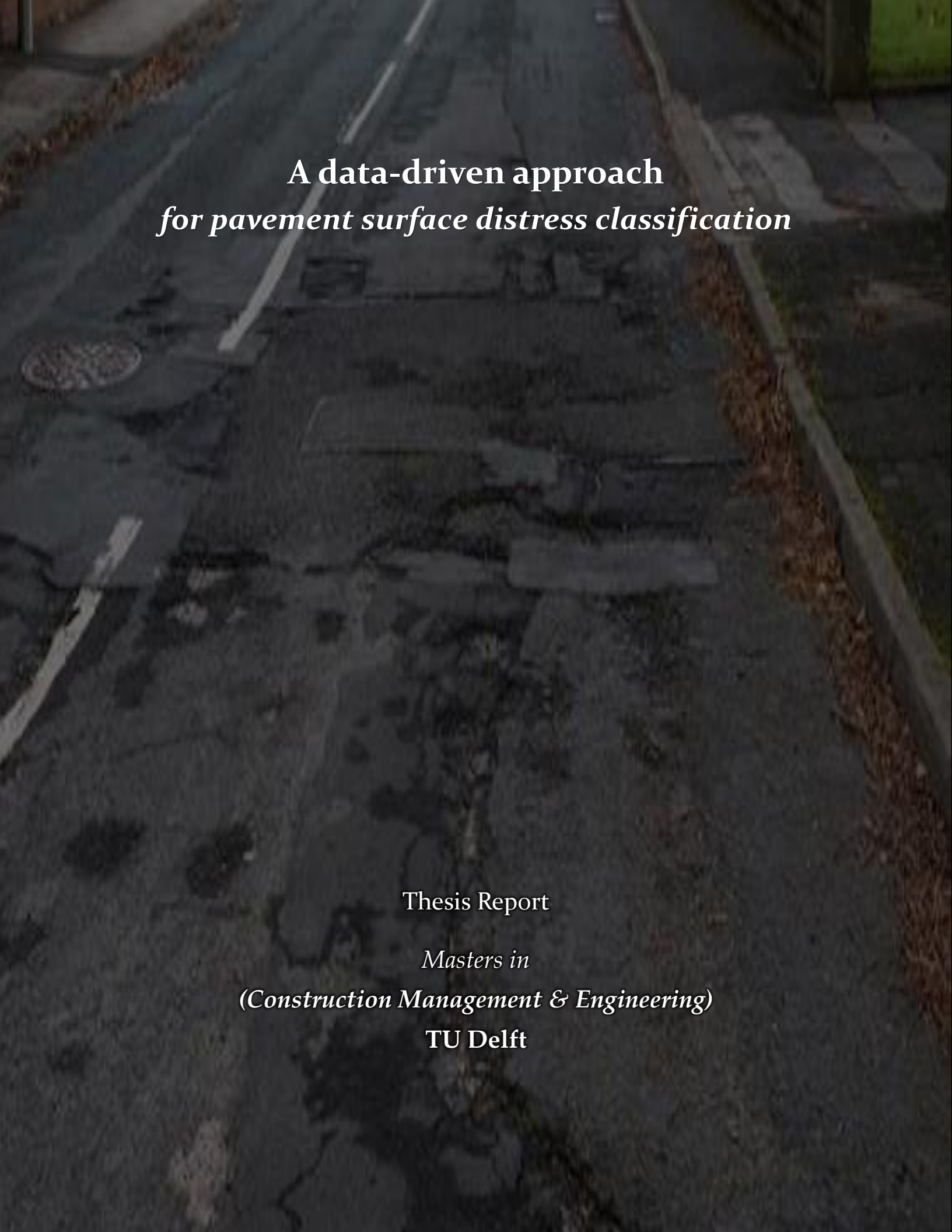


A data-driven approach
for pavement surface distress classification

Thesis Report

Masters in

(Construction Management & Engineering)

TU Delft

[*This-page-is-left-blank*]

Colophon

Student:

Author	: Saurabh Sharma
Student Number	: 4609247
Date	: November 2019
Email address	: s.ssau91@gamil.com
Date of final presentation	: February 4th 2020

University:

Name	: Delft University of Technology
Faculty	: Faculty of Civil Engineering and Geoscience
Master track	: Construction Management & Engineering (CME)

In collaboration with

Company	: RoadEO

Graduation Committee:

Daily Supervisor	: Dr. K. Anupam (Faculty of CiTG)
Chairman	: Prof. Sandra Erkens (Faculty of CiTG)
External Supervisor	: Dr. Mottaqiallah Taouil (Faculty of EEMCS)
Company Advisor	: Mr. Alexander Gunkel

Acknowledgment

An African proverb goes like:

If you want to go fast,
Go Alone.
If you want to go far,
Go Together.

The pursuit of this long (*far)*and arduous journey couldn't have been possible if I haven't had the support and blessings of the countless people who surround me. This mammoth task couldn't have been done if I were to do it alone. So, I am really blessed to meet such great people around this thing called life.

At first, this journey is an academic pursuit so my first and the warmest gratitude goes towards my thesis committee. It was my utmost pleasure to work with Dr. K. Anupam who served as the first supervisor for this research. He not only provided the right balance of guidance and direction for making this project successful but also provided a very healthy and passionate atmosphere every time. I also truly appreciate Prof. Sandra Erkens and Dr. Mottaqiallah Taouil for their time and support. The research wouldn't have shined if your viewpoints and understanding wasn't available. I wish you all a healthy and joyful life.

F.R.I.E.N.D.S. The people who make the most subtle and the biggest part of your life. From emotional support to the economical, from good food till the fun leisure time, from solving the smallest issues till the academical, I am so so grateful to have you all. Mentioning a few names wouldn't be fair to others, so I am gonna thank you all for everything you have done for me.

If this wasn't an academic acknowledgment, I would have credited everything to my family at first. Knowingly or unknowingly they have always been around me. It always tears me up when I think about their sacrifice and unconditional love that they have provided me. Thus I want to dedicate all of my achievement to my family and loved ones. Thank you all for being there because this was a collective accomplishment.

Executive Summary

Pavement undergoes a fast deterioration process either due to the damages induced by weather conditions, an increase in traffic flow and load, or passive factors like aging of infrastructure. Causing in reduced performance, safety, and comfort for the riders. This requires periodic rehabilitation measures to maintain the optimum condition of the underlying asset.

The Netherlands has one of the world's densest road networks. Of about 130,000km which is hard-surfaced asphalt pavement, 5,000km contributes to national trunk roads. Since damages on such a dense asphalt road network, imposes economic setbacks and concern for the users, governmental authorities are looking for a proactive approach to detect and classify distresses in their "early" stages. As a reason, Province of Zuid-Holland (PZH) conducts inspection procedures to assess the condition of the entire road network every year once, which has proved optimal until now, but as the traffic flows are increasing and weather conditions are worsening, a new approach is required to mitigate the need for a frequent, cost-effective and reliable inspection method.

Modern data sources such as smartphones are the biggest crowdsourced data generators. Having an intriguing number of sensors and in-built features, governmental and private authorities are just starting to acknowledge the potential of such crowdsourced data generators. Smartphones can serve as an auxiliary companion, by providing a cost-effective and a frequent substitute, to pre-existing inspection devices and systems. The research objective is to analyze if such modern data sources can be utilized to detect, classify and segment road surface distresses.

Based on the research problem, scope and objective a research question is formulated which is:

Can we utilize modern data-sources such as smartphones for detecting, classifying and segmenting pavement surface-distresses?

To answer the main research question, three data-driven models were built to assess the applicability and reliability of the proposed method by utilizing modern data sources such as smartphones. At first, vibration data collected via an in-built accelerometer sensor was put to use for detecting anomalies and predicting International Roughness Index (IRI) values, as reliable as Automatic Road Analyzer (ARAN) measurements. Secondly, a data-driven image classification model was built to automate the mundane task of manually classifying pavement

surface distresses. Lastly, a semantic segmentation model is built to objectively estimate the severity of the given surface distress.

In this research data-types like vibration and imagery were gathered and synthesized to assess the efficiency and accuracy of the 3 machine learning models. The research followed a 7-step methodology to build the three data-driven models in an unsupervised, semi-supervised and fully supervised approach. A Random Forest decision tree is trained with an input of accelerometer and gyroscope data to predict the roughness index of the road, in a completely unsupervised manner. The model performance graphs (Fig. 4.1) and the predicted IRI table (Table 4.2) suggest that the model could estimate a good range of IRI values (error rate +-0.5), except the upper (above 4.8) and lower extremes (below 0.8).

The use-case for the classification model was to build a Deep Learning model that can automate the mundane task of manual classification of images for pavement surface distresses. A total of 15,750 images were labelled in a semi-supervised manner. Methods of Transfer Learning were utilized to minimize the computational power required to train and build the Convolutional Neural Network (CNN) model from ground-up. A custom sequential model is utilized to replace the last layers of the VGG16 model for classifying 6 object classes. The multi-classification model is trained on NVIDIA GPU GeForce 940M for 25 iterations, the best validation accuracy achieved was ~86% with the corresponding validation loss of ~0.43.

If road damages are detected and classified, the PZH further needs to diagnose the severity of the classified distresses as objectively as possible. As cracks are the most commonly occurring distress among the primary and secondary road networks of the Netherlands, a pixel-level semantic segmentation model was constructed to solve the issue at hand. A total of 6 EnDec deep neural network architecture were built to classify and segment the given surface distress. A total of 22,996 images were annotated resulting in more than (~7 billion) pixels and (13 to 54 million) network parameters. By utilizing multi-threading opportunities provided by the Dutch supercomputer called *Cartesius*, the six models were trained over 6 to 16 times for about 4hrs to 26hrs of time. Lightweight models like U-net-VGG16 could provide timely performance but with lesser accuracy. Heavyweight models on the other side had higher precision and recall accuracy (F1) and also performed quite satisfactorily on the Intersection over Union (IoU) scores.

[*This-page-is-left-blank*]

Table of Contents

List of Figures

List of Tables

List of Abbreviations

PMS	: Pavement Management System
PZH	: Province of Zuid-Holland
ARAN	: Automatic Road Analyzer
LRS	: Linear Referencing System
GPS	: Global Positioning System
IRI	: International Roughness Index
GPR	: Ground Penetrating Radar
XVP	: Laser Transverse Profile
POS LV	: Position and Orientation System for Land Vehicles
DMI	: Distance Measuring Instrument
CSIRO	: Commonwealth Scientific and Industrial Research Organization
ML	: Machine Learning
SVM	: Support Vector Machine
FWD	: Falling Weight Deflectometer
FT	: Fourier Transforms
WT	: Wavelet Transforms
SWT	: Stationary Wavelet Transform
DWT	: Discrete Wavelet Transform
SSL	: Semi Supervised Learning
DL	: Deep Learning
CV	: Computer Vision
CNN	: Convolutional Neural Network

TL : Transfer Learning

VGG : Visual Geometry Group

ILSVRC : Imagnet Large Scale Visual Recognition Challenge

DNN : Deep Neural Network

CBIR : Content-Based Image Retrieval

FCN : Fully Convolutional Networks

EnDec : Encoder-Decoder

RNN : Recurrent Neural Network

LW : Light Weight

MW : Medium Weight

HW : Heavy Weight

ISBI : International Symposium on Biomedical Imaging

MSE : Mean Squared Error

LR : Learning Rate

SGD : Stochastic Gradient Descent

GPU : Graphics Processing Unit

CPU : Central Processing Unit

SBU : Strategical Business Unit

HPC : High-Performance Computing

RAM : Random Access Memory

IoU : Intersection over Union

F1 : Precision recall metrics

[*This-page-is-left-blank*]

Chapter 1: Research

1.1 CONTENT

This chapter presents the research motivation in Section 1.2 of the research introduction, research problems that are prohibiting further improvements in the industry are discussed in Section 1.3. In Section 1.4 objective and scope of the research are determined. Based on the formulation of the research so far the main research question is formulated in Section 1.4. As the direction of research is determined, the methodology to execute the research is discussed in Sec 1.6.

1.2 MOTIVATION

Decisions are implemented and executed, in every task or aspect. But in order to make the right decisions, a system needs to go through an arduous and painstaking amount of time and efforts by identifying resources, gathering information, and assessing alternative resolutions to come up with a solution. This is defined as the decision-making process and the addressed system is the research is the Pavement Management System (PMS).

Until now, public authorities like in The Netherlands (Rijkswaterstraat and the Province of South Holland) have relied upon process-driven (top-down) strategies to make the *right* decisions. In order to monitor the performance of all the underlying assets (expressways, highways, provincial roads, etc.) the authorities quest to obtain data that is accurate, reliable and complete. The entire life-cycle of all the public assets (i.e pavements, bridges, locks, etc.) depends upon the acquired data and how it influences the decisions taken by the authorities. To do so, a PMS utilizes a performance model to support their decision-making process because it estimates the long-range investment requirement and the consequences of budget allocation for maintenance treatments of a particular road segment on the future pavement condition (Amin, 2015).

Essentially, the accuracy and reliability of a performance model (as seen in Fig. 1.1) depends upon the collective insight provided by the three types (probabilistic, regression and classification) of models, for assessing the pavement condition. The purpose to assess the pavement condition is to plan maintenance and rehabilitation requirements and to monitor the pavement network condition (Hein & Watt, 2005a).

Deterioration in the pavement surface(s) can be induced by several factors (Section 1.2) thus demanding periodic monitoring procedures for acquiring pavement condition data. As a result, manual and automatic pavement monitoring methods are deployed to assess the pavement condition to maintain the dense network above the minimum acceptable condition.

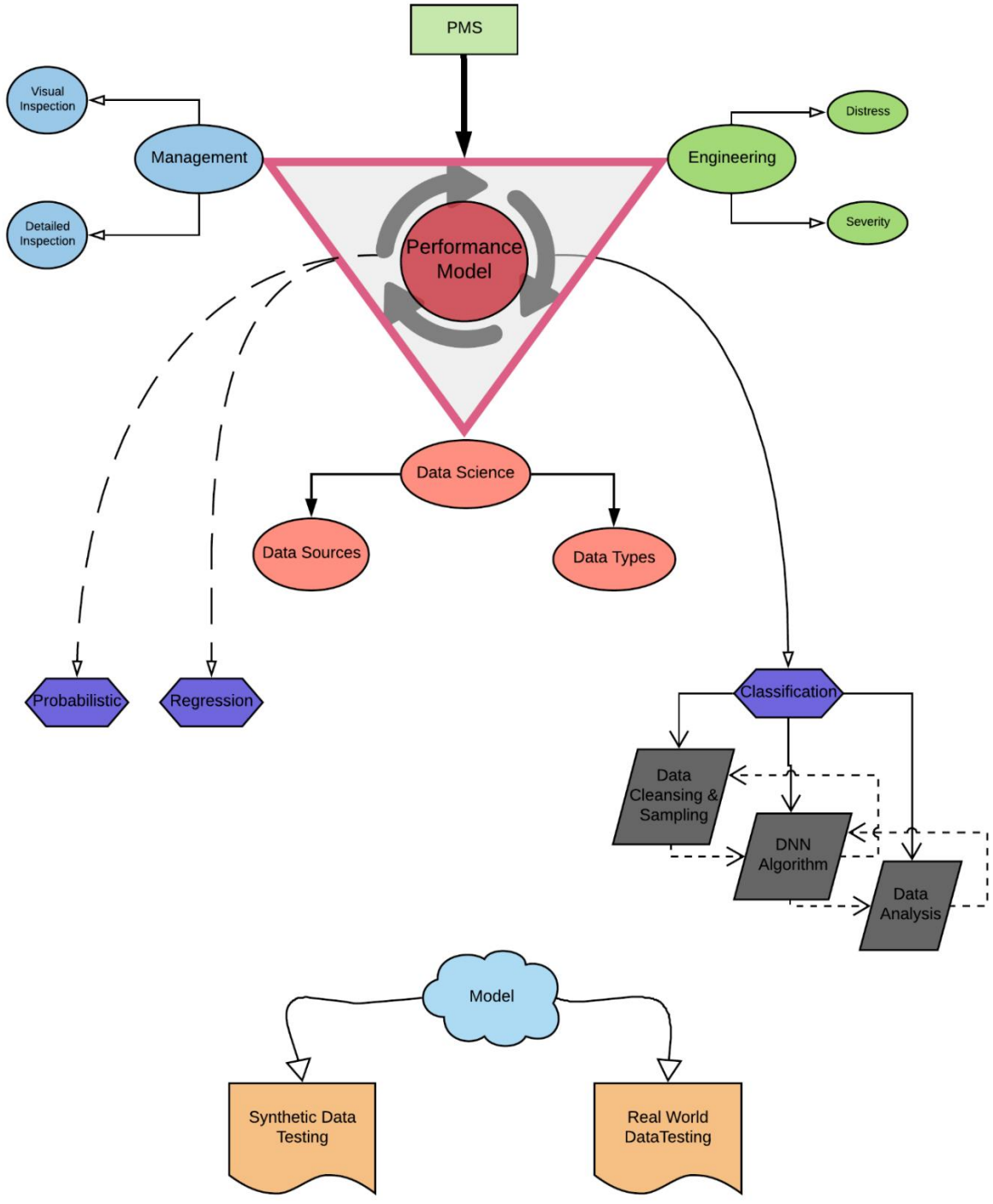


Fig. 1.1 Research Introduction

Classification of pavements based on their condition and their performance is a subject undergoing intense study. Many techniques have been used for pavement condition classification, where (Eldin & Senouci, 1995) developed pavement condition rating index based on back-propagation neural network method, (Sagheer, Kubaisy, Y.A., & Awad, 2008) developed a knowledge-based system to diagnose distresses in flexible pavements using logic programming, (Terzi, 2006) utilized data mining techniques for flexible pavement to model pavement serviceability index.

Many indices that combine the effects of all distresses found in pavement sections are being used by highway authorities for estimating maintenance need priorities for each section (Mahmood, Rahman, Nolle, & Mathavan, 2013). Thus, the classification models of pavements and of the entire network, based on their condition and distresses governs further rehabilitation planning and decision-making procedure. To objectively analyze the reliability and accuracy of the performance model for the Province of Zuid-Holland (PZH), a data-driven (machine-learning) approach is utilized to classify the pavement roughness index (Section 3.2, 3.3) and the surface distresses (Section 3.4).

1.3 PROBLEM STATEMENT

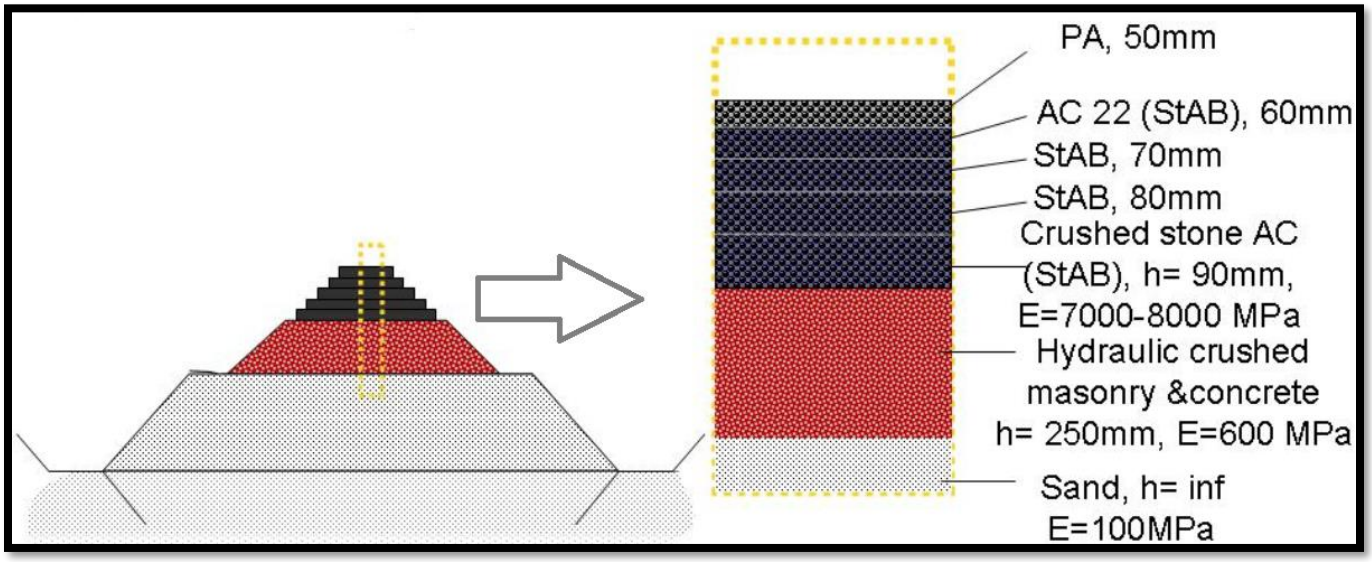


Fig 1.2 Typical cross-section example of Dutch asphalt pavement

Distress in pavements can result from multiple factors (heavy traffic load, weather conditions, aging, etc.) and can affect the entire pavement at different layers resulting in the further deterioration process. Figure 1.2 displays a cross-section view of a typical flexible pavement describing the surface and sub-base layer of a standard flexible pavement. In a hierarchal

manner, the PMS has further sub-divided the distresses into two main categories: structural and surface distresses. When structural elements i.e. the foundation, cannot handle the exerted force caused by vehicle loading, structural distress appears. Surface related damages are those that are caused by wearing out of the binding component, the bitumen, in asphalt pavements due to contact with the vehicle and natural elements (i.e. ultraviolet rays, moisture, oxygen) and incidents (TNO, 2010).

Deterioration in asphalt roads can be caused by a combination of multiple factors, for example, traffic loads, environmental factors, poor road design methods and poor quality of construction. Environmentally induced distress such as weathering, moisture, and aging of the pavement. Loading causes structural-oriented distress. In this way, pavement failure happens from both loading and weathering (Lavin, 2003).

The Netherlands has one of the world's densest road networks. Of about 130,000 km is hard-surfaced, including 5,000 km of national trunk roads as per ("Asphalt and road building | TNO," 2019). For the past three decades, detailed and accurate information about road conditions has served as the basis for transportation planning and management (Radopoulou & Brilakis, 2016). The responsibility of all the asset managers, asset owners, public and private organizations throughout the Netherlands is to keep track of such densely populated road network and its deterioration.

A road network structure is among the most critical elements of road safety and can usually affect the quality of life. Therefore, road networks should be regularly maintained and repaired to reduce accidents; however, maintenance is highly expensive, labor-intensive and time-consuming (Schnebele, Tanyu, Cervone, & Waters, 2015). Not only, the factors influencing the pavement deterioration has remained a challenging task for the field engineers to analyze but also the subjectivity with which this surface distresses are inspected and graded.

Considering all the aspects stated above, it becomes extremely important to regularly inspect the asphalt pavements in an objective manner for the identification and classification of distresses, in a way that can be reliable, cost-effective and frequent compared with traditional visual inspection methods (Section 2.1.1).

1.4 OBJECTIVE & SCOPE

The purpose behind pavement condition assessment for the PMS is two-folded: first to monitor the pavement network condition and secondly to recognize maintenance and rehabilitation requirements (Hein & Watt, 2005b). Efficient condition monitoring strategies can aid in a

significant reduction in pavement life-cycle maintenance costs by developing appropriate scheduling of pavement maintenance and repair activities (Gopalakrishnan, Khaitan, Choudhary, & Agrawal, 2017). Visual inspection methodologies (Section 2.1) are thus implemented in a process-driven procedure to detect, classify and assess the severity of the damages.

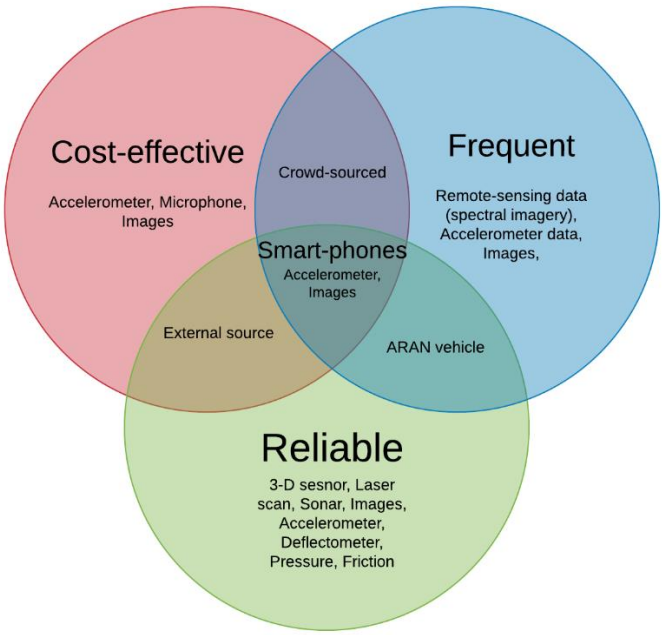


Fig 1.3 Research Scope

The general goal of all practices is to detect damages in 'early' stages but in a cost-effective, reliable and frequent manner. Thus the research objective is to investigate the possibility to develop an automatic data-driven model that is reliable and accurate enough to detect and classify pavement surface distresses in a completely objective manner. The research proposal includes a vibration and vision-based road inspection model that is capable of detecting, classifying, and segmenting road distresses. As smartphones are becoming the next modern data-source capable of generating a vast amount of data with their inbuilt sensors (Section 2.4.1) and optical instruments (Section 2.4.2), the research scope (as seen in Fig. 1.3) will be focussed upon studying the data-types (like accelerometer and images) for analyzing the pavement surface distresses which are commonly found in PZH.

1.5 RESEARCH QUESTIONS

In order to resolve the research problem explained in Section 1.2 and based on the scope and objective from Section 1.3, the main research question will be formulated. To answer the main research question, it will be further sub-divided into sub-research questions which will be studied chronologically. The main research question is:

Research Question:

Can we utilize modern data-sources such as smartphones for detecting, classifying and segmenting pavement surface-distresses?

Sub-questions:

I. *How can accelerometer data be utilized for anomaly detection and roughness prediction?*

II. *How can image data be utilized for classifying pavement surface distresses?*

III. *How can an image data be utilized for segmenting pavement surface distress?*

1.6 METHODOLOGY

The research methodology which will be utilized to answer the aforementioned research questions can be seen in Fig 1.3. The entire methodology will be iteratively executed in a 5 step procedure:

1. Literature survey: Desk research on the surface and structural distress distinguished by earlier studies along with the inspection strategies utilized by the authorities in the Netherlands to manage the distresses.
2. Data collection: Accelerometer data (vibration) and imagery data will be collected by utilizing modern data-sources such as smartphones.
3. Model development: Three Machine-Learning approaches will be utilized for predicting IRI, classifying distresses and their segmentation for severity ranking.
4. Prototype: The model will be trained and tested on the given dataset by tuning the relevant hyper-parameters.
5. Conclusion: The implications and outputs of the model will be presented.

If necessary, the proposed methodology will be utilized in an iterative manner with the corresponding dataset type.

Fig 1.4 Proposed Research Methodology

Chapter 2: Literature review & current practices

2.1 INSPECTION STRATEGIES

The PZH serves as a tier of government, situated right between the municipal authorities and the national government. Like all other provincial authorities, they are responsible for the construction and maintenance of roads. In order to provide optimum performance, in the matter of riders' safety and comfort, the road network needs to be regularly inspected and maintained. As a reason, PMS develop inspection strategies and create maintenance plans based on the different types of visual inspection methodology required for the use-case.

In order to make decisions about pavement maintenance planning, the asset managers need to regularly assess the pavement condition of the road network. Fig. 2.1 shows the road network that has been considered in this research. The color contours of the figure show an overview of the visual inspection outcomes from the 2018 database of PZH. In practice, such a plot serves as a guiding tool for coming up with adequate pavement maintenance planning & strategies.

2.2 VISUAL INSPECTION METHODOLOGY

Visual inspections can either be performed manually or automatically. The purpose of visual inspections is to assess, recognize and record structural and surface damages. To do so, the roads are examined on technical measures and the damages are assessed in qualitative and quantitative terms. According to the CROW2011(a) standards (CROW, 2011), followed by the provincial authorities, visual inspections vary broadly, in order to introduce different inspection strategies to monitor the road network. The different types of Visual inspection strategies can further be categorized into four major categories:

1. Global inspection
2. Minor maintenance inspection
3. Detailed inspection
4. Safety inspection

Below is a brief description of all the visual inspection strategies and implementation techniques by PZH.

Global Inspection:

Within the road management methodology of PZH, global inspections play an important aspect. The purpose of this type of inspection is to collect information quickly and efficiently at a network level about the pavement condition of the entire road network as seen in Fig. 2.1 It is mainly followed for two reasons:

Fig. 2.1 Visual Inspection Data from ARAN 2018 Database

I. To develop network plans and budget
II. To develop quality and safety overviews.

According to the CROW2011(a) standards (CROW, 2011), it is recommended to conduct the global inspection annually for the entire road network. This helps in providing acute information about the location and the extent of the damage (as seen in Fig. 2.2). As a suggestion, the need for minor maintenance is also indicated during the global inspections. For a detailed explanation, a separate manual has also been issued for the practical implementation of global inspection (CROW publication 146b 'Manual for global visual inspection').

Fig. 2.2 Inspection in progress (CROW, 2011)

Minor maintenance inspection

Minor maintenance inspections are carried out to detect and record minor defects in a structural manner, which if detected, must be repaired in the corresponding financial year. The objectives fulfilled by this type of inspections are:

I. Prevention of reduced traffic safety and driveability
II. Preserving the structural value of the road construction.
III. Minimizing risks in the context of liability statements.

The benefits of minor maintenance are to repair and preserve the damages before they spread and reduce the quality of the road section. Thus the lifespan of the road section is extended. As a reason, the minor maintenance inspection is carried out a number of times per year. To conduct the minor inspections every manager is entitled to use system utilities, which fits within the organization of his service.

Detailed Inspection

It is considered as the most extensive and accurate inspections, out of all other Visual Inspection methods. The purpose of the detailed inspection is to accurately record the visual condition of the pavement at the project level. The definitions of the severity classes that have been drawn up for this type of inspection have also been used for global inspection and the small maintenance inspection. Although the detailed inspections itself have no place within the PMS, it is used for specific purposes like creating liability statements, paving advice and for research at a project level. Also, there is no fixed frequency of execution. This type of inspection is carried out according to the specific needs of the manager.

Safety Inspection

The regular safety inspection is carried out throughout the year. In fact, the method of implementation is rather cost-free. This type of inspection aims at the safety of road users and the prevention of claims and liability. The main concern is the identification of damages that could result in an accident. The identification of such damages does mean that the damage must be repaired immediately.

Up until now, the different types of visual inspection techniques that are formulated by the PZH are described. In order to further understand their execution procedure, it is vital to recognize factors that influence (directly or indirectly) the outcome from the visual inspections.

2.3 FACTORS INFLUENCING VISUAL INSPECTION

Factors that determine the quality of inspection is also dependent on the method of implementation. The quality of results received through the inspections depends on a number of factors, namely:

I. The quality of the manual and the damage catalog
II. The conditions of performance or inspection
III. The quality of inspectors
IV. The accuracy of the database.

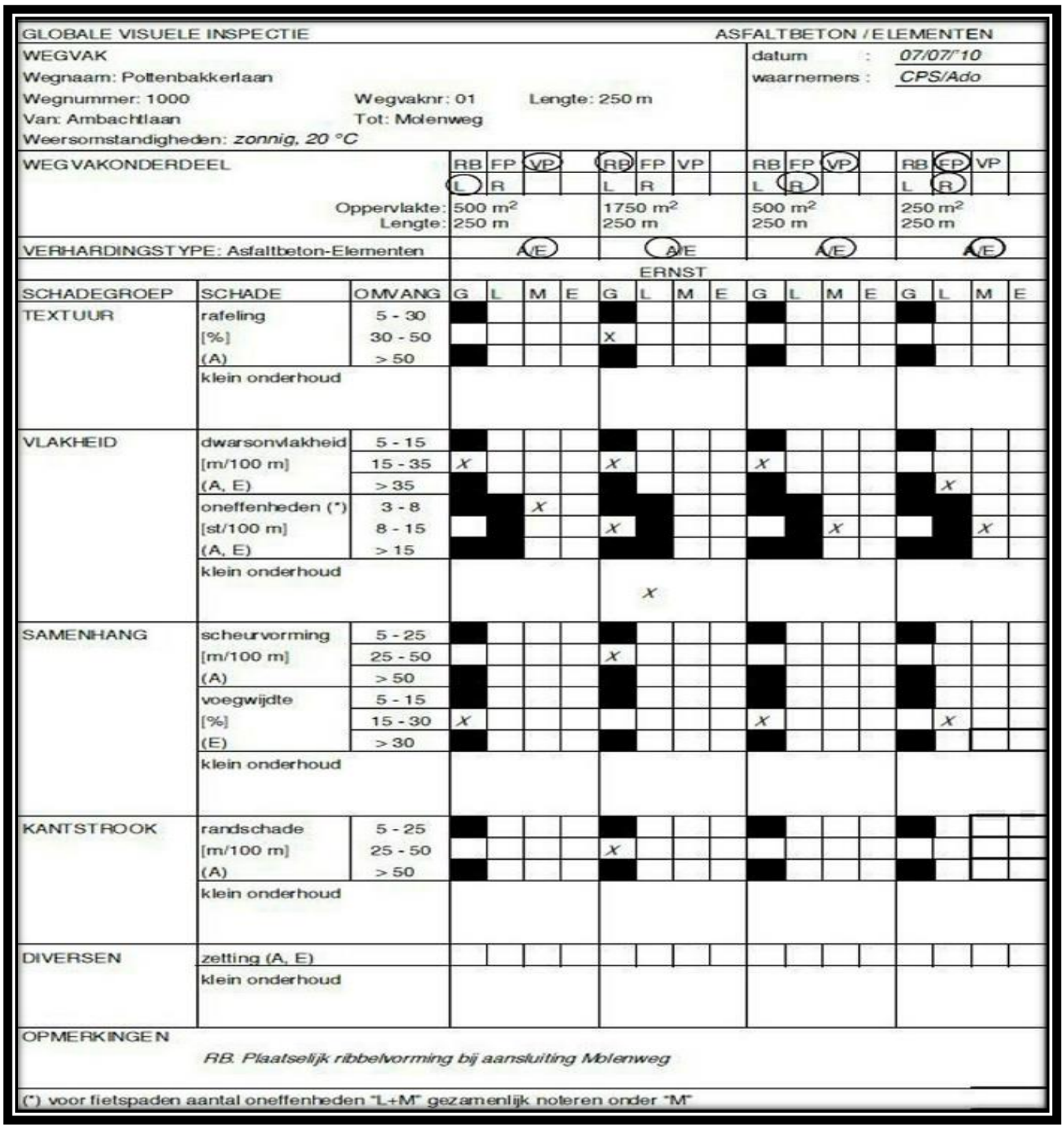
A detailed overview of the Global Inspection techniques is included in (Appendix A). Here, we will summarize the governing/crucial factors for conducting a global inspection and will identify the inspection templates method utilized by the authorities to assess pavement distresses.

The quality of the manual and the damage catalog is dependent on the road sections. In principle, the section must be homogenous and the length of the section has an influence on the accuracy of the inspection result. The zeroth point (start) and the boundary (end) are recorded unambiguously.

The condition of performance or inspection is also directly dependent on weather influence, frequency period and the number of equipment desired to conduct such maintenance strategies. As per the guidelines specifications, the inspections must be carried out on a dry road surface for the proper inspection of the damages. The frequency for global inspection is set at once every two years and can only be executed in a definite season interval of a year (March till October). Winter months are avoided due to poor inspection conditions (moisture, frost, leaf fall, etc). The types of equipment that are needed to perform such inspections are extensive (for e.g. Global visual Inspection manual, inspection forms, spirit level, measuring rod, safety vest/jackets, a photo camera, permits and exemptions, flashlight, etc).

The quality of inspectors is crucial and the subjective part of the inspections. According to the guidelines, there should be a team of two inspectors in a moving car or by the inspector(s) on foot. A sequence or route should be pre-decided before the inspection should begin. It is recommended to perform the measurements for eight hours a day with a single hour of break. The output should be around 20-40 km of road length for the outskirts and 10-20 km for the residential area. Also, it is extremely important for the inspectors to be at least of a technical MBO level and are trained periodically for accurate judgments of distresses.

The last dependency for performing a visual inspection is on the accuracy of the database that is created and maintained for performing maintenance on the damaged roads, which needs to be repaired in the current financial year. The following Fig 2.3 reflects a filled (completed) template for the global inspection:



GLOBALE VISUELE INSPECTIE — ASFALTBETON / ELEMENTEN

WEGVAK
Wegnaam: Pottenbakkerlaan
Wegnummer: 1000 Wegvaknr: 01 Lengte: 250 m
Van: Ambachtlaan Tot: Molenweg
Weersomstandigheden: *zonnig, 20 °C*

datum : *07/07/'10*
waarnemers : *CPS/Ado*

WEGVAKONDERDEEL	RB (L)	FP R	VP	RB L	FP R	VP	RB L	FP (R)	VP	RB L	FP (R)	VP
Oppervlakte:	500 m²			1750 m²			500 m²			250 m²		
Lengte:	250 m			250 m			250 m			250 m		
VERHARDINGSTYPE: Asfaltbeton-Elementen		(A/E)			(A/E)			(A/E)			(A/E)	

SCHADEGROEP	SCHADE	OMVANG	ERNST G	L	M	E	G	L	M	E	G	L	M	E	G	L	M	E
TEXTUUR	rafeling	5 - 30	■				■				■				■			
	[%]	30 - 50					X											
	(A)	> 50	■				■				■				■			
	klein onderhoud																	
VLAKHEID	dwarsonvlakheid	5 - 15	■				■				■				■			
	[m/100 m]	15 - 35	X				X				X							
	(A, E)	> 35	■				■				■				■	X		
	oneffenheden (*)	3 - 8	■	■	X		■	■			■	■			■	■		
	[st/100 m]	8 - 15		■			X	■				■	X			■	X	
	(A, E)	> 15	■	■			■	■			■	■			■	■		
	klein onderhoud							X										
SAMENHANG	scheurvorming	5 - 25	■				■				■				■			
	[m/100 m]	25 - 50					X											
	(A)	> 50	■				■				■				■			
	voegwijdte	5 - 15																
	[%]	15 - 30	X								X					X		
	(E)	> 30	■				■				■				■			
	klein onderhoud																	
KANTSTROOK	randschade	5 - 25	■				■				■				■			
	[m/100 m]	25 - 50					X											
	(A)	> 50	■				■				■				■			
	klein onderhoud																	
DIVERSEN	zetting (A, E)																	
	klein onderhoud																	

OPMERKINGEN

RB. Plaatselijk ribbelvorming bij aansluiting Molenweg

(*) voor fietspaden aantal oneffenheden "L+M" gezamenlijk noteren onder "M"

Fig. 2.3 A filled template from the Global inspection manual (CROW, 2011)

Fundamentally, for every road-section (*wegenvakonderdeel*), a road-type (*verhardings type*) is noted which is either a roadway=RB, cycle path=FP or a footpath=VP. Then the damages are grouped (*schadegroep)* by their characteristics based on texture (*textuur),* flatness (*vlakheid),* consistency *(samenhang),* edge strips *(kantstrook)* and others *(diverse).* For every group, a damage type *(schade)* is recorded with respect to their severity/size *(omvang).* The damages/distresses that

can be recorded by utilizing the above-mentioned template are Ravelling, Cross fall, Crack formation, Edge damages, and others.

Although the global inspection techniques require extreme care during surveillance and for recording damage information in a traditional way, they are still implemented and utilized for drawing up budget plans. But in order to accurately analyze damage assessment, a detailed inspection is deemed necessary.

In global inspections, a counter technique is utilized to overcome the shortcomings of manual inspection methodology. Automatic pavement damage detection is a method in which highly sophisticated vehicles like ARAN (Automatic Road Analyzer) are utilized.

To analyze the failure modes of pavements, these ARAN vehicles (Fig. 2.4) are supported with a dedicated modular system of hardware and software add-ons to meet the various infrastructure data collection needs. The available sub-systems in an ARAN vehicle are as follows:

- Linear Referencing System (LRS) positioning.
- High-definition cameras to complete the right-of-way panoramic view.
- GPS positioning
- Roughness (IRI)/ longitudinal processing
- Rutting/transverse profile
- Texture data
- Distress testing
- Ground-penetrating radar (GPR)
- Pavement-marking reflectivity
- Geometrics
- Lidar
- Video-imaging
- Laser SDP
- Laser Transverse profiler (XVP)
- Pave 2D
- Pave 3D
- Position and orientation system for land vehicles (POS LV)
- Distance measuring instrument (DMI)

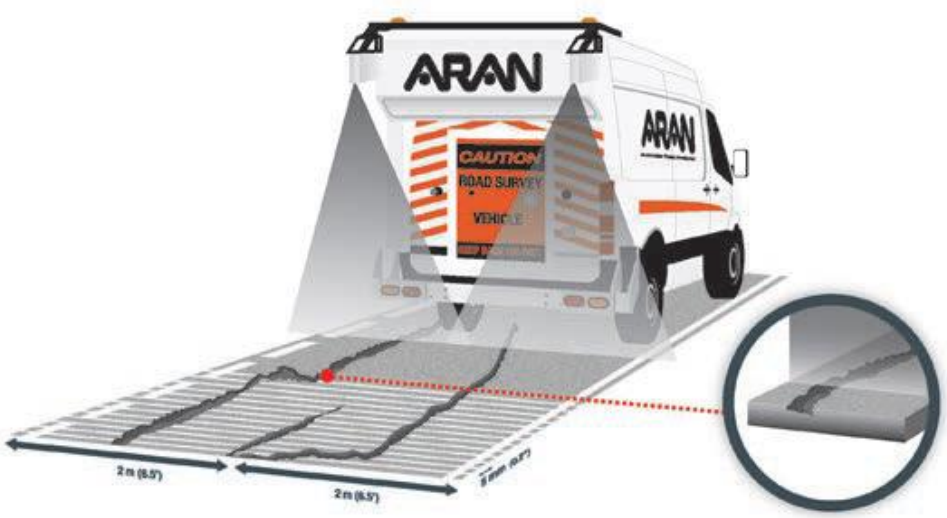


Fig 2.4 ARAN vehicle (Fugro, 2019)

2.3 DISTRESS TYPES

Whether it is the global inspection taken manually or an automated detailed inspection taken by an ARAN vehicle, distress monitoring is the essence for every type of visual inspection. Based on the extensive literature the various type of pavement distresses are shortlisted below:

I. Cracking
    a. Alligator Cracking
    b. Block Cracking
    c. Joint Cracking
    d. Longitudinal Cracking
    e. Slippage Cracking
    f. Transverse Cracking
II. Raveling
III. Bleeding
IV. Polished aggregate
V. Rutting
VI. Bumps
VII. Corrugation-shoving
VIII. Potholes
IX. Other Road Performance Indices:
    a. IRI (International Roughness Index)
    b. Skid Resistance

c. Patching
d. Manhole
e. Railroad crossing

As these distresses (mentioned above) have different behavioural characteristics they are further categorized based on the impact they have on the surface, sub-structure or the performance of the pavement. The classification table is based on the category of distress, their severity ranking, the type of information needed to assess the distress and their impact on the safety and comfort of the rider. The table describing the distresses into a categorical manner with their corresponding severity ranking, impact on safety, impact on comfort and the type of information required to assess them is listed under:

Category	Distress	Sub-category	Severity	Impact on Safety	Impact on Comfort	Type of Information
Surface Distress	Cracks	Block Cracking	L	0	0	2D 3D
			M	0	1	
			H	1	2	
		Joint Cracking	L	0	0	2D 3D
			M	0	1	
			H	1	2	
		Longitudinal Cracking	L	0	0	2D 3D
			M	0	1	
			H	1	2	
		Slippage Cracking	L	0	0	2D 3D
			M	0	0	
			H	1	1	
		Transverse Cracking	L	0	0	2D 3D
			M	0	1	
			H	1	2	
	Corrugation/Shoving		L	0	1	2D 3D
			M	1	1	
			H	2	2	
	Rutting		L	0	1	3D
			M	0	1	
			H	1	2	
	Raveling		L	1	0	2D
			M	1	0	
			H	2	0	
	Polished Aggregate		L	1	0	2D
			M	1	0	
			H	2	0	
	Bleeding		L	1	0	2D
			M	1	0	
			H	2	0	
Structural/ Visco Plastic Distress	Potholes		L	1	1	2D 3D
			M	1	1	
			H	2	2	
	Bumps		L	0	1	3D
			M	0	1	
			H	1	2	
	Aligator cracks		L	0	0	2D 3D
			M	0	1	
			H	1	2	
Others Road Perfromance Indices	IRI			1	1	2D 3D
	Patching			1	1	
	Skid Resistance			1	1	
	Manhole			1	1	
	Railroad Crossing			1	1	

Table 2.1 Pavement Distress Category overview

2.4 DATA TYPES

For all the types of distresses as classified above, public authorities have come up with distress detection methods like (visual inspections), which in turn utilizes dedicated instruments/sensors/data-equipments/data-generators to assess a single or multiple types of damage(s). (Coenen & Golroo, 2017) can further correlate particular distress types with their corresponding data-type(s)/equipments, which have been commercially utilized and scientifically researched in the pavement distress detection domain. The table is as shown under:

Distress and their detection equipment relation

	Camera	Accelerometer	3-D sensor	Microphone	Sonar	Pressure	Friction	Deflectometer
Cracking	C/P		C/P					
Patching	C/P		C					
Potholes	C/P	C/P	C/P	P	P	C/P		
Rutting	P		C/P		C			
Shoving	C	P	C/P			P		
Bleeding	P		C/P					
Polished aggregate			P	C/P				
Ravelling			C/P	C				
Edge drop-off			C		C			
Water bleeding and pumping			C	C				
Additional road information								
IRI	C	C/P	C/P	C/P		C/P	C	
Skid resistance			C				C	
Substructure			C/P					C
Other	C	C	C/P	P	C			C

Notes: C = commercial development; P = from scientific research papers.

Table 2.2 Distress and data types (red box symbolizes the research scope) (Coenen & Golroo, 2017)

Multiple ways and techniques are utilized to identify and detect pavement distresses by utilizing modern data sources/generators. Optical instruments and sensors like a camera and accelerometer sensors-based methods are currently being assessed and studied by multiple research communities and are put in practice by commercial organizations over the globe. According to the research scope (as in Section 1.3), a brief understanding of the two selected techniques and how they can be utilized for their associated distress is further explained:

1. **Accelerometer (Vibration-based) method**
2. **Camera-based method**

2.4.1 Accelerometer (Vibration-based method)

Distress that creates a vertical movement in the rider's comfort is a symbol of increased vibration. If the vibration is caused by the road's roughness then it can be measured by monitoring the tire sound, tire pressure, vibration in the tire, the axel or the car as a whole system. Henceforth, the most commonly used sensor for identifying the vibration anomalies caused by the roughness of the road will be studied.

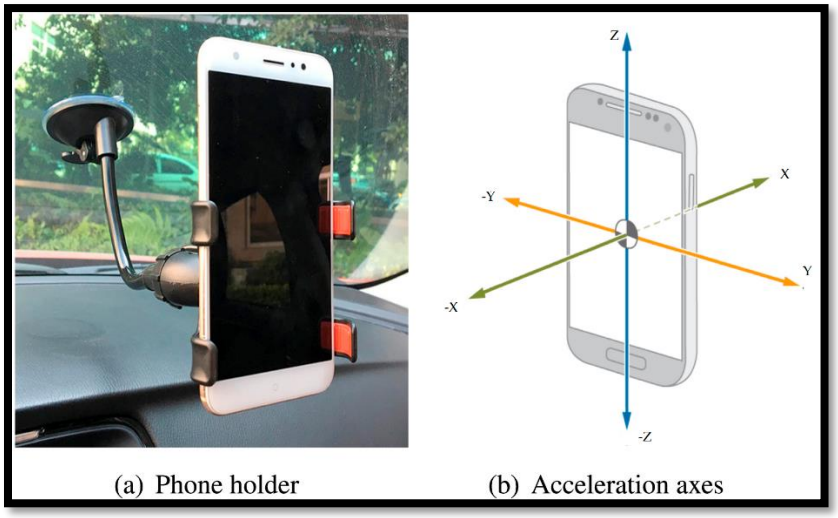


Fig. 2.5 Accelerometer axis diagram (Souza, 2018)

Accelerometer sensors measure the acceleration in three dimensions of a moving object as can be seen in Fig. 2.5. Multiple research has tried utilizing the accelerometer sensor to detect potholes (Casas-avellaneda & Lopez-parra, 2016) which are in-built sensors in most of the smartphones while riding in a vehicle. Though, there are few challenges like the pothole could have missed one of the axels or the entire vehicle path.

Companies as for example Roadscanners ("Home - Roadscanners," 1998), PaVision ("PaVision® Automated Pavement Condition Assessment System | www.ara.com," 2015), DynaTest ("Dynatest | Pavement Evaluation & Consulting," 1976), ARRB Hawkeye ("Hawkeye Software - ARRB Group," 2014), ROMDAS ("ROMDAS | Road Survey Equipment , Pavement Condition Survey," 1990), and International Cybernetics ("About Us | International Cybernetics," 1975) are deploying accelerometer sensor data as a support tool for calculating the International Roughness Index (IRI) of roads (Coenen & Golroo, 2017).

2.4.2 Camera-based method

Methods of collecting images via optical instruments like a single camera (2-D camera), video camera, line-scan camera, infrared camera, stereo-imaging camera, photometric stereo camera, 3-D depth sensor camera, etc. have remained in practice for detecting road damages.

Cameras have been commonly used for detecting different distresses on the basis of two-dimensional (2-D) images all over the globe. The single mounted, high-speed charge-coupled device (CCD) is used from the past 20 years to detect distresses and in particular crack and potholes (Li, Q., Zou, Q., Zhang, D., & Mao, 2011), (Mancini, A., Malinverni, E. S., Frontoni, E., & Zingaretti, 2013), (Moghadas Nejad, F., & Zakeri, 2011), (Premachandra, C., Premachandra, H. W. H., Parape, C. D. & Kawanaka, 2015), (Puan, O. C., Mustaffar, M., & Ling, 2007), (M. Wang, Birken, & Shahini Shamsabadi, 2015), (Yoo, H.-S., & Kim, 2016)

Studies like in the one conducted by Commonwealth Scientific and Industrial Research Organization CSIRO ("Road crack detection," 2019) implemented high-speed cameras mounted underneath the vehicle to capture high-resolution images in small patches, these are later consolidated into bigger images of half-meter interval. The CSIRO system could detect cracks in a fully automated way while combining machine learning and computer vision techniques.

In order to clearly visualize cracks, unwanted shadows have to excluded and extra noise from the pictures was also removed by using very strong illumination (Mathavan, S., Kamal, K., & Rahman, 2015). Also, (Cord, A., & Chambon, 2012) and (Oliveira, H., & Correia, 2013) have utilized built-in smartphone cameras for image collection, which had limitations over lens quality thus only focusing upon image processing rather than data collection.

2.5 SUMMARY

Newly paved roads deteriorate very slowly and almost imperceptibly in the first ten to fifteen years of their life, and then deteriorate much more rapidly unless timely maintenance is undertaken. One of the planning strategies developed by authorities like visual inspection requires a team of dedicated trained members to inspect the roads, and then later face their subjective opinion on quantifying the pavement distresses despite the pre-defined distress templates.

For all the types of distresses which can occur on roads and are identified above (in Section 2.3), there are specific data types and instruments which are vastly utilized for commercial purposes and for research methods as well (in Section 2.4). Acceleration (vibration) based methods have

been studied to detect anomalies in the road profile. Moreover, they are also studies showing the relevance for calculating the international roughness index (IRI) by measuring the raw data input from the accelerometer sensor. Multiple organizations and research studies have also focused upon camera-based methods for detecting pavement distress on pavement surfaces. Cracks and potholes detection were the major types of distresses that were studied in the camera-based techniques but certain limitations (Section 2.4.2) have restricted further development in the field.

So far, all the visual inspection strategies have remained a core maintenance strategy for the public authorities to inspect and manage pavement distresses through the dense network of roads. Further in this research, the author will further investigate how a PMS can utilize a data-driven approach for anomaly detection and pavement surface distress classification, utilizing the most abundant and crowdsourced data generator i.e smartphones. The research will also provide validation and accuracy measurements for each studied data type and its corresponding distress.

Chapter 3: Research methodology

3.1 DATA-DRIVEN MODEL

The process to construct a data-driven model is to first define the domain (Use-case) and the implication it ought to achieve. Once that is affirmed, the second step is to explore the underlying data and associate it with the *right* approach (i.e. supervised, unsupervised, etc.). The next step is to choose an appropriate algorithm based on the applicability and performance. Once the algorithm is chosen, the fourth step is to extract the needed features from the data by utilizing techniques like (transforming, augmenting, pre-processing). Based on the extracted features and needs of the use-case, the ML model is then constructed in step 5. The second last step involves iterating the algorithm with extracted features for multiple times until the model achieves its best accuracy. The last step is to evaluate the performance of the model and its applicability. The 3 steps (post the results' step) are associated with building the model for production purposes and hence will not be considered further in the research. The basic steps to build any ML models from scratch until the end can be visualized in Fig 3.1 below:

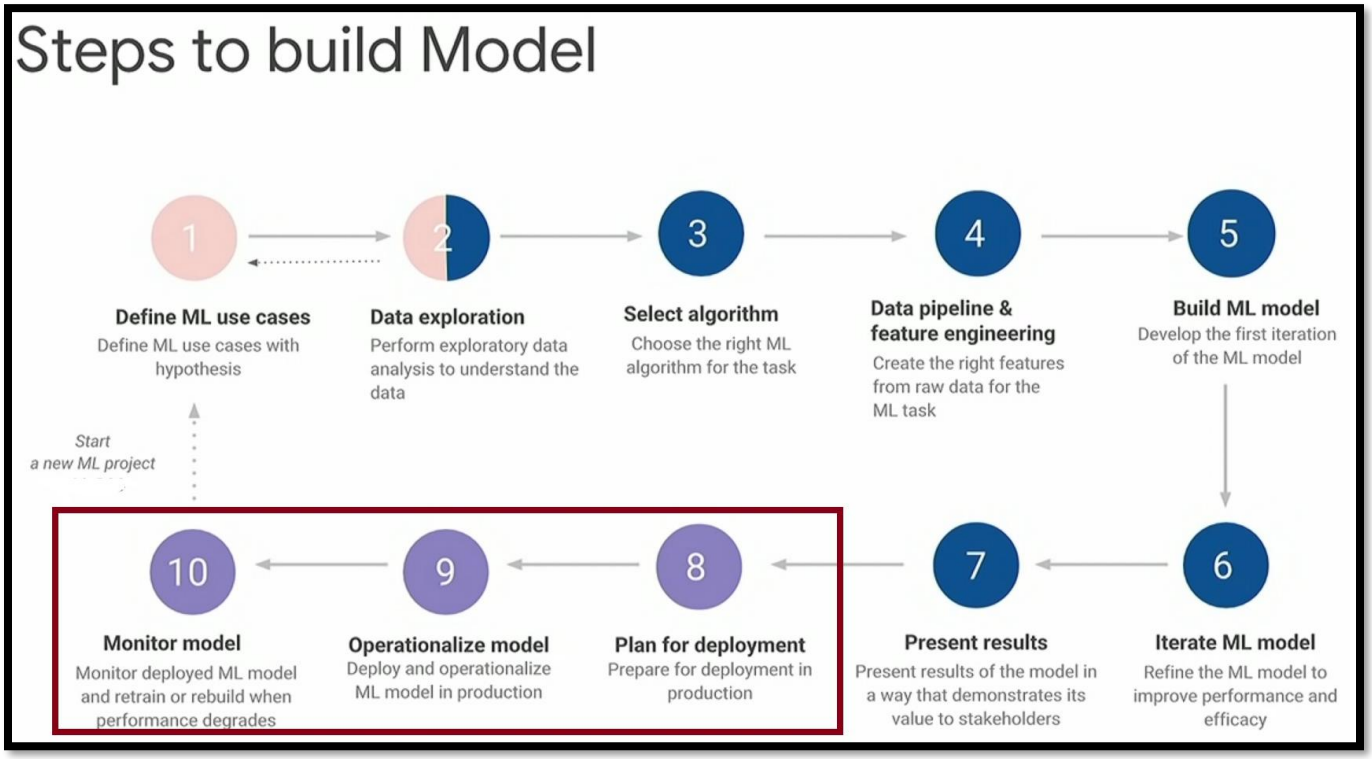


Fig 3.1 A flow diagram showing the steps to build an ML model

As the process to create a model is determined, a particular data-approach will be utilized depending on the dataset availability. The research will utilize an unsupervised learning approach for the accelerometer data (in Section 3.3), a semi-supervised approach for distress classification utilizing images (in Section 3.4) and lastly for distress segmentation a complete supervised approach has been implemented to get the desired output (in Section 3.5).

3.2 DATA-DRIVEN APPROACH

A data-driven model is based on the analysis of the data about a specific system. The main concept of the data-driven model is to find relationships between the system state variables (input and output) without explicit knowledge of the physical behavior of the system (Solomatine & Shrestha, 2009). Machine-Learning (ML) theory is related to pattern recognition and statistical inference wherein a model is capable of learning to improve its performance of a task on the basis of its own previous experience (Mjolsness & DeCoste, 2001).

The best practices within the field of machine learning to build a data-driven model is determined by the underlying data. Based on the data structure and amount, data-models can be further categorized mainly on 4 commonly known strategies:

1. Unsupervised Learning
2. Semi-supervised Learning
3. Supervised Learning
4. Reinforcement Learning

For this research, three (unsupervised, semi-supervised, supervised) learning methods were utilized to build the models later in (Section 3.3, 3.4, 3.5). At first, a brief introduction of the (selected) data-driven approaches will be provided for understanding the merits and the demerits over one-another.

3.3 UNSUPERVISED LEARNING APPROACH

Unsupervised refers to the ability to learn and organize information without providing an error signal to evaluate the potential solution (Sathya & Abraham, 2013). The lack of direction for the learning algorithm in unsupervised learning can sometimes be advantageous since it lets the algorithm to look back for patterns that have not been previously considered (Kohonen & Simula, 1996). For unsupervised-learning processes, the relationship of causality among the variables is unknown. Therefore there are no outputs, but only inputs for which the ML algorithm determines the relationship among them (Ticlavilca Torres, Alfonso, 2008). Classification algorithms like K-Nearest Neighbor (Laaksonen & Oja, 1996), Random Forest (Breiman, 2001), Support Vector Machines (SVM) (Tong & Koller, 2011) are the few renowned algorithms in the unsupervised learning approach. Decision trees and random forest are common classifiers with widespread use (Wu, Feng, Naehrig, & Lauter, 2016). A decision tree can be mathematically expressed according to Equation 3.1.

$$h(x) = \sum_{j=1}^{L} b_j . 1_{R_j}, \tag{3.1}$$

where:

$h(x)$ = decision tree function;

L = total number of leaves;

R_j = j^{th} disjoint region of the tree;

b_j = constant assigned to the region R_j:

$\mathbf{1}_{Rj}$ = indicator function in the region R_j.

The indicator function is defined in Equation 3.2 as follows:

$$\mathbf{1}_A(x) = \begin{cases} 0, \ if \ \ x \notin A, \\ 1, \ if \ \ x \in A. \end{cases} \tag{3.2}$$

where:

A = mathematical set.

Accelerometer Data

In this section, the unsupervised learning approach is utilized to construct a data-driven model based on the input data type:

1) Use-case

Fig 3.2 Accelerometer Data Collection

The purpose of this section is to determine if the data-driven model can identify road anomalies by utilizing a smartphones' accelerometer data as an input. In data mining terms, anomaly detection (also outlier detection) is the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of data.

For the Province of Zuid Holland (PZH) the first priority is to identify anomalies in the entire road network and carefully pick and associate them with actual road performance indices. A Falling weight deflectometer (FWD) (Nega, Nikraz, & Al-Qadi, 2016) is the world standard dynamic plate bearing test for the non-destructive testing of the flexible pavement, which is utilized during the yearly inspection carried out by the public authorities. Although the precision of an FWD sensor could be the most suited for determining the road profile, it still can't be implemented frequently because of its related hefty cost and manual labor associated with the equipment.

The use-case for this research is an attempt to develop and suggest a supportive data-driven approach, to detect road anomalies automatically, by utilizing an abundant/ crowdsourced modern data-source i.e. smartphones. Fig 3.2 displays a simple mechanism that had been utilized for the data collection procedure. The accelerometer data was captured by utilizing an android application built by RoadEO.

2) Dataset collection & exploration

The dataset has been collected on a road network (Fig. 3.3.) comprising of two primary roads (N206, N208) and one secondary road (N443) in the south Holland part of the Netherlands as can be seen in Fig 3.3. The combined length of the road network together contributed to approximately 30 kilometers. Where the stretch of 19.3 km from N206, 6 km from N208 and 3.9km from N443 was recorded.

Fig 3.3 Road Network for Accelerometer data

The entire stretch of the road network was recorded 3 times to obtain the raw accelerometer values at a rate of 50Hz while driving a rented gasoline car via *green wheels*. The device type utilized for the data collection was an LG-H870 smartphone that could record accelerometer vibration data within an average range of 3 to 5 meters on location accuracy. The fluctuations in GPS accuracy varied inversely as the speed of the vehicle increased or decreased.

The raw values generated during the data collection through the inbuilt accelerometer sensor can be visualized in Fig 3.4. The recorded values (change in acceleration) are in three-dimension where x is in the forward/moving direction of the car, y is when the vehicle moved in a transverse direction (i.e. during lane changing) and z is the movement in the vertical direction from the rider's perspective.

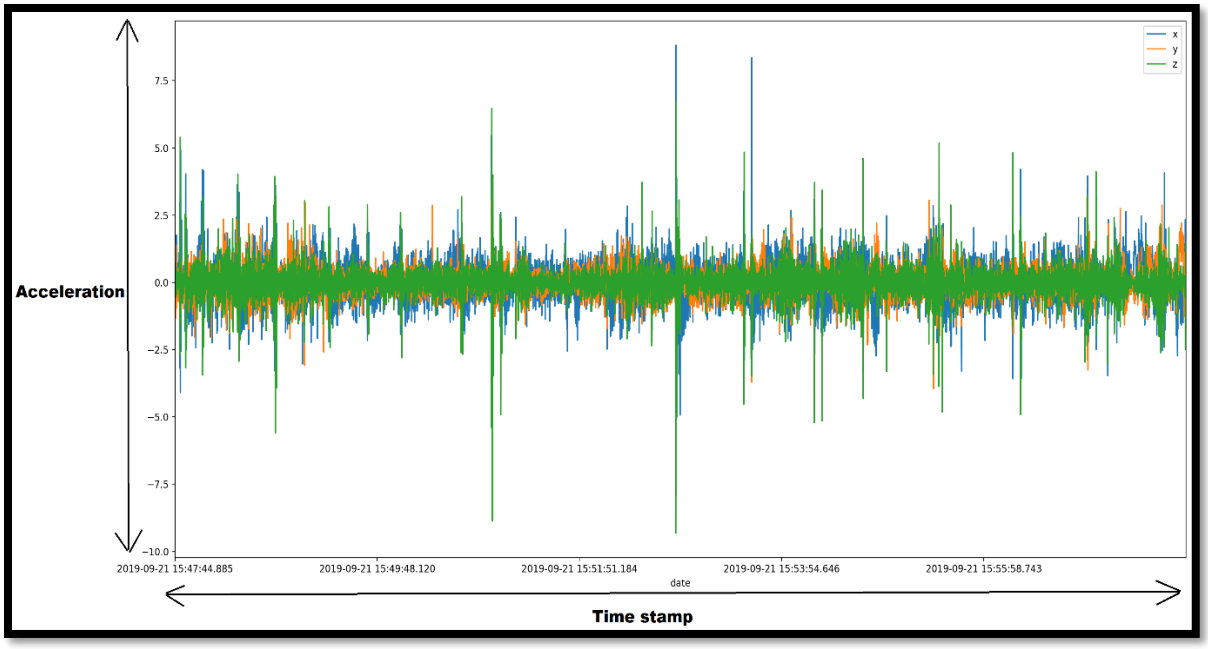


Fig. 3.4 Accelerometer data visualization (x-y-z) axis

Once the road network for the use-case is decided and the raw data is generated by using a smartphone from its inbuilt accelerometer sensor, the next step is to select an appropriate algorithm.

3) Algorithm

The algorithm selected for this use-case is the one that is commonly used in unsupervised machine learning tasks. The algorithm with a multitude of applicability and performance in this domain are K-Nearest Neighbor, Random Forest, Support Vector Machines (SVM) as seen in Section 3.3. For the further data-analysis, Random Forest or random decision forest is used. It is a technique of performing both regression and classification tasks with the use of multiple decision trees commonly known as bagging.

Features that are useful for the and are well recognized for the Random Forest algorithm are (Breiman & Cutler, 2019):

- The algorithm runs efficiently on large databases as the use-case dataset is.
- Performance-wise the other algorithms in the unsupervised learning approach achieve lesser accuracy than the random forest.
- It generates an internal unbiased estimate of the generalization error as the making of decision tree forest proceeds, which is appropriate for a large and unstructured dataset.

- Built Forest (decision nodes) can be further populated as more data is added to the use-case.
- Implements an effective method for estimating missing data and maintains accuracy when a large proportion of the data is missing, which is the nature of the dataset present for the use-case.
- Above mentioned capabilities can be extended to unlabeled data, leading to unsupervised clustering, data views, and outlier detection.

Based on the features mentioned above, the eminent algorithm (Random Forest or Random Decision Forest) has been chosen for the predefined use-case. Now, before the algorithm can be populated with its input variables feature extraction will be performed by using an appropriate transform function.

4) Feature extraction

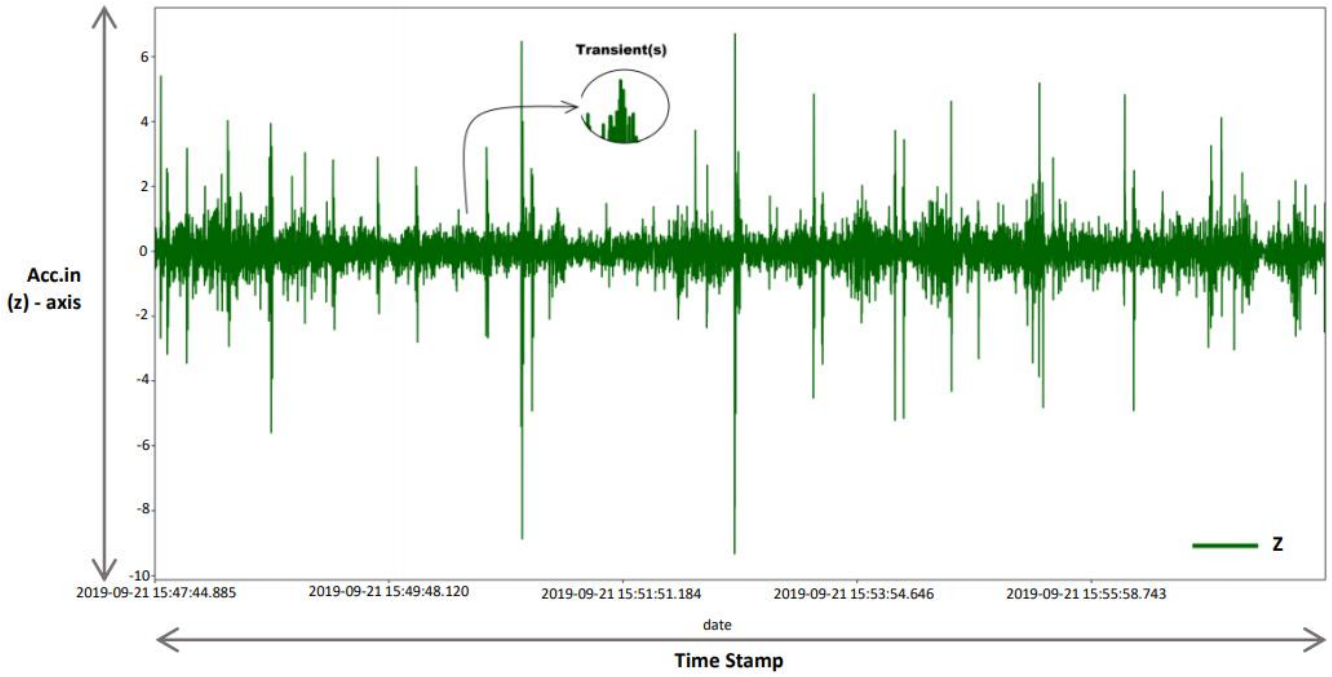


Fig. 3.5 Transient(s) in the z-axis

Feature selection or extraction is defined as a method that selects and/or combine variables into features, by effectively reducing the amount of data that must be processed, while still accurately and completely describing the original dataset.

Real-world data or signals frequently exhibit slowly changing trends or oscillation punctuated with transients as can be seen in Fig 3.5. The figure displays the change in acceleration in the z-axis (after the removal of acceleration incorporated by gravity)

alongside time. The abrupt changes/transient(s) are an interesting part of the data both perceptually and in the information they provide.

Fourier Transforms (FT) is a powerful tool for data analysis. However, it does not represent abrupt changes efficiently. The reason for such demerit is that FT represents data as a sum of sine waves that are not localized in time or space and also the sine waves oscillate forever and the accelerometer data is a time-bound three-dimensional data. Therefore, to accurately analyze vibration signals emitted through the accelerometer sensor there is a need to use a new class of functions that are well localized in time and frequency. This calls the attention towards Wavelet Transform (WT).

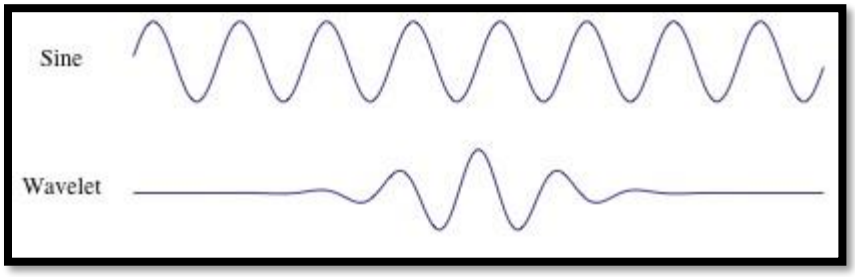


Fig. 3.6 Sine wave and a Wavelet

Wavelet transforms can be seen as an integral product of a signal to be analyzed with a scaled and a dilation version of another signal (E. & G., 2016):

$$WT = a^{-\frac{1}{2}} \int x(t)\psi^* \left(\frac{t-\tau}{a} \right) \tag{3.3}$$

A general expression for Wavelet function can be given as follows:

$$\psi_{(a,b)}(t) = |a|^{-\frac{1}{2}} \psi^*((t-b)/a) \tag{3.4}$$

A wavelet transform is a rapidly decaying wave-like oscillations that have zero mean. Unlike, sinusoid which extends to infinity (∞), a wavelet exists for a finite duration. As can be seen in Fig. 3.6 the main difference between a sine-wave and wavelet is that the former is not localized in time while the latter is localized in the time domain. Higher scale-factor (longer wavelet) corresponds with smaller frequency, thus by scaling the wavelet in time-domain smaller frequencies can be analyzed (achieving a higher resolution) and vice versa, by utilizing a smaller scale more detailed can be represented in the time domain. So, scales are principally the inverse of the frequency.

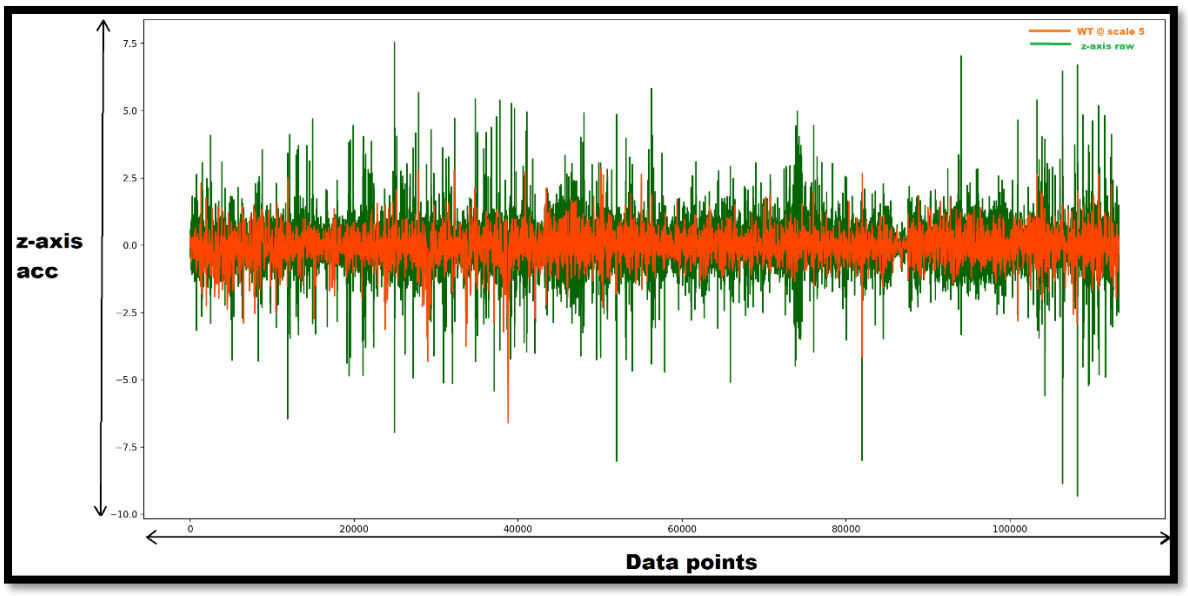


Fig 3.7 Raw & demodulated z-axis

To decompose the signal Stationary Wavelet Transform (SWT) was utilized which is a form of non-decimated Discrete Wavelet Transform (DWT). Multiple experiments were performed using different wavelet families and range. A combination of Sym5 wavelet from *symlet* wavelet family and, and at level 5 (scale) provided the best result. The raw signal and the transformed signal output can be seen in Fig. 3.7. The selection of the wavelet family and the approximate range was also validated by (Seraj, Van Der Zwaag, Dilo, Luarasi, & Havinga, 2016) studies.

5) Model development

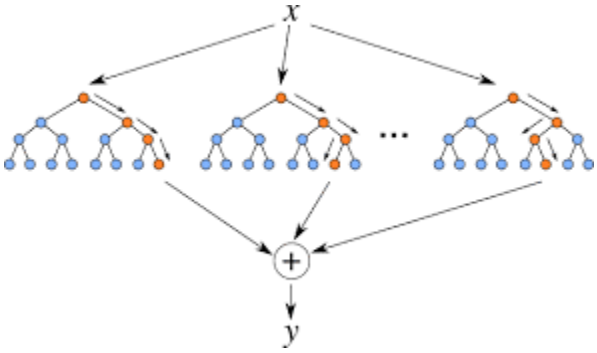


Fig 3.8. Decision Tree

Once the features are extracted using the Sym5 wavelet transform (H. Y. Lin, Liang, Ho, Lin, & Ma, 2014) over the z-axis and the algorithm is chosen (in Section 3.3.3), its time to build the decision tree. In general, the more trees (decision nodes) are there in

the forest the more robust the built model will be. The approach (Breiman & Cutler, 2019) utilized to build a decision tree (as seen in Fig 3.8) can be summarized as:

1) At first, *X* (input) data points are selected, picked or generated from the dataset at random.
2) Among the '*X*' features, calculate the decision nodes using the best split method.
3) The decision nodes are later split into daughter nodes to create a complex tree and repeat steps 1 & 2.
4) Each one of the decision nodes is utilized to predict the value of Y (output) from the *X* data points, and assign the new data point the average across all of the predicted values of Y.

To build a robust decision tree model, multiple data input variables are required to predict the output (*Y*). For the use-case, the z-axis raw values were transformed to feed the tree. Theoretically, the change in acceleration recorded, only in the vertical axis (z) should play a major role in determining the road anomalies, yet the *x* and *y*-axis data points were also utilized. The reason to incorporate both the axis was essentially done to avoid/ignore the anomalies created due to the rider's behavior. Sudden acceleration or deceleration in the direction of the vehicle (i.e x-axis) can contribute to some unexpected high and low () values. Also, if the peaks in the vertical axis are resulting from the abrupt movements in the (*y*-axis) direction i.e lane changing, it also needs to be considered in order to discard the unnecessary transients.

To further populate the decision tree, 82 features were added from the corresponding values from the z-axis (after the wavelet transform), y-axis and the x-axis. For example, this was done by introducing more variables by calculating the 'maximum', 'minimum', 'mean', 'standard deviation', 'variance', 'median absolute deviation' from the 0.05 percentile till the 0.95 percentile of values for all the axis. Similarly, the absolute values from the gyroscope sensor data were also introduced/experimented by calculating the above-mentioned variables so that the decision tree output prediction is as generalized as possible.

3.4 SEMI-SUPERVISED LEARNING APPROACH

A semi-supervised learning (SSL) approach is a technique that utilizes a typically small amount of labeled data with a large amount of labeled data. Labeled data acts as the orientation for data training and testing exercises. Labeling is a process of assigning meaningful tags and information about the data. SSL is particularly devoted to application domains in which unlabeled data are plentiful, such as image processing, information retrieval, and bioinformatics (Shabajee, Hannabuss, & Tilsed, 1998).

Distress classification

In this part, the pre-defined data-modeling technique described in Section 3.1 was utilized, comprising of a 7 step procedure to build a data-driven model. The first step is to define a use-case which is as follows:

(1) Use-case

Earlier in Section 3.3 the proposed model could only detect anomalies and correlate it with the IRI index which could only generalize the road profile into the good, bad or worse category. Although it is a useful technique to monitor the dense network to find initial signs on existing or upcoming distresses, it could not provide detailed information on the type of distresses that resulted in an anomaly.

An anomaly could result from multiple factors like (speed bumps, manhole, railroad crossing, road markings, potholes, cracks, raveling, etc.) but to precisely pick the resulting cause is also a challenging and an essential task for PZH. There are multiple distresses which can be visually detected and public authorities thus issue manual visual inspection plans to perform the job by dedicated inspectors.

To automate such a crucial task the author will focus on a different kind of data-type (*Imagery*) to classify pavement surface distresses utilizing a semi-supervised data-driven approach. Since the availability of labeled image data is scarce for classifying distresses, the author has performed a data collection procedure by utilizing a portable camera (GoPro 6) to collect images in bulk and quality as can be seen in Fig. 3.9. The vehicle utilized to perform data collection was provided by the PZH.

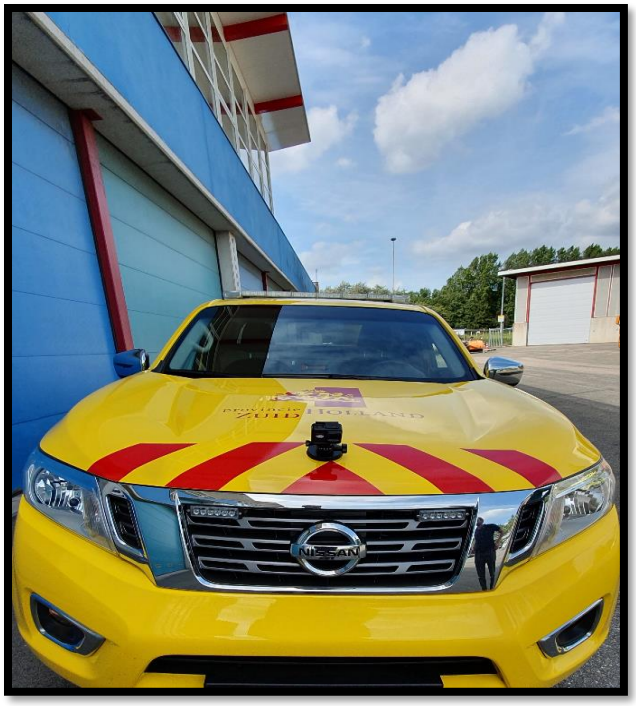


Fig. 3.9. Image Data collection

(2) Dataset collection & Exploration

An image, digital image, or still image is a binary representation of visual information. An image generally comprises of 3 channels (RGB) as seen in Fig 3.10. Each pixel coordinate contains 3 values ranging for intensities of 0 to 255 for an (8-bit) picture. Mixing different intensities of each color generates the full-spectrum of an image that is visualized by humans.

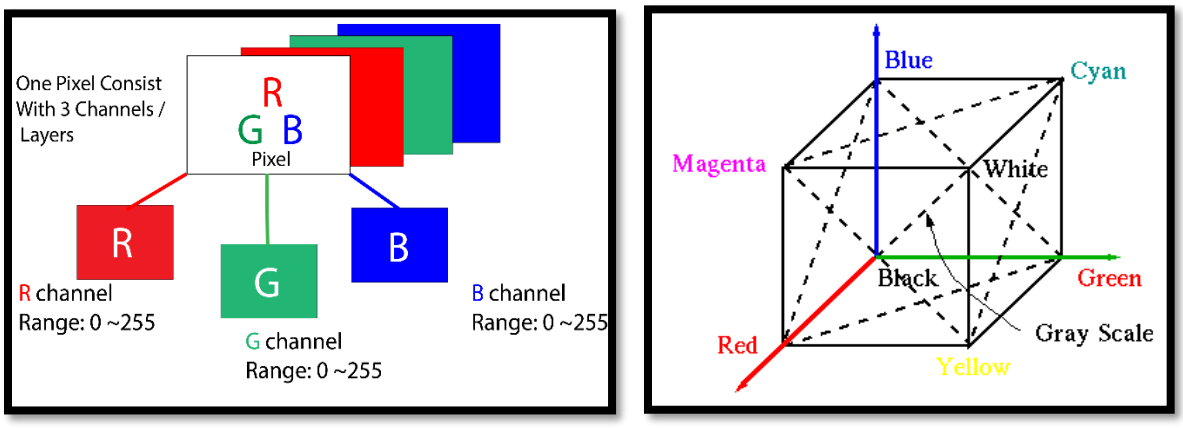


Fig. 3.10 Image channels and greyscale vector

An image f is a mapping from a spatial domain E to a range V, i.e. f:E→V and thus each element x∈E is mapped onto a value f(x)∈V, i.e. x∈E↦f(x)∈V. The domain of an image is the continuous Euclidean space (in principle we may position the image samples at all points in space): E=Rd. For a color image at every point, $x \in R^2$ three vales are measured: the red, green and blue intensity values. The image can be represented as a vector-valued function:

$$x \in R^2 \rightarrow f(x) = \begin{pmatrix} r(x) \\ g(x) \\ b(x) \end{pmatrix} \in R^3 \tag{3.5}$$

Human vision is a complex process, and mimicking the task has always remained a challenging process for computers. Through the use of traditional machine learning techniques and recent advancements in Deep Learning (DL), there had been significant progress being made in computers being able to interpret and react to what they "see".

Computer Vision (CV) is a scientific and interdisciplinary domain that concerns the theory and technology for building artificial systems that can obtain information from images or multi-dimensional data. CV follows a series of tasks to include methods for acquiring, processing, analyzing and understanding digital images for decision-making processes. The major tasks in the CV domain are image classification, image classification, and localization, object detection, object segmentation, etc.

Computer vision problems like image classification and object detection have traditionally been approached using hand-engineered features like SIFT (Lowe, 2004) and HoG (Dalal & Triggs, 2005). Representations based on the bag-of-visual-words descriptor (Yang, Jiang, Hauptmann, & C.-W. Ngo, 2007), in particular, has achieved remarkable success in image classification tasks. Image classification refers to the task of extracting information classes from a multiband raster image. The resulting raster from image classification can be used to create thematic maps. Depending on the interaction between the analyst and the computer during classification ("ArcGIS Help | ArcGIS Desktop," 2019).

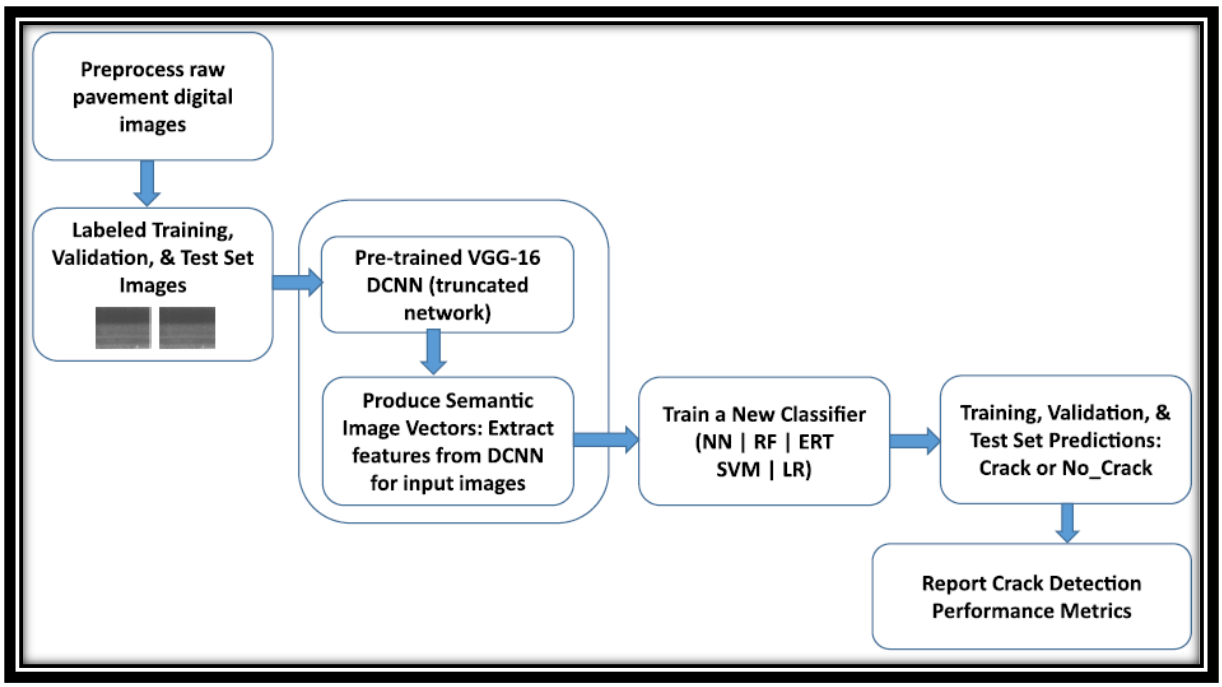


Fig 3.11 Pavement Distress Detection methodology proposed by (Gopalakrishnan et al., 2017)

Automated distress detection in pavements using imagery has remained one of the highest-priority research areas for transportation and pavement management systems. One of the proposed methodology in Fig. 3.11 is by (Gopalakrishnan et al., 2017) which have similar model deployment strategies (7 steps) as with this research but had only 2 types (crack and no-crack) of the labeled dataset to study from. This research is a step further, performing a multi-label classification task and with a larger dataset.

The data distribution (in%) per class can be seen in Fig. 3.12, comprising categorical data with 6 labeled classes: Crack, Potholes, Asphalt (without distress), shadow, manhole, and road markings. Each class of the dataset is manually labeled and segregated to perform the classification task.

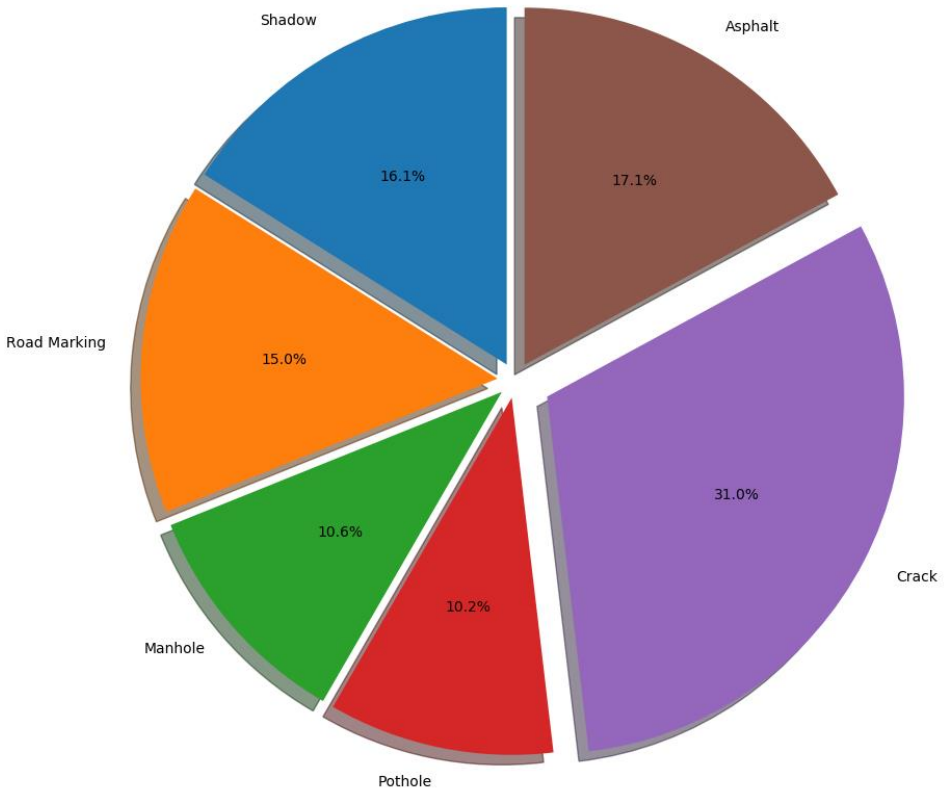


Fig. 3.12 Data distribution for distress classification.

The collected dataset comprises a sequence of 3 channels (RGB) images extracted from the mounted camera containing relevant and non-relevant (noisy) images for the distress classification task. A total of 15,750 images was then selected and segregated manually to perform the image classification task. The distribution per class in number is as below:

1) Crack: 4,889 images (31%)
2) Asphalt (without distress): 2,691 images (17.1 %)
3) Pothole: 1,608 images (10.2%)
4) Shadow: 2,530 images (16.1%)
5) Manhole: 1,663 images (10.6%)
6) Road Marking 2, 364 images (15.0%)

To maintain the consistency in the dataset, all multi-channel images were first pre-processed to 3 (RGB) channels and the author also maintained a dataset structure ("Kaggle: Your Home for Data Science," 2019) that is commonly used for conducting global competition throughout the globe.

(3) Algorithm

Convolutional Neural Networks (CNN) are now capable of outperforming humans on some CV tasks, such as classifying images. Despite all the attractive qualities of CNNs and the relative efficiency of their local architecture, they have still been prohibitively expensive to apply on a large-scale or for high-resolution images (Jansen & Zhang, 2007). Meanwhile, researchers from the machine learning community started working on developing denser models which typically consisted of multiple layers of non-linearity. This resulted in the early development of the first deep learning models (Srinivas et al., 2017) which showed promising results even from a 'small' dataset.

Deep learning's greed for large amounts of training data poses a challenge for image classification tasks, which can be alleviated by recycling knowledge from models trained on different tasks, in a scheme called transfer learning (Menegola et al., 2017). Multiple studies could achieve state-of-art results in image classification tasks based on transfer learning (Krizhevsky, Sutskever, & Hinton, 2012), (K Simonyan & Zisserman, 2014), (He, Zhang, Ren, & J, 2016). The aim of transfer learning (TL) is to improve the performance of a target solution with the help of already existing knowledge.

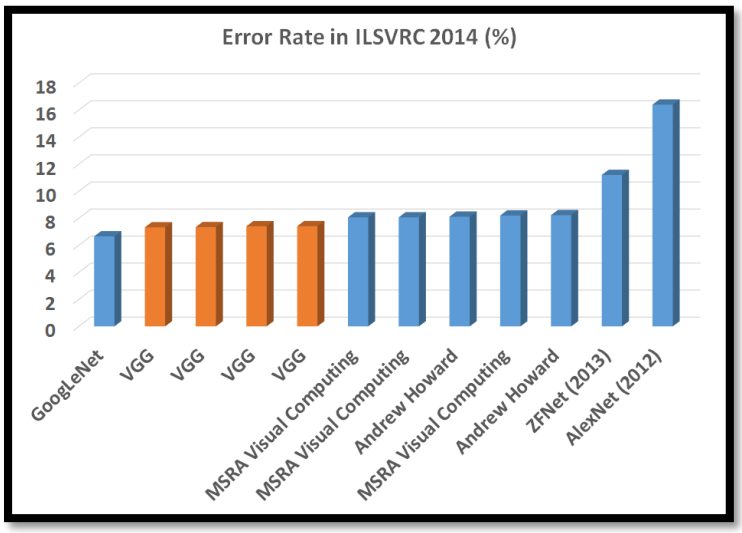


Fig. 3.13 Lowest Error rate in VGG networks for ILSVRC 2014 challenge

A common and highly effective approach towards deep learning on an image dataset is to leverage a pre-trained network. A pre-trained deep neural network developed and trained by Oxford's renowned Visual Geometry Group ("Visual Geometry Group - University of Oxford," 2014) achieved the highest accuracy and the lowest error rate

in the Imagenet Large Scale Visual Recognition Challenge in 2014 (ILSVRC2014) as can be seen In Fig. 3.13. ImageNet which is an on-going research project ("ImageNet," 2019), is organized according to the WorldNet hierarchy. Based on the statistics, there are 14,197,122 (~14 million images) in the dataset comprising of 21,000 categories (*synsets)* and a little more than 1 million images containing the bounding box annotations i.e image localization. The project is developed and maintained (by Princeton, Stanford, and other American universities) to benefit research communities by providing large-scale image datasets to overcome the challenges of Computer Vision like Image Classification and further. The dataset is been trained on multiple networks like: GoogleNet, VGG16, VGG19, Alexnet, Inception (v1, v2, v3), ResNet, Densenet, etc.

The research will utilize a deep neural network VGG16 which is trained on the ImageNet dataset by the process called Transfer Learning to handle the multi-label classification task for pavement surface distress classification.

(4) Feature extraction

Since it's a semi-supervised approach, a complete description of the images per category isn't required by the model rather just clustering the dataset and just associating the correct labels of the images should suffice. Before the pre-trained model can be fed with the dataset, the images need to be resized to fit the requirements of the neural network. The average size distribution of pixel vs density graph in Fig 3.14 suggests that the entire dataset comprises a vide ranging pixels, starting from the shape of 280*280 till 350*350 (width*height).

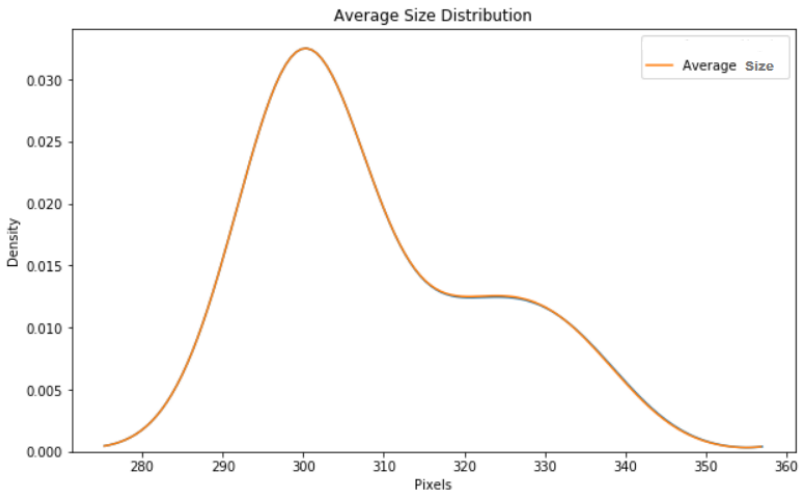


Fig. 3.14 Average size distribution of the Image dataset.

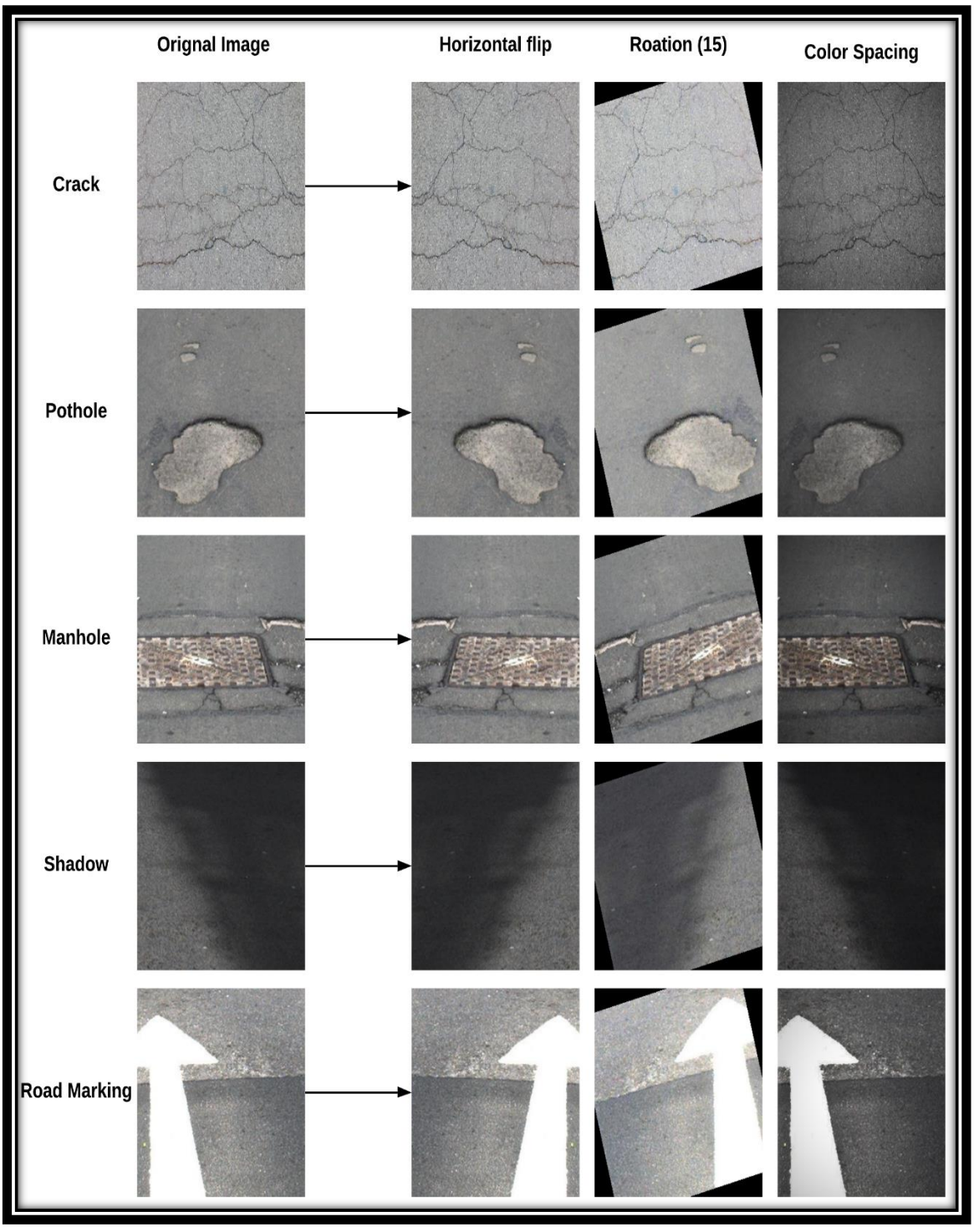


Fig. 3.15 Image Data augmentation

The generally accepted notion is that bigger datasets result in better Deep Learning models (Halevy, Norvig, & Pereira, 2009). However, creating an enormous dataset is a daunting task due to the manual efforts required for collecting and labeling the

underlying data. Thus data augmentation techniques are utilized to create a larger set of available data by performing the basic image manipulations techniques like geometric transformations, flipping, color space, cropping, rotation, translation, Gaussian noise injection, etc. (Chilamkurthy et al., 2018).

Each image from the dataset was first resized to the shape of 256*256 then each reshaped picture was randomly rotated 15 degrees around the horizontal axis, later the color space (brightness) of the images were altered to encounter the lightning biases. A random horizontal flip was introduced and then a center cropping function was utilized for converting the shape of the images to 224*224 (required by the VGG16 network). Every image was then normalized (between 0 to 1) for each channel, by subtracting the mean value and divide it by the standard deviation. The few data augmentation technique can be visualized in fig. 3.15.

(5) Model Development

PyTorch is an open-source library designed to enable rapid research in the domain of machine learning (Adam Paszke et al., 2017), which consists of several pre-trained models that have been trained on ImageNet dataset (Karen Simonyan & Zisserman, 2014). The approach for using a pre-trained model is well established and as follows:

1) At first, pre-trained model weights are loaded from a network that is trained on a large dataset.
2) The second step is to freeze all the weights in the lower (convolutional) layers depending on the similarity of the task to the size of the dataset.
3) The third step is to construct and replace the fully connected layers part of the network with a custom classifier, where the number of outputs must be set equal to the number of classes.
4) Lastly, only the custom classifier (fully connected) layers for the task are trained.

The pre-trained network (VGG16) (Karen Simonyan & Zisserman, 2014) with the pre-trained weights from (ImageNet) (Krizhevsky et al., 2012) serves as the first part of the neural network which extracts features that are relevant for many image recognition tasks. Pre-trained networks have proven to be reasonably successful for a variety of tasks, and result in a significant reduction in training time and usually increases in performance (Gu et al., 2015) (Sze et al., 2017).

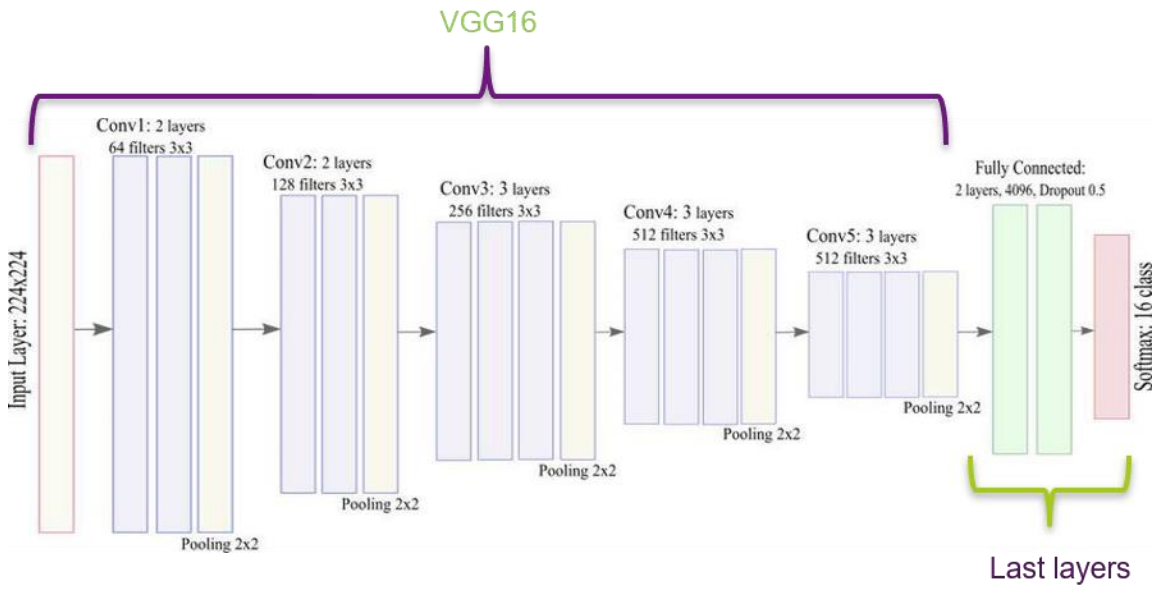


Fig. 3.16 Building the custom classifier (example)

The fully-connected convolutional neural network added in the last layers of the DNN which will be trained for the task consists of following layers (after multiple combinations and testings):

(1): Linear(in_features=4096, out_features=256, bias=True)
(2): ReLU()
(3): Dropout (p=0.2), inplace=False)
(4): Linear(in_features=256, out_features=6, bias=True)
(5): LogSoftmax()

To build the custom image classifier a *sequential()* module from the PyTorch library was utilized to specify each layer one after the other. The calculation of gradient is set to true by default for the last layers since only the last part of the network needs to be trained. The newly extracted features of the (use-case) dataset then can be trained on the new classifier.

3.5 SUPERVISED LEARNING APPROACH

For supervised-learning algorithms, there is a relationship of causality predefined by the user or modeler that is specified before providing the variables to the ML models (Ticlavilca Torres, Alfonso, 2008). Supervised learning is based on training a data sample from a data source with the correct classification already assigned (Sathya & Abraham, 2013). Most regression techniques are a supervised-learning approach where the researcher or the user defines the variables (outputs-inputs). Classification tasks also utilize this approach given that the casualty among the variables is pre-defined by the user (type and number of labels or groups).

Distress segmentation

The section will utilize the pre-defined 7 step approach (as studied in Section 3.1) to build a data-driven model using a supervised machine learning approach for pavement surface distress segmentation:

1) Use-case

If road damages are detected and classified, the PMS further needs to diagnose the severity of the classified distresses as objectively as possible. In contrast, damages are assessed by road inspectors in a completely subjective manner during the global visual inspection. PMS invests considerable resources towards monitoring the quality of roads due to the necessity of crack detection, to prevent water penetrating through the cracks which further leads to structural or service degradation of the road (KICT, 2017).

To tackle the problem at hand, the author will utilize one of the challenging domains from the list of CV tasks i.e. semantic segmentation. Until now the convolutional neural networks are typically used for classification tasks, where the output to an image is a single label class. However, in semantic segmentation, the desired output should include localization, i.e., a class label needs to be assigned for every pixel of an image as can be seen in Fig. 3.17.

Fig. 3.17 Semantic segmentation example (L. C. Chen, Zhu, Papandreou, Schroff, & Adam, 2018)

The progression of the CV task is to progress from coarse to fine inference. The starting point is a classification, which consists of making predictions for a whole input. Object localization/detection serves as a next step which provides not only the classes but also additional information regarding the spatial location of the corresponding classes. Finally, semantic segmentation contributes by achieving fine-grained level classification to extract pixel-level features of the desired output. Earlier computer vision problems only found elements such as edges (lines and curves) or gradients, but they never quite provided an understanding of images at a pixel level, in the way a human perceives it ("Semantic Segmentation," 2018). The applications of pixel-wise segmentation tasks extend to further domains like autonomous driving (Kaymak & Uçar, 2017), facial recognition (Saito, Li, & Li, 2016), biomedical image segmentation (Menegola et al., 2017), precision agriculture (Milioto, Lottes, & Stachniss, 2018), remote sensing (Ardeshir, Collins-Sibley, & Shah, 2015), etc.

2) Dataset

Constructing a large dataset of annotated images with the desired objects is a costly and lengthy endeavor. Image Annotation is the process of building datasets for computer vision models. This enables machines to learn how to automatically assign metadata into a digital image using captioning or keywords. According to (Deligiannidis & Arabnia, 2014), image annotation can be categorized into three types:

I. Retrieval-based
II. Classification-based
III. Probabilistic based

The basic notion behind retrieval-based annotation is that sematic-relevant images are composed of similar visual features. Traditional techniques like content-based image retrieval (CBIR) has been proposed by (Kato, 1992). CBIR is applied by the use of image features including shape, color, and texture.

A classification-based approach on the other side is completely a supervised approach where annotations are treated as classification using multiple classifiers. Images are classified based on the extracted features where each semantic concept has an independent class.

The last type intends to solve the problem of 'synonym' and 'homograph' by estimating the correlations between images and concepts while emphasizing the term-term relationship. Notwithstanding the efforts made on the enhancement of annotation quality, the aforementioned approaches tend to assign the same annotation to all the images in the same cluster thereafter they suffered the lack of customized annotation for each image (Huang, 2014).

As the use-case is also a supervised approach towards segmenting pavement surface distresses from 2D images, the second type (classification-based) approach will be utilized for the image annotation technique. Conventionally, datasets are built by a single research group and are tailored to solve a specific problem. To achieve this, a web-based annotation tool LabelMe (Russell, Torralba, Murphy, & Freeman, 2008) was utilized for building up the synthetic dataset.

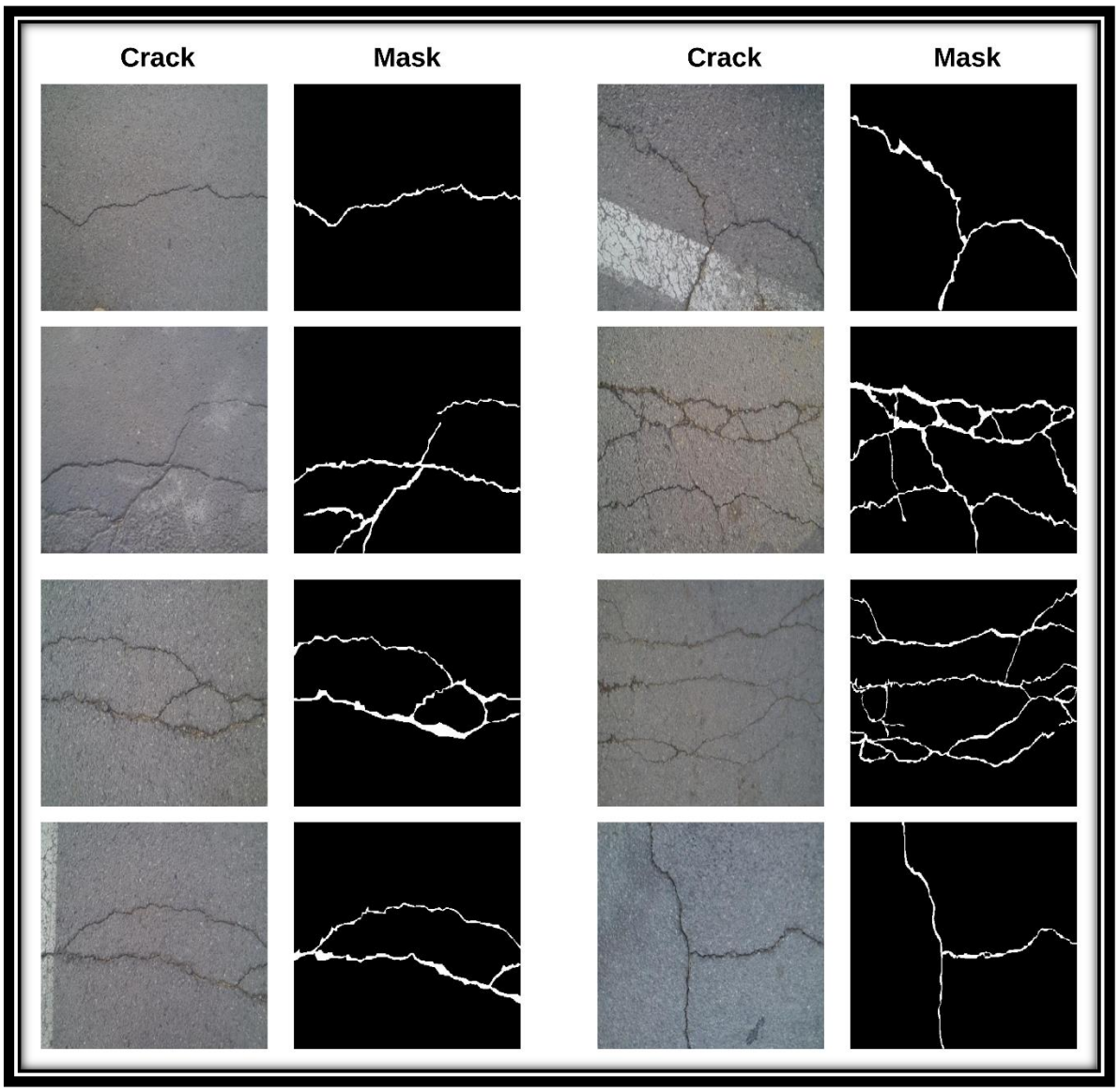


Fig. 3.18 Crack segmentation dataset (sample)

The prepared dataset (as seen in Fig. 3.18) comprises of a single class pavement surface distress, i.e. crack. As mentioned in Section 3.5.1 it is essential for PMS to detect and seal cracks as early as possible. Also, the most commonly found surface distress in the PZH network was cracking, thus the database was pivoted towards one type of class.

After, several hours of manually annotating the image dataset a total of 22,996 images were prepared to feed the model. Further demarcation of the dataset is below:

i. RGB images of crack (11,498)
ii. Binary masks of segmented cracks (11,498)

iii. Data shape (320*320*3), width*height*channel
iv. Total no of pixels 7,064,371,200 (~7 billion)

3) Algorithm

Unlike in a multi-label classification (as in Section 3.3) where the output of the model is a single-digit value (label) for the specified class, a segmentation model re-generates a mask from the given input image as can be seen in Fig. 3.19.

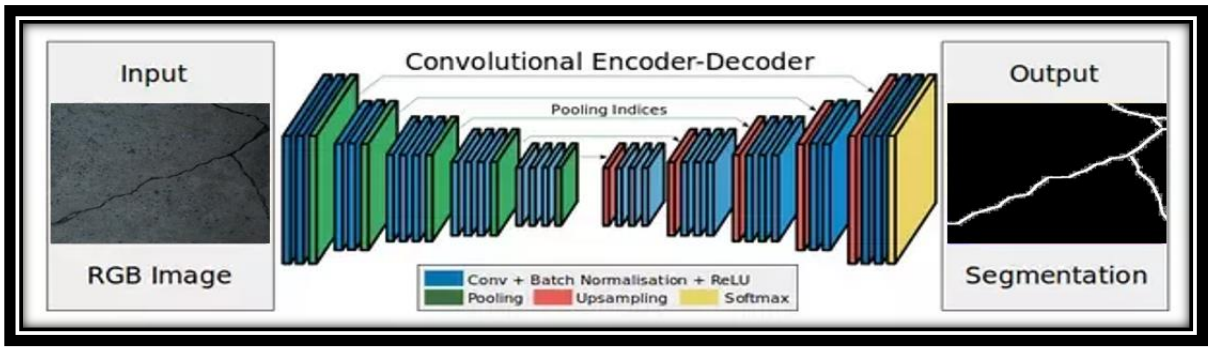


Fig. 3.19 Encoder-Decoder architecture example

Several models based on Fully Convolutional Networks (FCNs) (Sermanet et al., 2014) (Long, Shelhamer, & Darrell, 2015) can exploit the contextual information for segmentation tasks. The Encoder-Decoder (EnDec) architecture has shown tremendous success in multiple CV tasks, including object detection (T.Y et al., 2017), human pose estimation (Newell, Yang, & Deng, 2016), and even semantic segmentation (Noh, Hong, & Han, 2015) (G. Lin, Milan, Shen, & Reid, 2017).

Typically an EnDec architecture contains (1) an encoder network module that gradually reduces the feature maps and captures higher semantic information, and (2) a decoder module that gradually recovers the spatial information (L. C. Chen et al., 2018). For example, a CNN, RNN, etc., can serve as an encoder (contraction path) that takes the input and output a feature map/vector/tensor. These feature vector hold the information, the features, that represents the input. The decoder (expansion path) is again a network (usually the same network structure as encoder but in opposite orientation) that takes the feature vector from the encoder and gives the best closest match to the actual input or intended output.

4) Feature extraction

It is common knowledge that more the data an ML algorithm has access to, the more effective it becomes. Data augmentation (as in Section 3.3.4) techniques were utilized to create 'new' data and reduce the chances of overfitting on the segmentation model. A very generic and accepted current practice for augmenting image data is to perform geometric and color augmentations, such as reflecting the image, cropping and translating the image, and changing the color palette of the image. All of the transformations are an affine transformation of the original image that takes the form (J. Wang & Perez, 2017):

$$y = Wx + b \tag{3.6}$$

As suggested above, data augmentation fulfills a necessary part for training any neural network as it increases the size and generalizes the dataset, i.e prevents model overfitting. Although it can be strategically applied at two different intervals. First, is to perform all the necessary transformations beforehand, which in turn increases the size of the given dataset. The second interval is to apply the transformations on mini-batches, just before feeding it to the model. The former method is known as offline augmentation and is preferred if the size of the dataset is relatively small. Latter approach is utilized when the dataset is large enough, preventing the sudden explosion of the dataset in size by a factor equal to the number of transformations performed.
Transformations were applied by using a Python library name *albumentations* ("albumentations · PyPI," 2019), which supports a diverse set of transformations that can be applied to images, masks, key points, and bounding boxes. Spatial-level transforms can simultaneously change both an input image as well as additional targets such as masks.

Since the dataset is large enough, a small number of augmentations techniques were required and applied:

- Flip: flips the input either horizontally, vertically or both. Both flips were utilized.
- Affine: places a grid of points on the input images and randomly move the neighborhood of the point around the applied affine transformation.
- Perspective: perform a random four-point perspective transform of the input.
- Brightness, contrast and color manipulations
- Image blurring and sharpening

- Gaussian noise.
- Random crops

5) Model deployment

In order to build an EnDec architecture, the network requires a contraction path (Encoder) and an expansion (Decoder) path to first converge the features and return output in the form pixelated object class. The fully connected CNN architecture was built by two sets of the network on each side and are explained as follows:

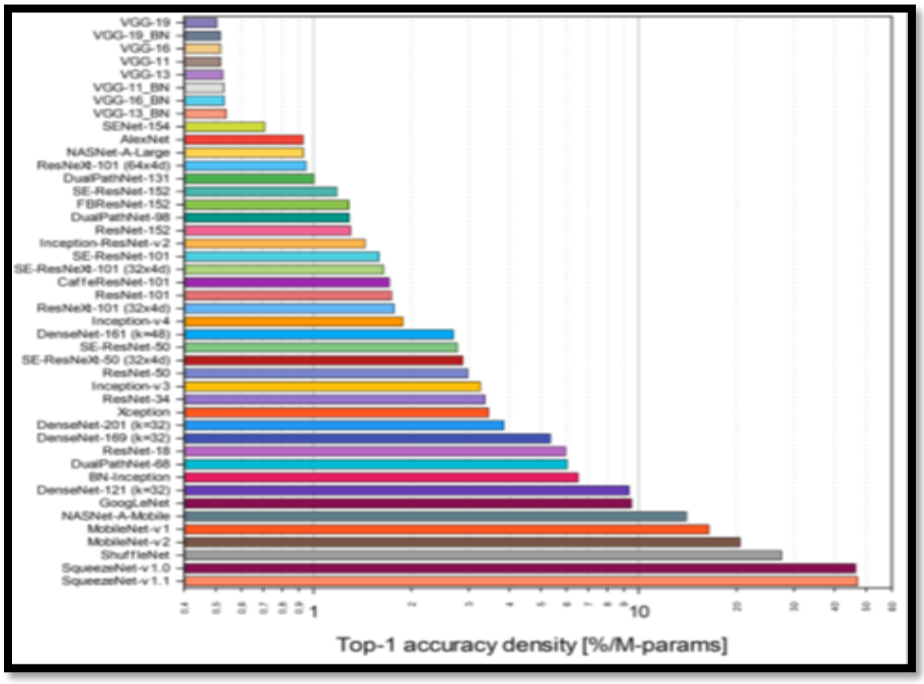


Fig 3.20. Top-1 accuracy density (%M-params) (Bianco, Cadene, Celona, & Napoletano, 2018)

Encoder network:

The first part plays the role of the encoder path which processes the image for extracting the features. Compare to the dataset of ImageNet, the use-cased dataset is still small. Thus, for converging the input images faster Transfer Learning techniques were again utilized to suffice data greed and accelerate convergence thus improving the model performance (Zamir et al., 2018). Similar studies in domains like construction equipment detection (Kim, H., H., Y.W., & Byun, 2018), pavement crack detection (Gopalakrishnan et al., 2017) (Zhang et al., 2018) (A. Zhang, Wang, C., et al., 2018), have shown promising results even with 'less' training data. The top-1 accuracy density of the encoder network is shown in Fig. 3.20.

List of deep convolutional neural architectures available as a backbone for the crack segmentation task is as shown in Table 3.1 below:

No.	Weights	Model	Acc@1	Acc@5	Source
1	Light Weight (LW)	**vgg16**	70.79	89.74	keras
2		vgg19	70.89	89.69	keras
3		resnet18	68.24	88.49	mxnet
4		resnet34	72.17	90.74	mxnet
5		resnet50	74.81	92.38	mxnet
6		resnet101	76.58	93.1	mxnet
7		resnet152	76.66	93.08	mxnet
8		resnet50v2	69.73	89.31	keras
9		resnet101v2	71.93	90.41	keras
10		resnet152v2	72.29	90.61	keras
11	Medium Weight (MW)	resnext50	77.36	93.48	keras
12		resnext101	78.48	94	keras
13		densenet121	74.67	92.04	keras
14		densenet169	75.85	92.93	keras
15		**densenet201**	77.13	93.43	keras
16		inceptionv3	77.55	93.48	keras
17		xception	78.87	94.2	keras
18	Heavy Weight (HW)	**inceptionresnetv2**	80.03	94.89	keras
19		seresnet18	69.41	88.84	pytorch
20		seresnet34	72.6	90.91	pytorch
21		seresnet50	76.44	93.02	pytorch
22		seresnet101	77.92	94	pytorch
23		seresnet152	78.34	94.08	pytorch
24		seresnext50	78.74	94.3	pytorch
25		seresnext101	79.88	94.87	pytorch
26		senet154	81.06	95.24	pytorch
27		nasnetlarge	82.12	95.72	keras
28		nasnetmobile	74.04	91.54	keras
29		mobilenet	70.36	89.39	keras
30		mobilenetv2	71.63	90.35	keras

Table 3.1 Available backbone architectures as contraction path encoder

One backbone from each category (based on the diversity, popularity and weight size of each architecture) was picked to serve as a contraction path encoder for the segmentation model:

1) Light Weight (LW): VGG16
2) Medium Weight (MW): densenet201
3) Heavy Weight (HW): inceptionresnetv2

Decoder Network:

In place for the expansion path where the function of the network is to enlarge the size of the features-maps to match it with the input image, two types of architecture were utilized:

1. **U-net**

   Generally in a semantic segmentation problem, the objective is to partition an image into a set of non-overlapping regions, which allows the homogeneous pixels to be clustered together (McGuinness & O'connor, 2010). Perhaps, the most ingenious aspect of the U-Net architecture is the introduction of skip connections (Ibtehaz & Rahman, 2019). These skip connections allow the network to retrieve the spatial information lost by pooling operations (Drozdzal, Vorontsov, Chartrand, Kadoury, & Pal, 2016). Fig. 3.21 displays a U-net architecture example:

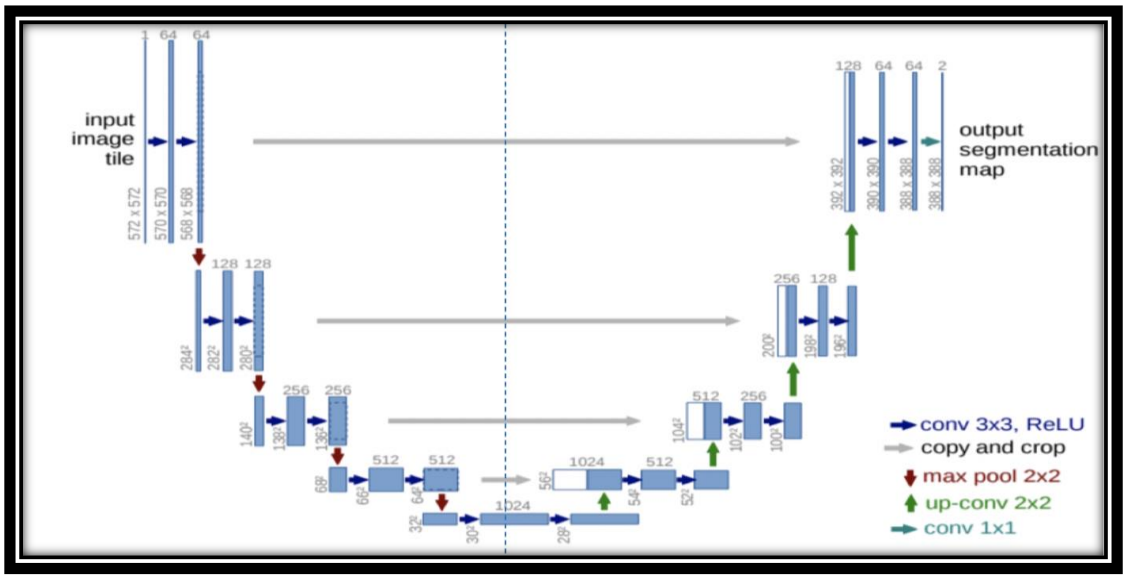


Fig.3.21 U-net architecture example ("U-Net: Image Segmentation Network," 2018)

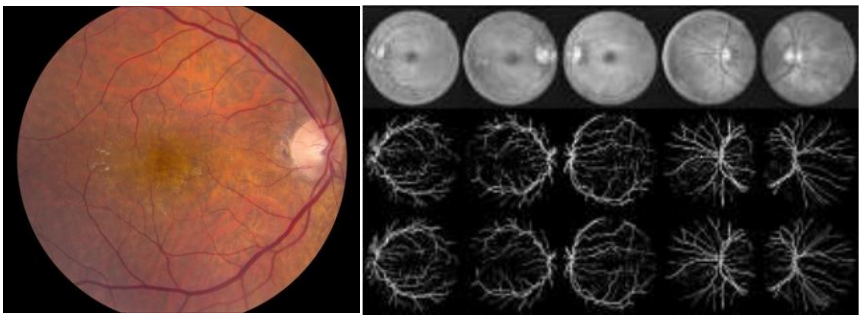
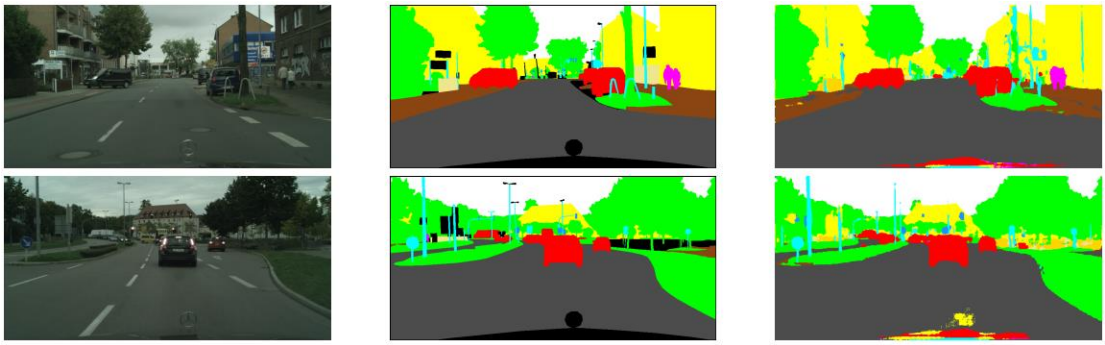


Fig. 3.22 Retina blood vessel segmentation using U-net (Xiancheng et al., 2018)

U-net convolutional network has also been utilized for International Symposium on Biomedical Imaging (ISBI) challenges, where it outperformed the prior best method (a sliding-window convolutional network) for segmentation of neuronal structures in electron microscopic stacks (Ronneberger, Fischer, & Brox, 2015) and retina blood vessel as seen in Fig. 3.22 (Xiancheng et al., 2018). Based on the model performance on tasks like segmenting cellular wall membranes and blood vessels in the human retina, it is ought to perform well on similar features characteristics from the use-case dataset.

2. LinkNet



(a) Input Image (b) Ground truth (c) Prediction

Fig. 3.23 LinkNet prediction on Cityscapes test set (Chaurasia & Culurciello, 2017)

Once the features are extracted from the input image by the encoder network, it is mapped back to the original image size by the decoder for dense prediction. The motivation for intensive predictions on the input image class is to ensure the speed without the loss of accuracy. LinkNet contains fewer parameters than the heavyweight architectures and the whole network are more efficient, and even real-time operation can be realized (Wulamu, Shi,

Zhang, & He, 2019). The decoder architecture comprises of four decoder blocks, each containing two convolutions and a full convolution layer. Followed by batch normalization in each CNN layer with additional Rectified Linear Unit (ReLU) (Fred Agarap, 2019) and two up-sampling layers. Various model performance had also been tested on the Cityscapes (Cordts et al., 2016) dataset (as seen in Fig. 3.23) against the popular SegNet (Badrinarayanan, Kendall, & Cipolla, 2015), ENet (A. Paszke, Chaurasia, Kim, & Culurciello, 2016), Dilation8/10 (Yu & Koltun, 2015), and Deep-Lab CRF (L.-C. Chen et al., 2016), out of which LinkNet performed the best (Chaurasia & Culurciello, 2017) in overall performance.

With three types of encoder architecture and two types of decoder architecture, the model can now be iterated on 6 different combinations with the use-case dataset for pixel-level segmentation of pavement surface cracks.

3.6 SUMMARY

Through all the techniques specified above, each approach has its own dependency upon the amount of labeled data the use-case has. Firstly, if the data is available in a huge amount and it is unlabeled the best approach would be the unsupervised learning method. This method is useful to provide useful insight when the data is ambiguous and unlabeled. Since there is no *known* output from the input values, the accuracy of the models tends to be lower than other approaches. Secondly, if the input data has a brief categorical or class information and the desired output is to filter or sort the unlabeled data, semi-supervised approaches work well. Lastly, the supervised learning technique scores high in accuracy and precision values but requires massive and detailed information about the input-output dataset.

Chapter 4: Iterations and Results

This chapter includes the last two steps of the flow diagram for building an ML model (as seen in Section 3.1). The iteration process to fine-tune the model for achieving better performance and results for each use-case is presented in Section 4.1 (accelerometer data), Section 4.2 (distress classification) and Section 4.3 (distress segmentation).

4.1 ACCELEROMETER DATA

Index	acc_data_x	acc_data_y	acc_data_z	accuracy	date	lat	locationtimestamp	lon	metadata
0	-0.015714	-0.011703249	-0.005272222	3.686	12:03.5	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
1	0.011795682	-0.022867623	-0.023249876	3.686	12:03.5	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
2	-0.00449002	-0.000691636	0.00151067	3.686	12:03.5	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
3	0.024965255	-0.038734667	0.008986248	3.686	12:03.6	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
4	0.059898184	-0.018520138	-0.01194505	3.686	12:03.6	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
5	0.038828201	0.025258577	0.01946399	3.686	12:03.7	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
6	0.061498041	0.01708107	-0.013957155	3.686	12:03.7	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
7	0.004614671	0.017390205	0.013577627	3.686	12:03.8	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
8	0.052325452	0.029235405	-0.011483541	3.686	12:03.8	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
9	-0.002614828	-0.078822247	-0.028078642	3.686	12:03.9	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
10	0.014700213	-0.005832565	-0.004613102	3.686	12:04.0	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'
11	-0.004286688	0.032431329	-0.031619181	3.686	12:04.0	52.16	1.56897E+12	4.496	{'deviceType': 'LG-H870'

Table 4.1 Sample Accelerometer Data template

Iterations

Table 4.1 represents a sample data template produced by the accelerometer sensor, comprising the 'accuracy', longitude', 'latitude', 'date', 'location timestamp' 'metadata' and many more. All the 82 variables were collected for the 125,325 data points in the entire road network. All this dataset was further split into two segments bu utilizing the commonly know Pareto principle (80/20 rule). The *training* dataset (80%) is used to fit the model and the *test* dataset (20%) is rather utilized for assessing the performance of the model. The training dataset was randomly shuffled and iterated over 40 times to populate the decision tree. To gauge the estimation of the model and assess the quality of the predictor a Mean Squared Error (MSE) function was utilized to create an unbiased estimate of error variance i.e. the residual sum of squares divided by the number of data points or also known as the degree of freedom. Thus the prediction error can be computed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{4.1}$$

where:

$$\text{MSE is the mean } \left(\frac{1}{n}\sum_{i=1}^{n}\right) \text{ of the squares of the errors } \left(Y_i - \hat{Y}_i\right)^2 \qquad (4.2)$$

Results

In an unsupervised classification task, like the one above, multiple input variables (X) are allocated with an unbiased *weight* factor by the model to generate a complex relationship between the input variables and the desired output (*Y*). In the pre-defined use-case (Section 3.3.1), the task at hand was to detect road anomalies by using accelerometer data and assign a relationship in the detected outliers as such, that the model could produce values as close as the governing road performance indices like International Roughness Index (IRI).

Table 4.2 Predicted IRI from accelerometer data

Aran ID	IRI	Predicted IRI	Location
1840	2.865	2.187125	POINT (91180.45 464554.03)
1841	2.545	2.300875	POINT (91178.46000000001 464554.28)
1842	2.255	2.138531	POINT (91168.52 464555.41)
1850	0.925	1.455437	POINT (91091.21000000001 464570.6)
1851	0.77	1.307313	POINT (91081.69 464573.81)
1852	0.985	1.083817	POINT (91072.12 464577.02)
1853	1.455	1.57025	POINT (91062.67999999999 464580.23)
1858	1.92	1.438146	POINT (91009.37 464597.01)
1859	1.55	1.438146	POINT (91005.31 464598.34)

For training the model and for validation purposes also, the ARAN 2018 database was utilized as a standard. Table 4.2 represents a portion of the predicted output from the model. Fig 4.1 is plotted to further visualize the predicted IRI over the actual IRI (x-axis) along with data points (on the y-axis) for all the three road networks:

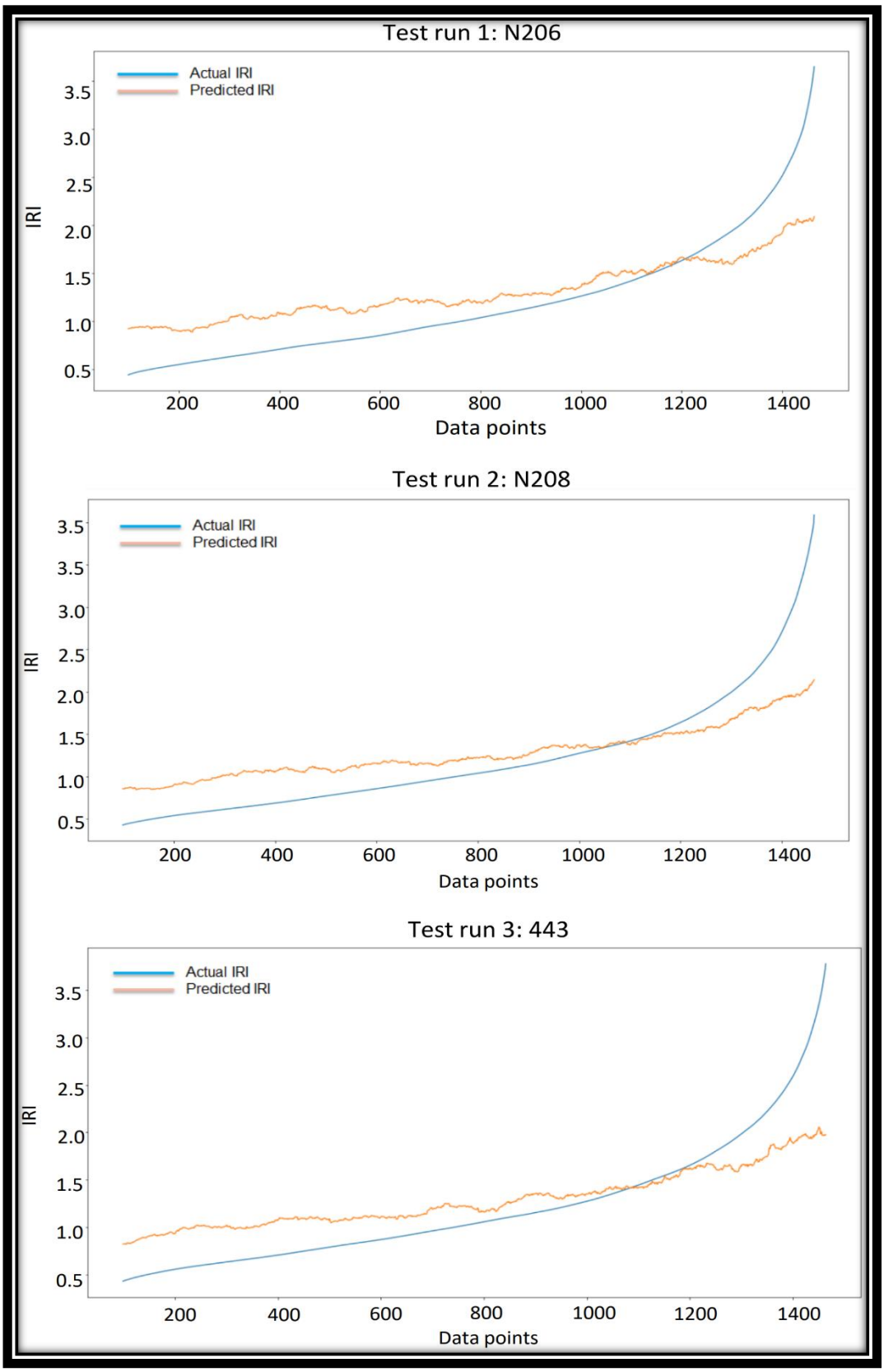


Fig. 4.1 Actual IRI vs Predicted IRI for the three test runs

With the actual IRI from ARAN and the predicted IRI from the model, a comparison map can be plotted. Since each data-point was geotagged with its longitude and latitude a precise classification of the road network based on the IRI scale can be plotted. Fig. 4.2 displays a network map in which every data point was indicated *green* (0.1<IRI<2.5) for good roads, *orange* (2.6<IRI<4.5) for bad roads and *red* (4.6<IRI6.5) for the bad roads.

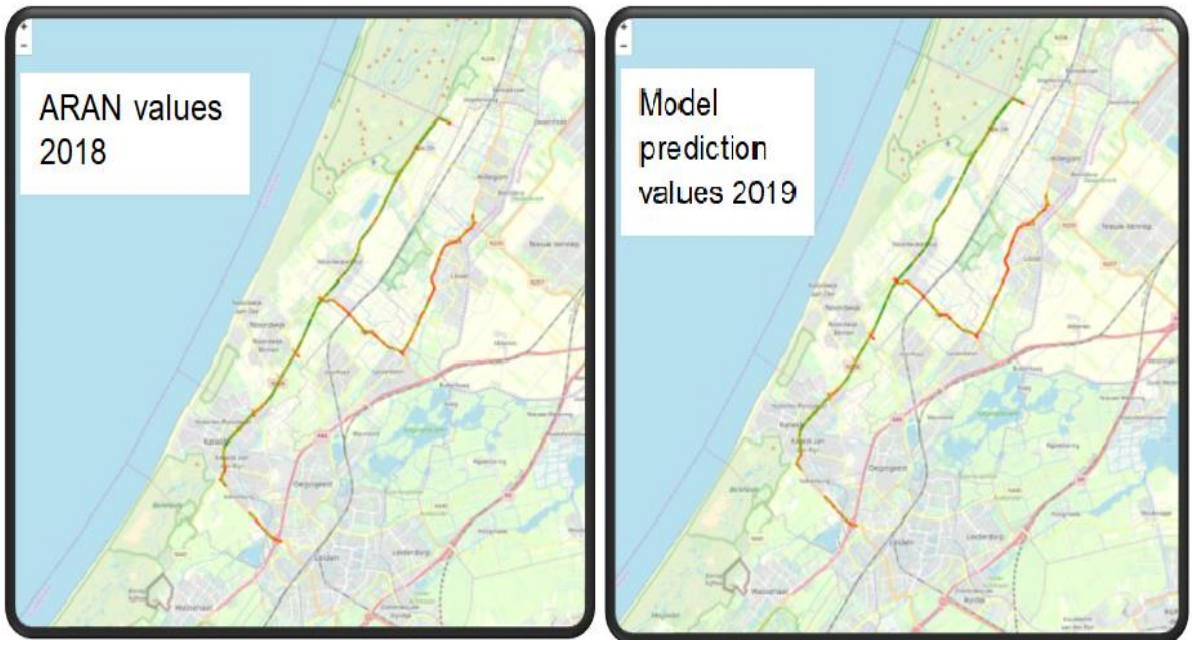


Fig. 4.2. Network map of ARAN values vs Predicted values of IRI

As seen in Fig 4.2, the predicted values of IRI from the inexpensive ML model are producing similar results as the affluent ARAN vehicle. Since it was an unsupervised task, the model's accuracy can only be evaluated only by comparing the average dispersion of IRI values. The average error per data point in just 3 test runs were recorded to be (+-0.5) values apart from the validation results.

In this section, a system was introduced that can detect road anomalies using smartphones equipped with inertial sensors like accelerometer and gyroscope. Classifying road anomalies is rather a difficult task and the expectancies to produce close to actual values of IRI on one pass are quite low. Thus, to achieve better results the test run was repeated 3 times to collect ample data points and later a wavelet transforms analysis was implemented to remove the effects of speed, slopes, drifts from the sensor signal. Moreover, the developed technique is useful irrespective of the orientation (horizontal/vertical) of the phone, the type of vehicle (gasoline/diesel) and

the smartphone type. The results can be further improvised if the training dataset could further incorporate the high IRI's (above 5.5) and the low IRI's (below 0.8). A conclusion and additional recommendations on the developed model will be summarized later in Section 5.2 and 5.3 respectively.

4.2 DISTRESS CLASSIFICATION

Iterations

The total number of parameters in the custom-built classifier is around 135,310,918 but only 1,050,374 needs to be trained from scratch. Because of the advantage, given by TL, the custom classifier can be trained on a local machine and in lesser time than to train its entire architecture from scratch.

Even with few layers to be trained, there are still over a million parameters (weights) that will be updated during training. In effect, the model is fine-tuned by performing some hyperparameter optimization to work on the custom problem (Feurer & Hutter, 2012). Just like in Section 4.1, the entire dataset is again split into two portions of the *training set* (80%) for fitting the model, and a *validation set* (20%) for assessing the performance of the custom-built classifier.

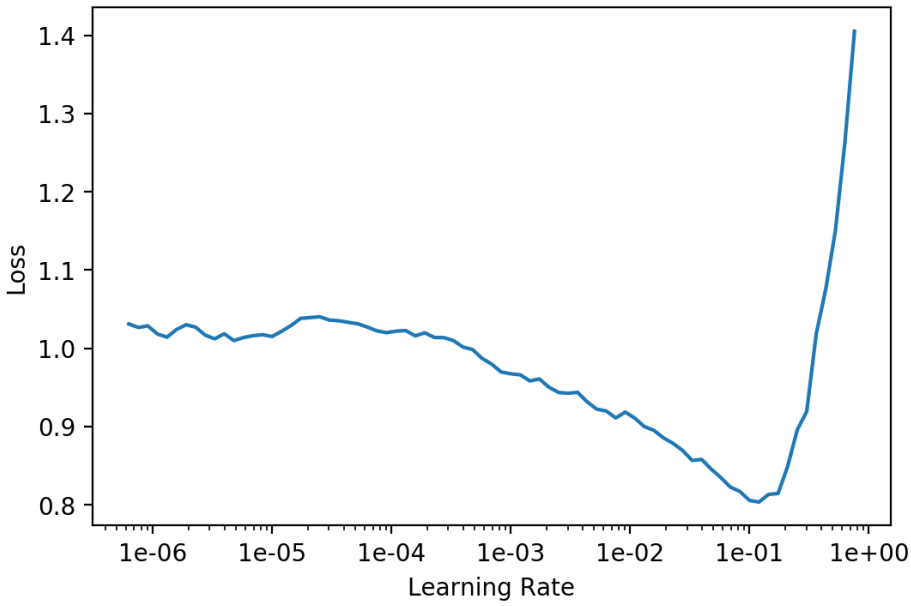


Fig. 4.3 Learning rate optimization

For training the model it is initially trained for a lesser number of iterations than it might actually require. Since the weights of the built deep neural network cannot be calculated using an analytical method. Instead, the weights must be discovered via an

empirical optimization procedure. A learning rate *(LR)* or step size schedule is a configurable hyperparameter used in the training of neural networks that is a positive scalar, often in the ranges from 0.0 to 1.0 (Ge, Kakade, Kidambi, & Netrapalli, 2019). Fig. 4.3 displays the model converging (empirical optimization) of the learning rate $1e^{-01}$, which fits the model accurately.

It is important to find an optimum value for the learning rate for the custom-built models since it also depends on the input features from the training dataset. Common learning rate schedules include constant decay, time-based decay, step-decay, and exponential decay. Since the use-case dataset is dissimilar to the pre-trained network which is trained on ImageNet, a time-based (1/t) Stochastic Gradient Descent (SGD) optimizer will be utilized (Cormode, Shkapenyuk, Srivastava, & Xu, 2015). The mathematical form is represented as:

$$\alpha = \frac{\alpha_0}{(1+kt)} \tag{4.3}$$

where:

α_0 and k are the hyperparameters and t is the iteration number.

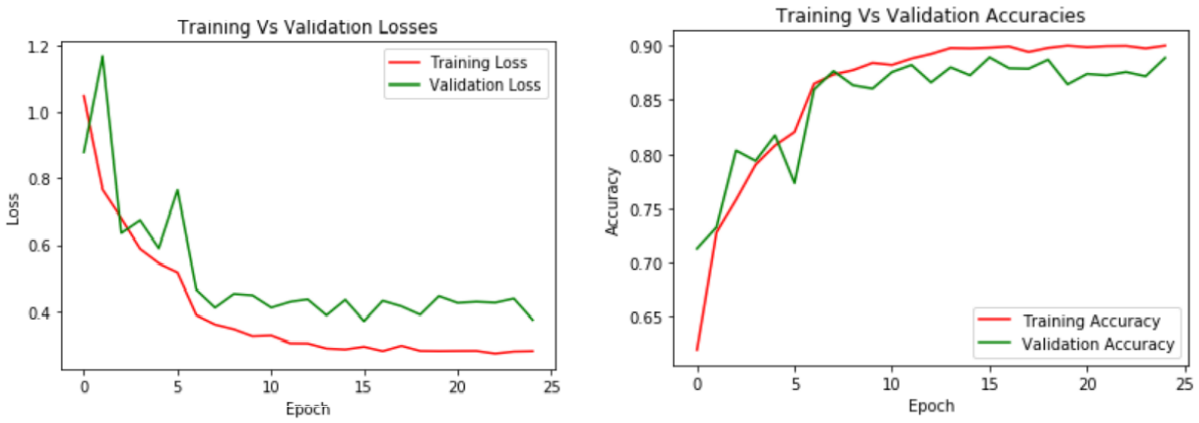


Fig. 4.4 Accuracy and Loss for Training and Validation dataset

After calculating the gradients, an optimizer function is called to update the model parameters with the gradients. This is performed to monitor the accuracy and error loss of the model on individual batches during the forward pass (as seen in Fig. 4.4). A validation dataset is prepared for fine-tuning the hyperparameters such that *early stopping* can be performed.

Early stopping halts the training when the validation loss has not decreased for a number of epochs. Each time the validation loss does decrease, the model weights are saved so that it can later be loaded for further iterations Early stopping is an effective method to prevent overfitting on the training data (Prechelt, 1997). If training continues, the training loss will continue to decrease, but the validation loss will increase because the model is starting to memorize the training data. Early stopping prevents this from happening, and, if the model is saved per epoch when the validation loss decreases, it is possible to retrieve the model that does best on the validation data.

Results

The model trained for 25 epochs (iterations) until the validation loss stopped plummeting. The overall time it took for training the model was near to 09 hours and 16 minutes on an NVIDIA GeForce 940M GPU. The best training accuracy achieved was at ~(89%) with ~(0.17) training loss at the end. Interestingly, the model performed also distinctively well on the validation accuracy ~(86%) with a slightly higher loss ~(0.43) than the training dataset.

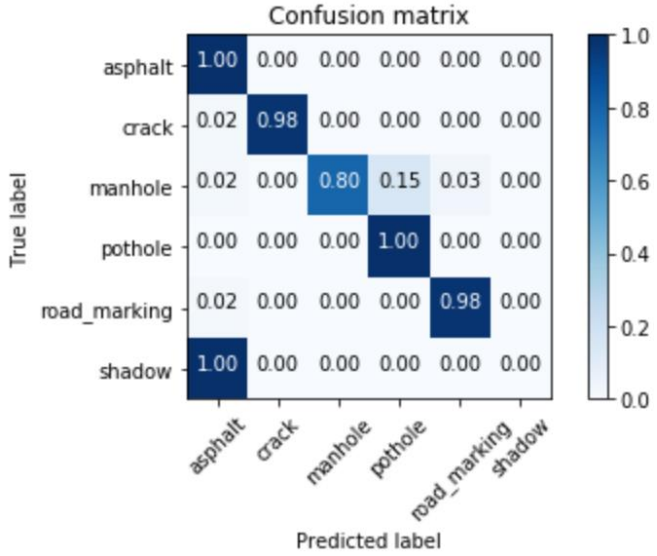


Fig. 4.5 Confusion Matrix per label class

As the use-case is of a multi-label classification task, the overall accuracy of the model doesn't justify how it performed per class. Thus to visualize the performance of the algorithm a confusion matrix is plotted (Fig. 4.5) against a test dataset for which the true values are known (Visa, Ramsay, Ralescu, & Van Der Knaap, 2011). The algorithm seems to perform quite satisfactory in all the classes but the *shadow*. Although the test dataset includes the images that the model has never encountered before, it still only

provides an overview of how the model will perform during production phase (Step 8, 9, 10 as studied in Section 3.1). Since *shadow* is not a relevant class for PMS to classify, thus no further fine-tuning or optimization will be performed to improve the accuracy per class. Below in Fig. 4.6 are the sample prediction results of the model per top 5 class:

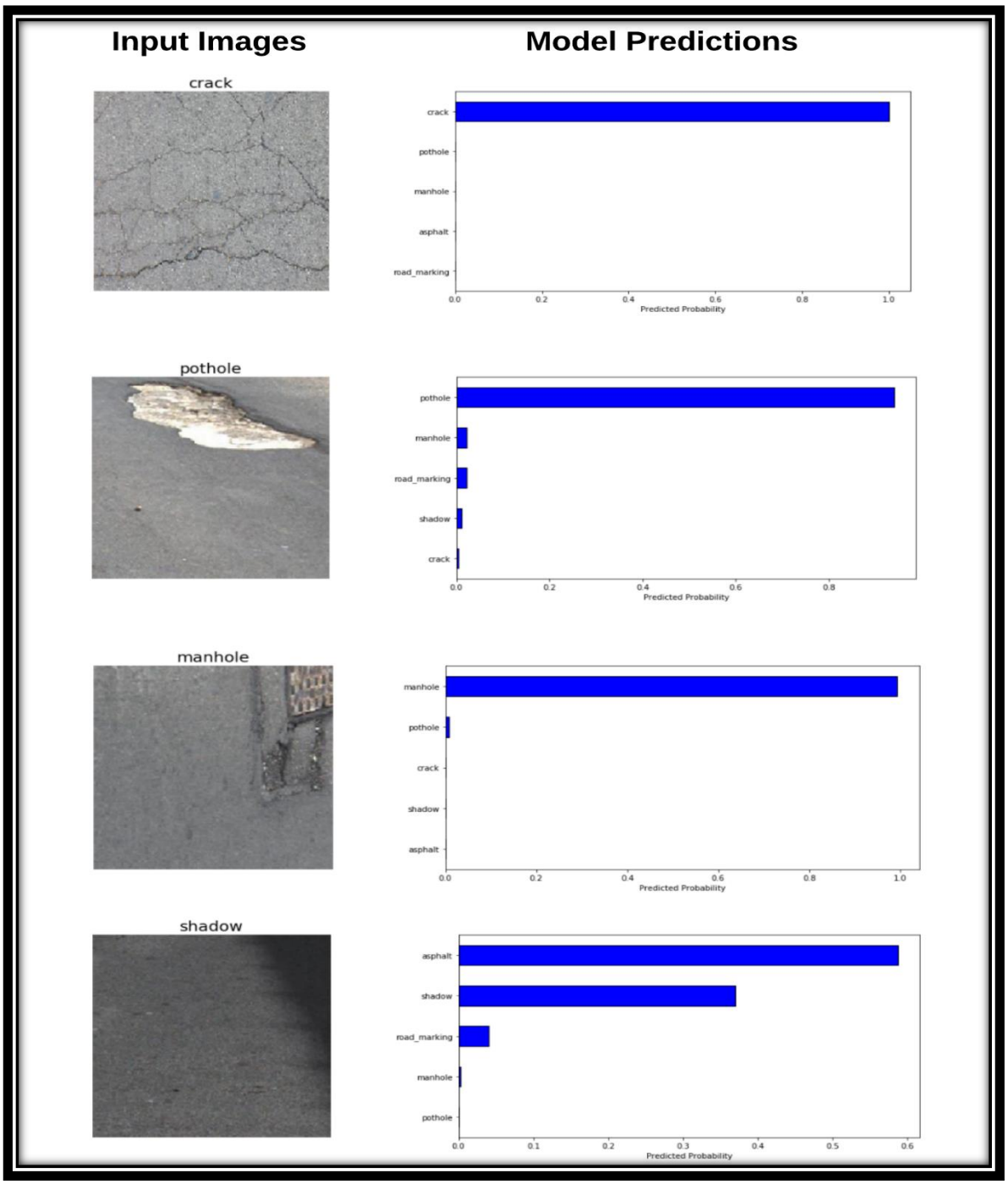


Fig. 4.6 Model Prediction (VGG16) for Image distress classification

The prediction results display a probability distribution graph with which the model predicts the value per class and classifies the input image. As can be seen in Fig 4.6 the class *shadow* is yet detected on the second-highest probability, but since the accuracy of the test dataset depends upon the first prediction, the model seems to perform poorly on the particular class.

Model	Weights	Architecture Backbone	Top Accuracy	Parameters	Base Model	Hardware	Learning Rate	Epoch	Time	Train Loss	Val Loss	Accuracy per Class	Training Accuracy	Validation Accuracy
Image Classification	LW	**GoogleNet**	0.779	**1.3 M**	Sequential	**NVIDIA GeForce 940M**	1.00E-03	25	**05:48:43**	0.263	0.458	1) 73% 2) 82% 3) 80% 4) 71% 5) 63% 6) 56%	71%	**69%**
	MW	**VGGnet**	0.713	**13.83 M**	Sequential	**NVIDIA GeForce 940M**	1.00E-03	25	**09:15:12**	0.177	0.439	1) 92% 2) 89% 3) 83% 4) 86% 5) 91% 6) 87%	89%	**86%**

Table 4.3 Image Classification Model performance

Similarly, another architecture named GoogleNet (Szegedy et al., 2015) which is a lightweight (LW) model (compared to VGG16) was also trained on the custom dataset to compare the accuracy of the model performance. Table 4.3 presents the results of both image classification model architectures. As the VGG16 model performed better than the lightweight model (GoogleNet), it will be discarded for further use and will serve only as a comparison model. Further recommendations and conclusions on the distress classification model will be summarized in Section 5.2 and 5.3.

4.3 DISTRESS SEGMENTATION

Iterations

Extracting and generating information of ~7 billion pixels values is quite a heavy task to be performed computationally and in time. To train the built EnDec model(s) on the annotated images, off the shelf High-Performance Computing (HPC) infrastructure facilities were utilized. 35,000 SBU(s) strategical business units were assigned as per the research author's estimation, for The Dutch National Supercomputer *Cartesius* (Fig. 4.7). It was only possible with the flagship cloud-computing services provided by SurfSara ("Dutch National Supercomputer Cartesius | SURF.nl," 2019), to overcome the major challenges of computing power, storage capacity and remote access in the field of Deep Learning (X. W. Chen & Lin, 2014) (Najafabadi et al., 2015).

Scaling up deep learning algorithms has been shown to lead to increased performance in benchmark tasks and to enable the discovery of complex high-level features. Recent efforts to train extremely large networks (with over 1 billion parameters) have relied on cloudlike computing infrastructure and thousands of CPU cores (Coates et al., 2013). Further details about the Cartesius system will follow in (Appendix D).

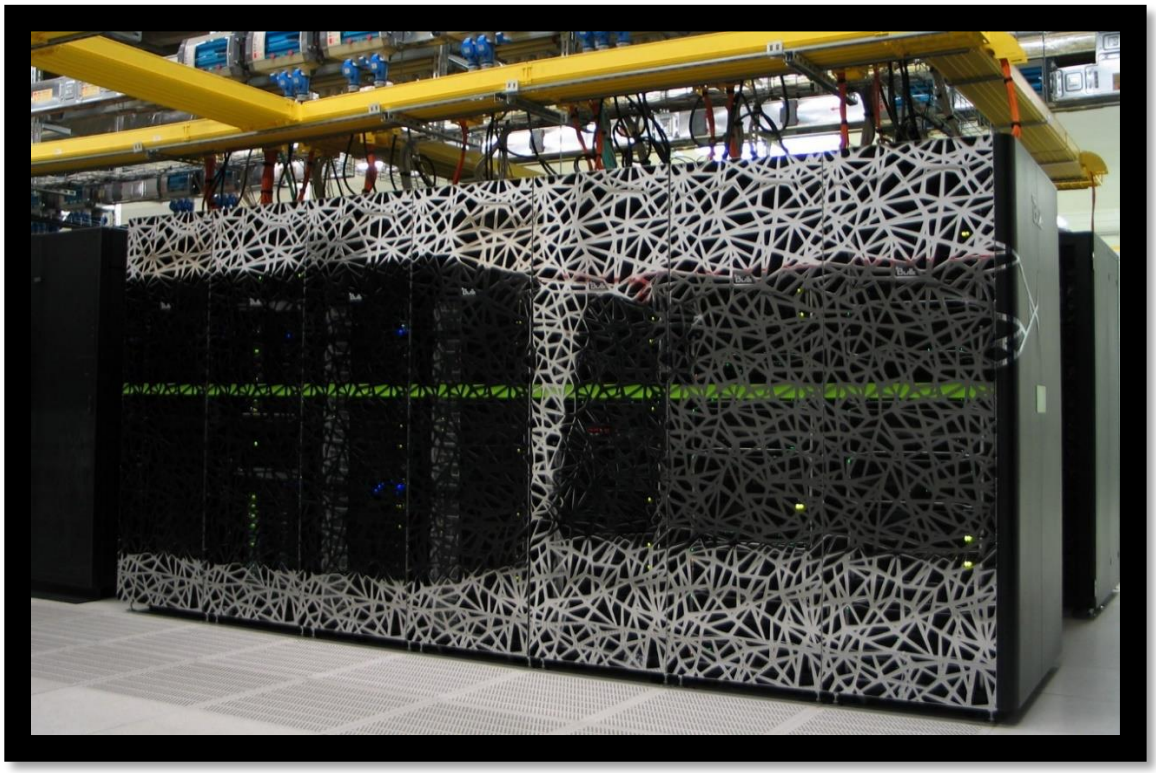


Fig. 4.7 The Dutch Supercomputer- Cartesius

The dataset for the segmentation task is split into three sets; *training* (70%), *validation* (20%), and *test* (10%). Until now as seen in Sec. 4.1 and 4.2 the training dataset was used to fit the model, and the validation set was used for fine-tuning and hyperparameter optimization. The reason for the test dataset split is to assess the performance of the algorithm by utilizing Intersection over Union (IoU) and Precision and Recall (F1) metrics. Later discussed in this section.

A step-based decaying learning rate at (1.00E-04) was scheduled, which in turn drop the *lr* by a dedicated factor every few epochs. It is represented as:

$$lr = \text{lr0} * \text{drop}^{\text{floor}(\text{epoch} / \text{epochs\_drop})}$$

To return the updated weights an Adam optimizer was implemented which is represented as:

$$\begin{cases} m_t = \beta_1 m_{t-1} + \left(1 - \beta_1\right) \nabla_\theta J\left(\theta\right) \\ \upsilon_t = \beta_2 \upsilon_{t-1} + \left(1 - \beta_2\right) \left(\nabla_\theta J\left(\theta\right)\right)^2 \\ \hat{m}_t = \frac{m_t}{1-\beta_1{}^t} \\ \hat{\upsilon}_t = \frac{\upsilon_t}{1-\beta_2{}^t} \\ \theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{\upsilon}_t + \varepsilon}} \hat{m}_t \end{cases} \tag{4.4}$$

where $\beta 1$ and $\beta 2$ represent the exponential decay rates for the moment estimates; t is the iteration index; η represents the learning rate;mt and vt represent the exponential moving averages of the first and second moments of the gradient, respectively; and $\hat{m}$ t and $\hat{v}t$ are the unbiased values of the mt and vt, respectively, calculated using the coefficients $\beta 1$ t and $\beta 2$ t (Kingma & Lei Ba, 2015).

To monitor the loss function for the majority of the classification task, a cross-entropy cost function was used which measures the similarity between the predicted value and the target value, as shown in the equation:

$$J(\theta) = -\frac{1}{N} \sum_{i=1}^{N} p_i \ln q_i + (1 - p_i) \ln(1 - q_i) \tag{4.5}$$

where $J(\theta)$ is the cost function, pi represents the target value of the ith pixel, qi represents the predicted value of the ith pixel, and N is the total number of pixels in the image (Z. Zhang & Sabuncu, 2018).

Results

To prevent overfitting of the model the learning rate was scheduled to 0.0001, and then every lightweight model was trained for 6 iterations, mediumweight for 12 iterations, and heavyweight for 16 iterations. Overall, 6 EnDec architectures were trained to perform the pixel-level segmentation for pavement surface cracks on a cloud computing infrastructure called *Cartesius.*

By utilizing multi-threading techniques (Kochura, Stirenko, Alienin, Novotarskiy, & Gordienko, 2018) each lightweight model was trained on a 64GB RAM, 24 cores of 2.4Ghz processors for approximately 5 hours of training time. For the mediumweight and heavyweight models, a 256GB RAM, 32 cores of 2.7Ghz processor was utilized, which required approximate training time of 19 hours and 28 hours respectively.

Model training was remotely monitored based on per epoch session to visualize the training loss and accuracy function. The training and validation loss along with the training and validation accuracy graphs for U-net are plotted (as seen in Fig. 4.8) and for LinkNet (as seen in Fig. 4.9 to visualize the performance as to how the 6 models performed.

The EnDec model(s) were evaluated with two sets of metrics IoU and F1. The precision and recall metrics calculate the low false positives and low false negatives, so that correct identification of classification score is not disturbed by false alarms. The F1 metric is calculated as:

$$\text{F1= 2x (precision*recall)/(precision+recall)} \tag{4.6}$$

The other metrics Intersection over Union (IoU) also known as the Jaccard Index compute the similarity between finite sample sets, divided by the size of the union from the *test* set. It is calculated to measure the model performance over object localization precision. The index is calculated as (Suphakit Niwattanakul*, Jatsada Singthongchai, 2013):

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{4.7}$$

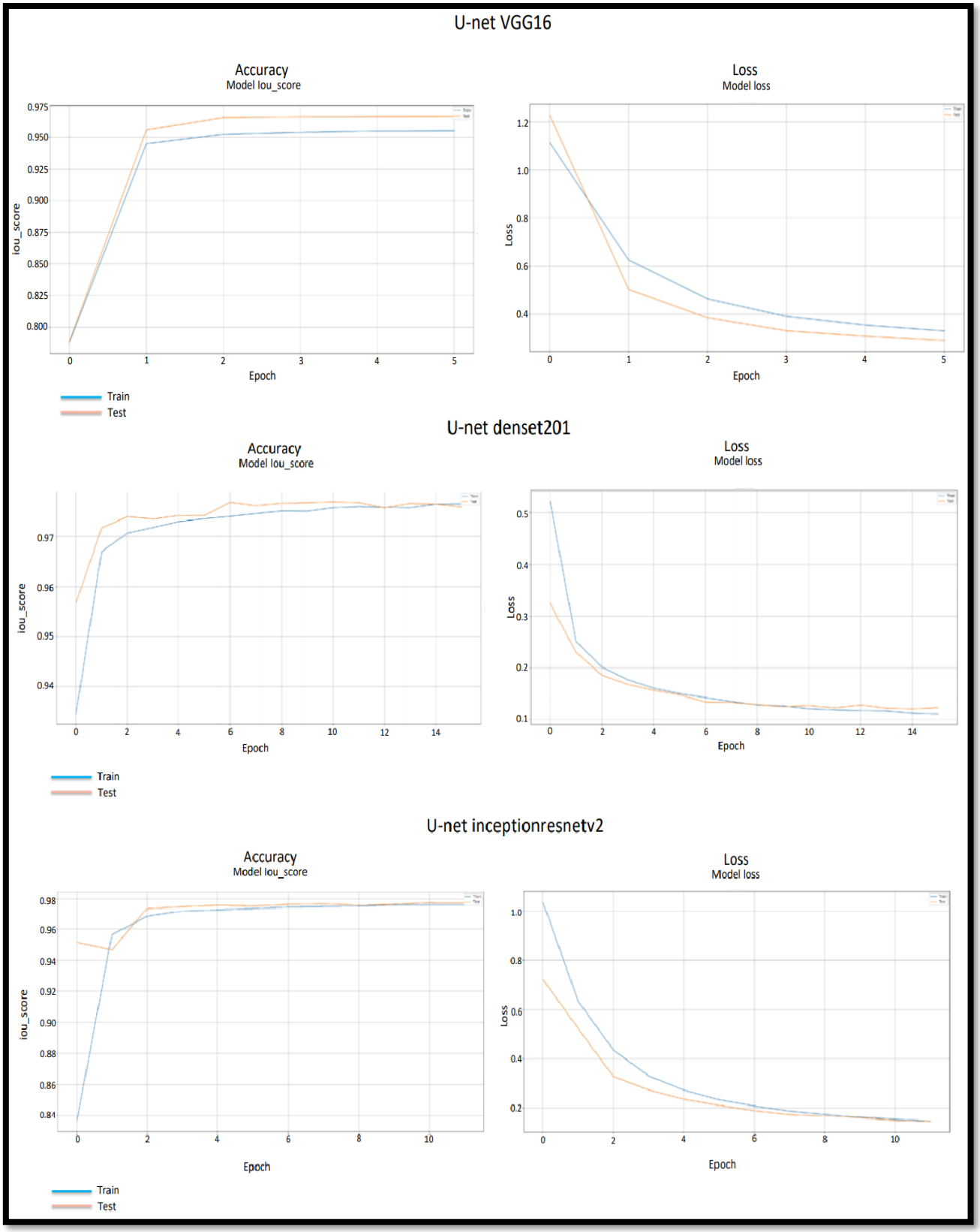


Fig. 4.8 U-net (VGG16, densenet201, inceptionresnetv2) accuracy and loss function graph

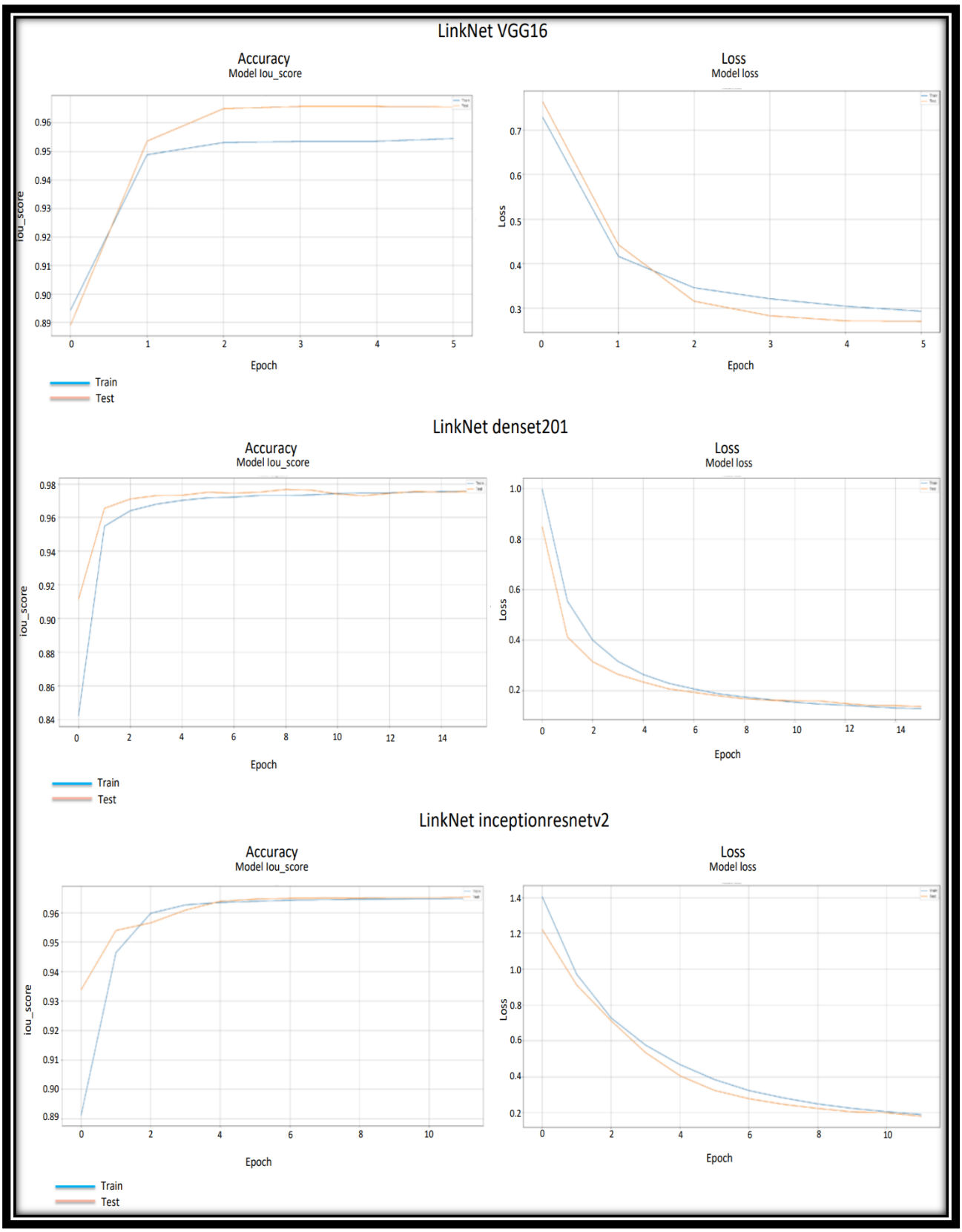


Fig. 4.9 LinkNet (VGG16, densenet201, inceptionresnetv2) accuracy and loss function graph

Model	Weights	Encoder	Top Accuracy	Parameters	Decoder	Hardware	Learning Rate	Epoch	Time	Train Loss	Val Loss	Test Accuracy	
			Imagenet			Cartesius			DD:HH:MM:SS			IoU	f1-score
Image Segmentation	LW	VGG16	0.713	13.8 M	U-net	(Thin node)	1.00E-04	6	00:04:16:10	0.289	0.374	0.949	0.974
	MW	Densenet201	0.77	25.6 M	U-net	(Fat node)	1.00E-04	12	00:19:39:42	0.143	0.207	0.96270	0.9810
	HW	Inceptionresnet (v3)	0.80	54.85 M	U-net	(Fat node)	1.00E-04	16	01:04:35:33	0.121	0.194	0.96250	0.9808
	LW	VGG16	0.713	13.8 M	LinkNet	(Thin node)	1.00E-04	6	00:04:59:02	0.27	0.372	0.948	0.973
	MW	Densenet201	0.713	25.6 M	LinkNet	(Fat node)	1.00E-04	12	00:18:20:35	0.137	0.208	0.961	0.980
	HW	Inceptionresnet (v3)	0.80	54.85 M	LinkNet	(Fat node)	1.00E-04	16	01:00:49:21	0.179	0.231	0.96400	0.9810

Table. 4.4 EnDec Image segmentation model results

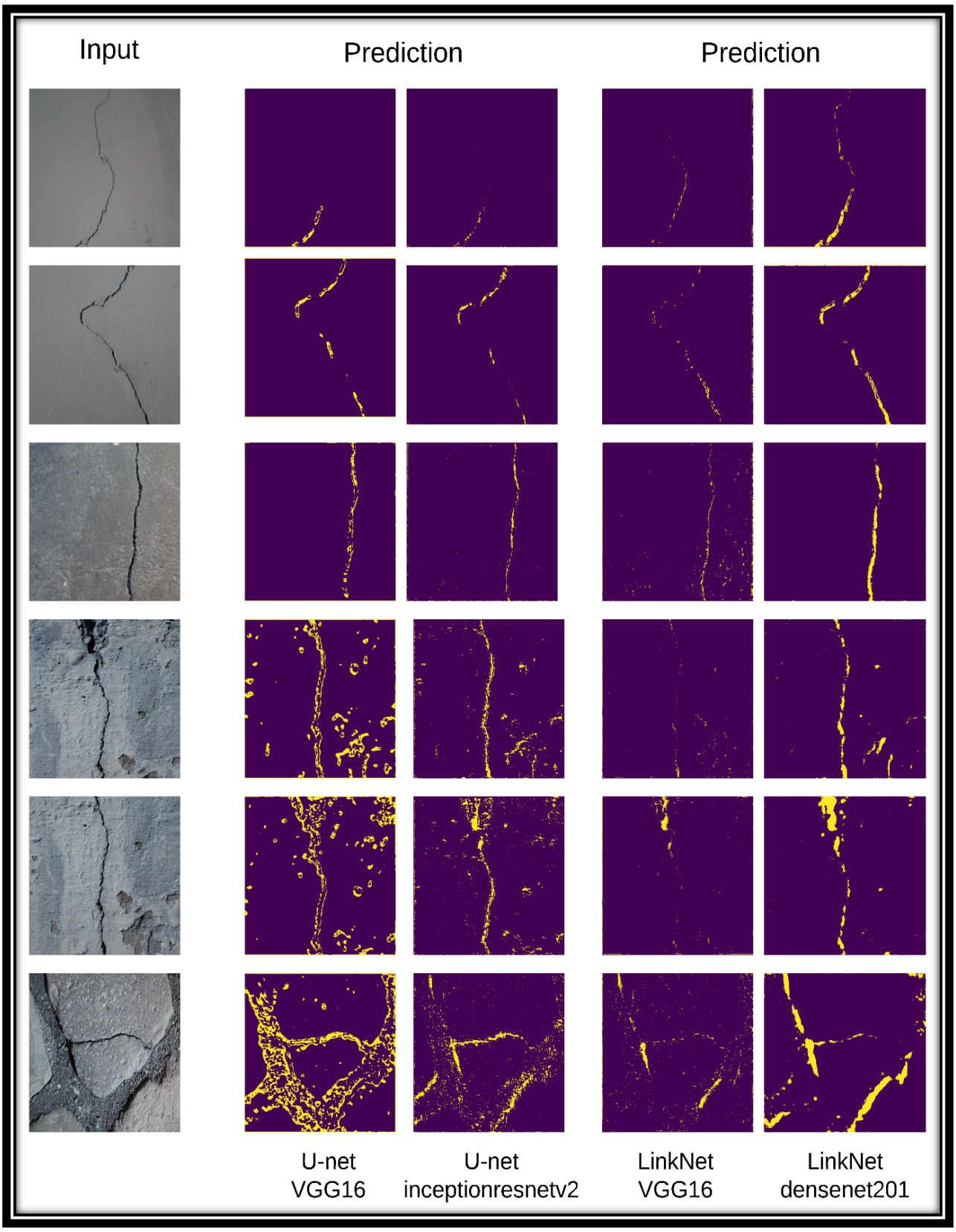


Fig. 4.10 Prediction masks output from the segmentation model

Table 4.4 provides an overview of all the 6 EnDec models that were trained for crack segmentation task. As can be seen, the lightweight models required lesser training time than the medium and heavyweight models but sabotaged on F1 and IoU accuracy. The minimum training loss (0.121) and validation loss (0.194) were obtained in *U-net-inceptionresenetv2* model. The highest classification accuracy (98.1%) was approximately similar in the case of *U-net-densenet201* and *LinkNet-Inceptionresenetv2*. Thus, suggesting that the model with a higher number of parameters performed well but also took more time for training and vice versa for the lightweight models.

The expected output of the segmentation model is to generate a binary mask of the object class (cracks). Some of the results produced as binary masks from the segmentation model can be visualized in Fig 4.10. Depending upon the output mask produced by each model further evaluation of severity ranking can be established. A dynamic range of thresholds could further categorize the severity ranking (*low, medium, high*) based on the pixels containing crack and non-crack portion of the image.

Chapter 5: Overview, Conclusion, Limitation & Further Research aspects

This chapter comprises of research overview and conclusions for the three data-driven approaches utilized in Section 3.3, 3.4 and 3.5. Later in this section, limitations, and recommendations for future research aspects will be summarized.

5.1 RESEARCH OVERVIEW

The pavement undergoes a fast deterioration process due to the damages induced by several factors like heavy traffic load, aging of the infrastructure, weather conditions, increase of traffic flow, construction and built quality, improper maintenance and many more. Periodic rehabilitation measures are thus needed to maintain the condition of the underlying asset above the minimum acceptable condition as can be seen in Fig. 5.1.

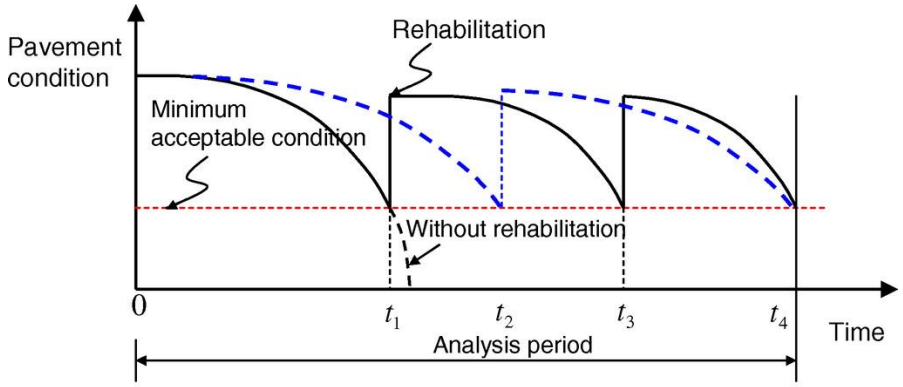


Fig. 5.1 Pavement Life-cycle (Bucher & Frangopol, 2006)

More often than not, the rehabilitation strategies are becoming more reactive as the road networks are getting denser by the day. Moreover, the extensive factors leading to all the types of distresses cost higher and reduces the life-cycle, if not rehabilitated in time. Thus, creating an urgency to detect and identify the distresses as early as possible.

Since damages on the road impose economic setbacks and concern with the safety and comfort of riders too, governmental authorities consider this as a high-priority issue. In order to repair the damages in their 'early' stages, PMS of the Province of Zuid Holland (PZH) deploys multiple process-driven inspection strategies. Such visual inspection strategies (Section 2.1) are conducted only in a specific time of interval, are cost-ineffective and unsafe for the road inspectors during implementation.

As a reason, there are immediate research needs to develop and suggest another way-of-approach towards the indispensable issue. In order to find a promising solution, the

research scope (Section 1.4) dilated towards selecting a data-driven approach for testing its applicability and reliability in detecting and classifying pavement surface distresses.

The research objective was to focus on modern data sources such as smartphones which are economical and frequent crowdsourced data generators. To attain the research objective, a set of research questions were formulated.

Research question:

Can we utilize modern data-sources such as smartphones for detecting, classifying and segmenting pavement surface-distresses?

Sub-questions:

I. *How can accelerometer data be utilized for anomaly detection and roughness prediction?*

II. *How can image data be utilized for classifying pavement surface distresses?*

III. *How can an image data be utilized for segmenting pavement surface distress?*

In Section 3.1 a 7 step procedure was established to develop the data-driven models and has been utilized for answering the three sub-questions of the research. A detailed outcome of each ML model and its performance results are crafted in Section 4.1, 4.2 and 4.3.

5.2 CONCLUSIONS

To answer the main research question, three data-driven models were built to assess the applicability and reliability of the proposed method by utilizing modern data sources such as smartphones. At first, accelerometer sensor data was put to use for detecting anomalies and predicting IRI values, as reliable as ARAN measurements. From, the results (Section 4.1), it was apparent that the approach showed promising results. Secondly, a data-driven image classification model was built to automate the mundane task of manually classifying pavement surface distresses. Lastly, a semantic segmentation model was built to objectively estimate the severity of the given surface distress.

Below is a general conclusion (for each model) from the author's perspective on how each model was developed and what research boundaries were faced. Also, the applicability and reliability it can provide for detecting and assessing pavement surface distress will be discussed as well.

Accelerometer Data

Vibration sensors have been commercially exploited and extensively researched due to their multitude of capabilities. Accelerometer sensors in specific have a high-frequency response, higher sensitivity, low cost, and mobile capabilities.

To build the data-driven model from scratch, a built-in smartphone accelerometer sensor was utilized for the data collection procedure. The motive behind manual data collection was to simulate an unstructured manner for gathering raw accelerometer data. The benefit of such a procedure is to mimic the generalized behavior of riders. In such a condition, where a smartphone is docked in a vehicle and readings from the accelerometer sensor has been recorded by the permission of the user, the model should be able to single-handedly detect anomalies in the road surface and classify the roughness index too.

The most prudent algorithm Random Forest (more in Section 3.3.3) was utilized for the unsupervised task at hand. The benefit of such algorithms is that it assigns weight factors to all the input variables in a completely unbiased manner. The disadvantage of such an approach is that it requires enormous data. To generate more data, not only the use-case network (Section 3.3.2) was driven over three times, but also various feature extraction techniques (Section 3.3.4) like applying a wavelet transform were also utilized.

Most importantly, after 40 times iterating the algorithm over 125,325 data points, the model seem to build a relation between the 82 input variables and IRI values measured from the visual inspection database of PZH from 2018. The validation of the model was put against the highly sophisticated visual inspection road analyzer called ARAN. The model performance graphs (Fig. 4.1) and the predicted IRI table (Table 4.2) suggest that the model could estimate a good range of IRI values (error rate +-0.5), except the upper (above 4.8) and lower extremes (below 0.8). A clear indication is that the more data gets collected with such crowdsourced appliances the better model predictions will become.

With the help of GPS location tagging a comparison map was plotted (Fig 4.2) against the validation dataset gathered from ARAN vehicle, which also symbolizes that the model reliability is close to the actual readings. Also, there is almost a year gap between when both the datasets were collected.

Limitations and further recommendations will follow in Section 5.3

Image Classification Data

Post visual inspections, the PZH appoints third parties to manually categorize the 2D image dataset based on the distresses which are visually perceivable. To perform this arduous and repetitive task several man-hours of work is demanded. Classification of images based on object categorization has long been studied in the CV domain. The use-case for this research was to build a data-driven model that can automate the task of manual classification of images for pavement surface distresses.

The goal at hand was to develop a model that could segregate images containing objects which could interest the PMS of PZH. To obtain a relevant size of the dataset for the use-case, high-definition portable cameras (GoPro) was utilized. A dataset of images containing 6 classes of objects like *crack*, *potholes*, *asphalt without distress*, *manhole*, *road-marking*, and *shadow* was structured in a way how Kaggle competitions suggest.

A total of 15,750 images were labeled in a semi-supervised manner, for the multi-label classification task at hand. Moreover, techniques of Transfer Learning were utilized to minimize the computational power required to train the model and the need to built a CNN from ground-up. Pre-trained weights from the ImageNet dataset proved to be beneficial in model convergence. A custom sequential classifier was used to replace the last layers of the VGG16 model (Section 3.4.5), to classify the 6 object classes.

With over 9 hours of training time on an NVIDIA GPU GeForce 940M and 25 iterations, early stopping was implemented to prevent model overfitting. The best validation accuracy achieved was ~86% with the corresponding validation loss of 0.43. A confusion matrix was plotted (Fig. 4.5) to assess the model performance per class category. The model was set to further testing by predicting on images that it has never encountered before. The results of predictions can be visualized in Fig 4.6.

Thus suggesting that the model can be put to use to automate the tedious task of manually classifying pavement surface distresses from 2D images. Limitations and further research aspects will follow in Section 5.3.

Image segmentation Data

Rehabilitation strategies depend on first analyzing the type of distress and secondly determining the severity of the classified distress to schedule proper maintenance plans. Since in the primary and secondary road networks of south Holland, cracks are the most commonly occurring distress, thus the use-case will develop a data-driven tool to objectively classify the dominant pavement surface damage.

Unlike in the classification task where the model output is a binary digit assigned to its label category, segmentation models not only have to understand the object patterns but also need to properly localize the object and generate a similar mask as the desired output. When an output from a model is pre-defined, it becomes a supervised task and the data labeling techniques need to be followed accordingly.

Since semantic segmentation is a pixel-level classification task, a total of 22,996 images were annotated per-pixel wise by utilizing a web-based annotation tool called LableMe. The dataset comprises an equal number of images containing *cracks* and their corresponding *masks* resulting in more than (~7 billion) pixels to train on. The EnDec architecture utilized (Section 3.5.5) comprises a contraction path and an expansion path of two neural networks attached one after the other. A total set of 3 encoders: one lightweight (VGG16), one mediumweight (densenet201) and one heavyweight (inceptionresnetv2) served as a contraction path. On the expansion path of the decoder, two profoundly used neural networks (U-net and LinkNet) were placed to output the features back to an image.

Training with such a large amount of pixels and network parameters (13 million to 54 million) wasn't possible to execute on a local machine, thus cloud-computing services provided by Surfsara played a crucial role. The Dutch super-computer *Cartesius* provided multi-threading (Section 4.3) opportunities to parallelly train the model on multiple cores of a remotely accessed CPU. The lightweight models like U-net-VGG16 could achieve their best performances in just about 6 epochs (~ 4 hours of training time) but compensate for lesser accuracy. Heavyweight models on the other side iterated for about 16 times over a period of (~26 hours) but resulted in higher precision and recall accuracy (F1) and also performed quite satisfactorily on the IoU (Intersection over Union) scores as well.

The performance of each model can be visualized in Fig. 4.8 and their generated output in Fig. 4.9. Also, intricate details of the learning rate, hardware requirements, training loss, validation loss, etc. of all the 6 models are illustrated in Table. 4.4.

Based on the output masks generated by the segmentation model of cracked pavements, the portion of cracked pixels was calculated, and a dynamic threshold value was assigned to further categorize the cracks by its severity level.

If the requirement is to accurately locate and generate cracks from the images (as low as hairline cracks) heavyweight models can perform the task pretty accurately. If the requirement is to quickly process bulk images while compensating for precision accuracy, lightweight models ought to perform significantly faster.

Clearly, the results of the research depend on the research scope boundaries that were pre-defined in Section 1.4. The limitations of the obtained results and further aspects where the research can proceed are further discussed below.

5.3 LIMITATIONS AND FUTURE RESEARCH ASPECTS

The section comprises of research limitations and future research aspects for every built model (in Section 3.3, 3.4, 3.5) is as under:

- Accelerometer Data

  - A diverse dataset (including different vehicles, different road material, etc.) along with a larger volume could result in better prediction of IRI.
  - Multiple accelerometer readings per tire/axel could create even more accurate and generalized IRI predictions.
  - The orientation of the phone was pivoted and kept horizontally assuming the exclusion of randomly created noise by the users.
  - Better location accuracy by GPS tagging could generate in a finer and detailed location map for the good, bad and worse road sections.
  - With the use of Federated Learning, governments can benefit by requesting road-users data to monitor the condition of the pavement, in exchange for value.

- Image Classification

  - Further categories of labeled data can be incorporated so that more objects or classes can be classified, as per the PMS demands.
  - Pre-trained weights from the ImagNet dataset seemed to work well, but more open-source datasets like MS COCO, PASCAL, CIFAR can be utilized and compared.
  - Better accuracy can be achieved in return for the training time. Denser networks like resnet50, alexnet, squeezenet can also be tested on the prepared dataset.
  - Outliers data in the training set which reduced the model performance can be removed/replaced for better accuracy.

- Image Segmentation

  - The dataset comprises of pictures that were either taken orthogonally or without noise. A diverse dataset (including dashcam videos) can also be utilized for further generalization.

- The output of the segmentation model totally depends on the type of dataset (supervised learning) it is been trained upon. One could prepare a much accurate and larger dataset.
- The severity ranking assigned in the research was based on the number of pixels where the distress was found, a dataset could be prepared where each pixel can be assigned to a specific area unit.
- GAPv2 dataset where multiple distresses with annotated with pixel-level details is available for research purposes, a similar approach can also be applied to configure the dataset for PZH.

Above mentioned aspects for further research will not only aid in improving the model accuracy but also its reliability for the PMS to adopt. In particular, using the rules of Federated learning, governments can benefit by leveraging road-user data through smartphones. By further training the ML models with larger and diverse datasets, it should contribute tremendously in increasing the reliability and accuracy.

About Us | International Cybernetics. (1975). Retrieved November 26, 2019, from https://www.internationalcybernetics.com/about-us/

albumentations · PyPI. (2019). Retrieved October 22, 2019, from https://pypi.org/project/albumentations/

Amin, M. S. R. (2015). The Pavement Performance Modeling: Deterministic vs. Stochastic Approaches. In *Numerical Methods for Reliability and Safety Assessment* (pp. 179–196). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-07167-1_5

ArcGIS Help | ArcGIS Desktop. (2019). Retrieved October 18, 2019, from https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/image-classification/what-is-image-classification-.htm

Ardeshir, S., Collins-Sibley, K. M., & Shah, M. (2015). *Geo-semantic Segmentation*.

Asphalt and road building | TNO. (2019). Retrieved September 25, 2019, from https://www.tno.nl/en/focus-areas/buildings-infrastructure-maritime/roadmaps/buildings-infrastructure/infrastructure/asphalt-and-road-building/

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *ArXiv Preprint ArXiv:1511.00561*.

Bianco, S., Cadene, R., Celona, L., & Napoletano, P. (2018). Benchmark analysis of representative deep neural network architectures. *IEEE Access*, *6*, 64270–64277. https://doi.org/10.1109/ACCESS.2018.2877890

Breiman, L. (2001). ST4_Method_Random_Forest. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1017/CBO9781107415324.004

Breiman, L., & Cutler, A. (2019). Random Forests. Retrieved October 15, 2019, from https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Bucher, C., & Frangopol, D. M. (2006). Optimization of lifetime maintenance strategies for deteriorating structures considering probabilities of violating safety, condition, and cost thresholds. *Probabilistic Engineering Mechanics*, *21*(1), 1–8. https://doi.org/10.1016/j.probengmech.2005.06.002

Casas-avellaneda, D. ., & Lopez-parra, J. . (2016). Detection and localization of potholes in roadways using smartphones. *Detección y Localización de Imperfecciones Viales Utilizando Smartphones. DYNA, 83, 44919*.

Chaurasia, A., & Culurciello, E. (2017). *LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation*.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., L., A., & Yuille. (2016). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *ArXiv Preprint ArXiv:1606.00915*.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous

separable convolution for semantic image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11211 LNCS*, 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

Chen, X. W., & Lin, X. (2014). Big data deep learning: Challenges and perspectives. *IEEE Access*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2014.2325029

Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Kumar Venugopal, V., … Ai, Q. (2018). *Development and Validation of Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans*.

Coates, A., Huval, B., Wang, T., Wu, D. J., Ng, A. Y., & Catanzaro, B. (2013). *Deep learning with COTS HPC systems*.

Coenen, T. B. J., & Golroo, A. (2017). A review on automated pavement distress detection methods. *Cogent Engineering*, *4*(1), 1–23. https://doi.org/10.1080/23311916.2017.1374822

Cord, A., & Chambon, S. (2012). Automatic road defect detection by textural pattern recognition based on adaboost. *Computer-Aided Civil and Infrastructure Engineering*, *27*, 224–259.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., … Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Cormode, G., Shkapenyuk, V., Srivastava, D., & Xu, B. (2015). *Forward Decay: A Practical Time Decay Model for Streaming Systems*.

CROW. (2011). Knowledge base. Retrieved November 26, 2019, from http://kennisbank.crow.nl/Kennismodule

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE Com-Puter Society Conference on Computer Vision and Pattern Recognition, Vol. 1, CVPR2005, San Diego, CA, USA, IEEE*, 886–893.

Deligiannidis, L., & Arabnia, H. R. (2014). *Emerging Trends in Image Processing, Computer Vision and Pattern Recognition*. *Emerging Trends in Image Processing, Computer Vision and Pattern Recognition*. Elsevier Inc. https://doi.org/10.1016/C2014-0-01692-9

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C. (2016). The importance of skip connections in biomedical image segmentation. *In Deep Learning and Data Labeling for Medical Applications*, 179–187.

Dutch National Supercomputer Cartesius | SURF.nl. (2019). Retrieved October 23, 2019, from https://www.surf.nl/en/dutch-national-supercomputer-cartesius

Dynatest | Pavement Evaluation & Consulting. (1976). Retrieved November 26, 2019, from https://www.dynatest.com/

E., R., & G., R. (2016). Comparative Study of Different Wavelet Transforms in Fusion of Multimodal Medical Images. *International Journal of Computer Applications*, *146*(11), 18–24. https://doi.org/10.5120/ijca2016910899

Eldin, N. ., & Senouci, A. B. (1995). Use of neural networks for condition rating of jointed concrete pavements. *Advances in Engineering Software*, *23*, 133–141.

Feurer, M., & Hutter, F. (2012). *Hyperparameter Optimization*.

Fred Agarap, A. M. (2019). *Deep Learning using Rectified Linear Units (ReLU)*. Retrieved from https://github.com/AFAgarap/relu-classifier.

Fugro. (2019). Equipment and software | Fugro. Retrieved November 26, 2019, from https://www.fugro.com/our-services/asset-integrity/roadware/equipment-and-software

Ge, R., Kakade, S. M., Kidambi, R., & Netrapalli, P. (2019). The Step Decay Schedule: A Near Optimal, Geometrically Decaying Learning Rate Procedure. Retrieved from http://arxiv.org/abs/1904.12838

Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., & Agrawal, A. (2017). Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and Building Materials*, *157*, 322–330. https://doi.org/10.1016/j.conbuildmat.2017.09.110

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, *24*(2), 8–12. https://doi.org/10.1109/MIS.2009.36

Hawkeye Software - ARRB Group. (2014). Retrieved November 26, 2019, from http://arrbgroup.net/products/hawkeye-software/

He, K., Zhang, X., Ren, S., & J, S. (2016). Deep residual learning for image recognition. *IEEE Conference*, 770–778.

Hein, D., & Watt, D. (2005a). Municipal Pavement Performance Prediction based on Pavement Condition Data. *Annual Conference of the Transportation Association of Canada, Calgary, Alberta*.

Hein, D., & Watt, D. (2005). (2005b). Municipal Pavement Performance Prediction based on Pavement Condition Data,. *Annual Conference of the Transportation Association of Canada , Calgary, Alberta.*

Home - Roadscanners. (1998). Retrieved November 26, 2019, from https://www.roadscanners.com/

Huang, C. M. (2014). Anecdotes extraction from webpage context as image annotation. In *Emerging Trends in Image Processing, Computer Vision and Pattern Recognition* (pp. 353–367). Elsevier Inc. https://doi.org/10.1016/B978-0-12-802045-6.00022-3

Ibtehaz, N., & Rahman, M. S. (2019). *MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation*.

ImageNet. (2019). Retrieved October 21, 2019, from http://www.image-net.org/

ImageNet Large Scale Visual Recognition Competition 2014 (ILSVRC2014). (2014). Retrieved October 21, 2019, from http://www.image-net.org/challenges/LSVRC/2014/results

Jansen, K., & Zhang, H. (2007). Scheduling malleable tasks. *Handbook of Approximation Algorithms and Metaheuristics*, 45-1-45–16. https://doi.org/10.1201/9781420010749

Kaggle: Your Home for Data Science. (2019). Retrieved October 18, 2019, from https://www.kaggle.com/

Kato, T. (1992). <title>Database architecture for content-based image retrieval</title>. In *Image Storage and Retrieval Systems* (Vol. 1662, pp. 112–123). SPIE. https://doi.org/10.1117/12.58497

Kaymak, Ç., & Uçar, A. (2017). *A Brief Survey and an Application of Semantic Image Segmentation for Autonomous Driving*.

KICT. (2017). *Final report of the National Highway Pavement Management System 2016*.

Kim, H., K., H., H., Y.W., & Byun, H. (2018). Detecting construction equipment using a region-based fully convolutional network and transfer learning. *Journal of Computing in Civil Engineering, 32(2), 04017082.*

Kingma, D. P., & Lei Ba, J. (2015). *ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION*.

Kochura, Y., Stirenko, S., Alienin, O., Novotarskiy, M., & Gordienko, Y. (2018). Performance analysis of open source machine learning frameworks for various parameters in single-threaded and multi-threaded modes. In *Advances in Intelligent Systems and Computing* (Vol. 689, pp. 243–256). Springer Verlag. https://doi.org/10.1007/978-3-319-70581-1_17

Kohonen, T., & Simula, O. (1996). "Engineering Applications of the SelfOrganizing Map." *Proceeding of the IEEE, Vol. 84, N*, pp.1354 – 1384.

Krizhevsky, A., Sutskever, I., & Hinton, G. . (2012). *Imagenet classification with deep convolutional neural networks*.

Laaksonen, J., & Oja, E. (1996). Classification with learning k-nearest neighbors. In *IEEE International Conference on Neural Networks - Conference Proceedings* (Vol. 3, pp. 1480–1483). IEEE. https://doi.org/10.1109/icnn.1996.549118

Lavin, P. (2003). Asphalt Pavements: A Practical Guide to Design, Production and Maintenance for Engineers and Architects. Retrieved September 25, 2019, from https://books.google.com/books?id=XuM3YYeQrk8C&pgis=1

Li, Q., Zou, Q., Zhang, D., & Mao, Q. (2011). FoSA: F* seedgrowing approach for crack-line detection from pavement images. *Image and Vision Computing*, *29*, 861–872.

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks with identity mappings for high-resolution semantic segmentation. *CVPR*.

Lin, H. Y., Liang, S. Y., Ho, Y. L., Lin, Y. H., & Ma, H. P. (2014). Discrete-wavelet-transform-based noise removal and feature extraction for ECG signals. *IRBM*, *35*(6), 351–361. https://doi.org/10.1016/j.irbm.2014.10.004

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *CVPR*.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput.Vis. 60 (2)*, 91–110.

Mahmood, M., Rahman, M., Nolle, L., & Mathavan, S. (2013). A fuzzy logic approach for pavement section classification. *International Journal of Pavement Research and Technology*, *6*(5), 620–626. https://doi.org/10.6135/ijprt.org.tw/2013.6(5).620

Mancini, A., Malinverni, E. S., Frontoni, E., & Zingaretti, P. (2013). Road pavement crack automatic detection by MMS images. In *In 2013 21st Mediterranean Conference on Control and Automation, MED 2013 - Conference Proceedings* (pp. 1589–1596). Platanias-Chania.

Mathavan, S., Kamal, K., & Rahman, M. (2015). A review of three-dimensional imaging technologies for pavement distress detection and measurements. *IEEE Transactions on Intelligent Transportation Systems*, *16*, 2353–2362.

McGuinness, K., & O'connor, N. E. (2010). A comparative evaluation of interactive segmentation algorithms. *Pattern Recognition, 43(2)*, 434–444.

Menegola, A., Fornaciali, M., Pires, R., Bittencourt, F. V., Avila, S., & Valle, E. (2017). Knowledge transfer for melanoma screening with deep learning. In *Proceedings - International Symposium on Biomedical Imaging* (pp. 297–300). IEEE Computer Society. https://doi.org/10.1109/ISBI.2017.7950523

Milioto, A., Lottes, P., & Stachniss, C. (2018). Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In *Proceedings - IEEE International Conference on Robotics and Automation* (pp. 2229–2235). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICRA.2018.8460962

Mjolsness, E., & DeCoste, D. (2001). Machine learning for science: state of the art and future prospects. *Science 293:2051–2055*.

Moghadas Nejad, F., & Zakeri, H. (2011). A comparison of multi-resolution methods for detection and isolation of pavement distress. *Expert Systems with Applications*, *38*, 2857–2872.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1). https://doi.org/10.1186/s40537-014-0007-7

Nega, A., Nikraz, H., & Al-Qadi, I. L. (2016). Dynamic analysis of falling weight deflectometer. *Journal of Traffic and Transportation Engineering (English Edition)*, *3*(5), 427–437. https://doi.org/10.1016/j.jtte.2016.09.010

Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. *EECV*.

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. *ICCV*.

Oliveira, H., & Correia, P. L. (2013). Automatic road crack detection and characterization. *In IEEE Transactions on Intelligent Transportation Systems*, *14*, 155–168.

Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E. (2016). Enet: A deep neural network architecture for real-time semantic segmentation. *ArXiv Preprint ArXiv:1606.02147*.

Paszke, Adam, Gross, S., Chintala, S., Chanan, G., Yang, E., Facebook, Z. D., … Lerer, A. (2017). *Automatic differentiation in PyTorch*.

PaVision® Automated Pavement Condition Assessment System | www.ara.com. (2015). Retrieved November 26, 2019, from https://www.ara.com/projects/pavision-automated-pavement-condition-assessment-system

Prechelt, L. (1997). *Early Stopping | but when?*

Premachandra, C., Premachandra, H. W. H., Parape, C. D., &, & Kawanaka, H. (2015). Road crack detection using color variance distribution and discriminant analysis for approaching smooth vehicle movement on non-smooth roads. *International Journal of Machine Learning and Cybernetics*, *6*, 545–553.

Puan, O. C., Mustaffar, M., & Ling, T.-C. (2007). Automated Pavement Imaging Program (APIP) for pavement cracks classification and quantification. *Malaysian Journal of Civil Engineering*, *19(1)*, 1–16.

Radopoulou, S. C., & Brilakis, I. (2016). Improving Road Asset Condition Monitoring. *Transportation Research Procedia*, *14*(0), 3004–3012. https://doi.org/10.1016/j.trpro.2016.05.436

Road crack detection. (2019). Retrieved from www.csiro.au

ROMDAS | Road Survey Equipment , Pavement Condition Survey. (1990). Retrieved November 26, 2019, from https://romdas.com/

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Retrieved from http://lmb.informatik.uni-freiburg.de/

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). *LabelMe: a database and web-based tool for image annotation*. *INTERNATIONAL JOURNAL OF COMPUTER VISION* (Vol. 77).

Sagheer, A. M., Kubaisy, A., Y.A., & Awad, H. . (2008). Diagnosis of Flexible Pavement Road Deterioration by Using Expert System. *Iraqi Journal of Civil Engineering*, *12*, 1–25.

Saito, S., Li, T., & Li, H. (2016). Real-time facial segmentation and performance capture from RGB input. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9912 LNCS, pp. 244–261). Springer Verlag. https://doi.org/10.1007/978-3-319-46484-8_15

Sathya, R., & Abraham, A. (2013). THE SCIENCE AND INFORMATION ORGANIZATION Editorial Preface. *INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE (IJARAI) International Journal of Advanced Research in Artificial Intelligence*, *2*(2), 34–38. Retrieved from www.ijarai.thesai.org

Schnebele, E., Tanyu, B. F., Cervone, G., & Waters, N. (2015). Review of remote sensing methodologies for pavement management and assessment. *European Transport Research Review*, *7*(2). https://doi.org/10.1007/s12544-015-0156-6

Semantic Segmentation. (2018). Retrieved October 22, 2019, from https://www.kdnuggets.com/2018/10/semantic-segmentation-wiki-applications-resources.html

Seraj, F., Van Der Zwaag, B. J., Dilo, A., Luarasi, T., & Havinga, P. (2016). Roads: A road pavement monitoring system for anomaly detection using smart phones. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9546*, 128–146. https://doi.org/10.1007/978-3-319-29009-6_7

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., F., R, & LeCun, Y. (2014). Overfeat: Integrated recognition, localization and detection using convolutional networks. *ICLR*.

Shabajee, P., Hannabuss, S., & Tilsed, I. (1998). Book Reviews. *Interactive Learning Environments*, *6*(3),

281–290. https://doi.org/10.1076/ilee.6.3.281.3600

Simonyan, K, & Zisserman, A. (2014). Very deep convoultional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Simonyan, Karen, & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition, 1–14. Retrieved from http://arxiv.org/abs/1409.1556

Solomatine, D., & Shrestha, D. (2009). A novel method to estimate model uncertainty using machine learning techniques. *Water Resources, Res 45:W00*.

Souza, V. M. A. (2018). Asphalt pavement classification using smartphone accelerometer and Complexity Invariant Distance. *Engineering Applications of Artificial Intelligence*, *74*, 198–211. https://doi.org/10.1016/j.engappai.2018.06.003

Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S. S., & Babu, R. V. (2017). *An Introduction to Deep Convolutional Neural Nets for Computer Vision*. *Deep Learning for Medical Image Analysis* (1st ed.). Elsevier Inc. https://doi.org/10.1016/B978-0-12-810408-8.00003-1

Suphakit Niwattanakul*, Jatsada Singthongchai, E. N. and S. W. (2013). Using of Jaccard Coefficient for Keywords Similarity. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2013*, *1*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 07-12-June-2015, pp. 1–9). IEEE Computer Society. https://doi.org/10.1109/CVPR.2015.7298594

T.Y, L., P, D., R, G., K, H., B, H., & Belongie. S. (2017). Feature pyramid networks for object detection. *CVPR*.

Terzi, S. (2006). Modelling the pavement present serviceability index of flexible highway pavement using data mining. *Journal of Applied Sciences*, *6*, 193–197.

Ticlavilca Torres, Alfonso, A. M. (2008). Data Driven Models and Machine Learning (ML) Approach in Water Resources Systems, (Ml).

TNO. (2010). *Auto als Sensor - quick-scan naar detectiemogelijkheden van wegdekschade*.

Tong, S., & Koller, D. (2011). Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research*, 45–66.

U-Net: Image Segmentation Network. (2018). Retrieved October 23, 2019, from https://neurohive.io/en/popular-networks/u-net/

Visa, S., Ramsay, B., Ralescu, A., & Van Der Knaap, E. (2011). *Confusion Matrix-based Feature Selection*.

Visual Geometry Group - University of Oxford. (2014). Retrieved October 21, 2019, from https://www.robots.ox.ac.uk/~vgg/

Wang, J., & Perez, L. (2017). *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*.

Wang, M., Birken, R., & Shahini Shamsabadi, S. (2015). Implementation of a multi-modal mobile sensor system for surface and subsurface assessment of roadways. In *Smart Sensor Phenomena, Technology, Networks, and Systems Integration 2015* (Vol. 9436, p. 943607). SPIE. https://doi.org/10.1117/12.2084852

Wu, D. J., Feng, T., Naehrig, M., & Lauter, K. (2016). Privately Evaluating Decision Trees and Random Forests. *Proceedings on Privacy Enhancing Technologies*, *2016*(4), 335–355. https://doi.org/10.1515/popets-2016-0043

Wulamu, A., Shi, Z., Zhang, D., & He, Z. (2019). Multiscale Road Extraction in Remote Sensing Images. *Computational Intelligence and Neuroscience*, *2019*, 1–9. https://doi.org/10.1155/2019/2373798

Xiancheng, W., Wei, L., Bingyi, M., He, J., Jiang, Z., Xu, W., … Zhaomeng, S. (2018). *Retina Blood Vessel Segmentation Using A U-Net Based Convolutional Neural Network*. Retrieved from www.sciencedirect.comProcediaComputerScience00

Yang, J., Jiang, .G., Hauptmann, .G., & C.-W. Ngo. (2007). Evaluating bag-of-visual-words rep-resentations in scene classification. *Proceedings of the International Workshop OnWorkshop on Multimedia Information Retrieva*, 197–206.

Yoo, H.-S., & Kim, Y.-S. (2016). Development of a crack recognition algorithm from non-routed pavement images using artificial neural network and binary logistic regression. *KSCE Journal of Civil Engineering*, *20*, 1151–1162.

Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *ArXiv Preprint ArXiv:1511.07122*.

Zamir, A., R., S., A., S., W., G., L. J., Malik, J., & & Savarese, S. (2018). Taskonomy: Disentangling task transfer learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3712–3722.

Zhang, A., Wang, K., C., F., Y., L., Y., T., S., C., & C., Li, B. (2018). Deep learning–based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet. *Journal of Computing in Civil Engineering, 32(5), 04018041.*

Zhang, A., Wang, K. C., Fei, Y., Liu, Y., C., C., Y., G., Q., & S. (2018). Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network. *Computer-Aided Civil and Infrastructure Engineering, 34*, 213–229.

Zhang, Z., & Sabuncu, M. R. (2018). *Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels*.

Appendix A

Road sections and road section parts

The inspection is carried out on road sections that consist of one or more road section components (roadway, footpath, cycle path, etc.). If the roadway is wider than 4.50 m, it is recommended to divide the roadway into at least two road section components (roadway on the left and roadway on the right) due to the observation width. The method of road management uses the principle that a road section must be homogeneous for both traffic function and cross-sectional profile, as well as in pavement construction. A road section is a part of a road where the road is divided in a longitudinal direction into one or more sections (sections). The length of the road section influences the accuracy of the inspection results. Due to the length of the road section it can happen that the amount of damage is insufficient for a damage assessment. For instance: with 5 m cracking in a 100 m road section the damage is noted. If the same crack occurs within a 200 m road section, there is insufficient scope for a damage assessment.

For the sake of clarity and completeness, the road section layout must be used for the inspection. The length of the section of the road section plays an important role in the detailed inspection. The inspector has so much to remember about the severity and extent of the various damages that a standard length of 100 m is advised. The length of 100 m also plays a role in the context of consistency in investigations and measurements and also fits in well with the often present hectometer on the high lane roads.

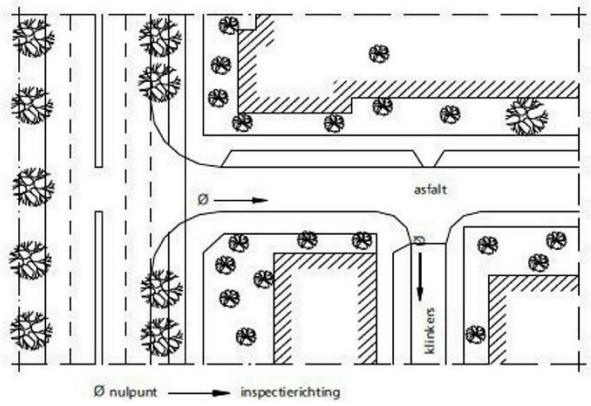


Figure 1. Zero recording

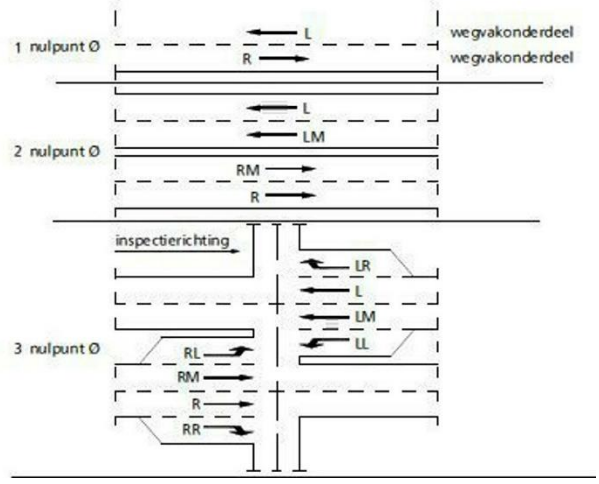


Figure 2. Examples of a strip layout in left, right, right and so on

The zero point and the boundaries of the road sections are precisely and unambiguously recorded in the data file, as well as the further indication of the road section part in the cross-section (see figure 2 and figure 3). If the inspector finds during the inspection that the data file is incorrect, he makes a comment about this.

Weather influences

The inspections must be carried out on a dry road surface. The weather influences during the inspection play a major role in recognizing the damage:

- With a dry road surface, most damage can be clearly distinguished.
- The light on the road must be taken into account. For example, if the inspection is carried out against the sun, many damages are more visible.
- On a wet or damp road surface there is a water film on the pavement, making inspection for asphalt and cement concrete pavements impossible. Crack formation, fraying and deterioration can then hardly be observed.
- With asphalt and cement concrete pavements, a drying road surface gives a distorted picture of the damage, because the cracking appears to be more serious than it actually is.
- Puddling makes inspection difficult.

Frequency and period

The minimum frequency of the global inspection is set at once every two years. To obtain results that can be compared with each other, it is desirable to always carry out this inspection in corresponding seasons. That is, in the period from March to October. It is not recommended to perform the global inspection in the winter months (November to February) due to poor inspection conditions (moisture, frost, leaf fall, etc.).

It is recommended to carry out the small maintenance inspection a number of times a year. The detailed inspection is not carried out with a fixed frequency, but according to the manager's needs.

Inspection equipment
The following inspection equipment is required:

- 'Global visual inspection manual';
- (digital) inspection forms;
- row, 1.20 m length with spirit level;
- wedge (with scale);
- (metal) pins of 3, 5, 10, 15 and 20 mm diameter;
- measuring wheel;
- measuring rod;
- safety vest or jacket;
- possibly a photo camera;
- 'ROAD INSPECTION' sign, clearly visible on the car;
- orange flashing light;
- any permits and exemptions;
- road barrier or arrow wagon, depending on the traffic situation.

Given the different nature of the various inspection forms, not all inspection requirements mentioned will be used for every inspection.

Figure 3. Inspection equipment

Implementation rules

The method of conducting the inspection depends on the type of inspection:

- The global inspection must be carried out by a team of two inspectors from a moving car or by an inspector on foot. Under certain circumstances (including city centers and cycle paths), inspection on foot or by bicycle is desirable or even necessary. During the inspection, the inspector will regularly have to get out of the car to determine the normative severity of the damage, or to observe the damage on a partially hidden footpath or cycle path. In order to obtain reliable inspection data, it is essential that the inspector measure the observed damage occasionally (with the use of a wedge and wedge).
- The detailed inspection must be carried out on foot by two inspectors. During the inspection, the damage is signed on the form. At the end of the inspected road section, the total amounts of damage found are entered on the form. The notation of the damage per road section is always done from the zero point of the road section layout.

This manual, in the form of a step-by-step plan, specifies precisely how the inspectors must assess and record the damage during the global and detailed inspection respectively. If a discussion arises or uncertainties arise during the inspection, then the 'Visual inspection manual' or the 'Global visual inspection manual' must be checked and the prescription applied correctly.

Day production and working hours

The daily production depends on the maintenance condition of the pavements and on the working hours that have been observed. It is recommended not to inspect more than eight hours a day including a one-hour break. Table 1 provides guidelines for daily production during the global inspection. The daily production during the small maintenance inspection is comparable to the daily production of the global inspection. There are no general guidelines for a daily production for the detailed inspection.

Table 1. Guidelines for daily production for global inspection

	The maximum length of the road section	Day production
Outside the built-up area	1000 m	20 - 40 km
In residential area	500 m	10 - 20 km

During the global inspection and the minor maintenance inspection, it is useful to establish a sequence (route) for the road sections prior to the inspection. A guideline for this can be a classification per neighborhood. Account must be taken of one-way roads. In this way the driving distance and the loss of time can be limited as much as possible.

Safety

The applicable safety regulations and health and safety legislation apply to all types of inspection. In this context, it is mandatory to wear a safety vest (this also applies if the inspector is in the car). During the global inspection and the minor maintenance inspection, the inspector will regularly have to get out of the car to see the road, which can be dangerous. In some cases, it may be necessary to inspect with a roadblock.

It is recommended to inquire about the local regulations and ensure that the necessary exemptions are present prior to the inspection.

1.3 Inspector

Preface

In addition to the quality of the manual and the method of implementation, it is primarily the human factor, the inspector himself, that is important for the quality of the inspection. Important aspects are the attitude, education, and experience of the inspector.

Inspector

Inspectors must be at least of a technical MBO level. When outsourcing a global inspection, the client may require that the inspectors are in possession of the 'Visual Inspector Roads' diploma.

Proper implementation of the inspection requires an accurate attitude and the ability to work in a long-term focused manner. By having two inspectors carry out the inspection, a discussion arises between the inspectors about the damage assessment and they can keep each other sharp. This improves the quality of the inspection. In particular, during the global inspection, the inspector must be able to detach himself from any maintenance measures that may be taken. Extensive experience in the execution of road works may not affect the objectivity of the inspector. Although the institution of the inspector is therefore extremely important, periodic training remains necessary.

Figure 4. An inspector

Table 2. Relevant damage groups and damage during the global inspection

Damage group	**Damage**		
	Asfaltbetonverhardingen	*Elementenverhardingen*	*Cementbetonverhardingen*
Texture	Rafeling		
flatness	Dwarsonvlakheid	Dwarsonvlakheid	
	Bumps	Bumps	Bumps
Consistency	Cracking		Cracking
Watertightness			Joint fill

Training and education

In order to obtain a clear assessment of the damage during inspection, clear descriptions of each damage are given in the damage catalog. Every inspector must follow the damage catalog. Only then is it possible to arrive at a clear assessment of the damage. Consistency in the assessment of damage is increased by obtaining the 'Visual Inspector Roads' diploma, which is issued by CROW. This diploma can be obtained by taking a theory exam followed by a practical exam. During the theory exam, the theoretical knowledge of both the detailed and the global inspection is tested. If the inspector has passed the theory test, he will receive a certificate as proof that he has mastered the theory. He can then take the practical exam to obtain the diploma. During the practical exam (aptitude test), the practical skills for the global inspection of various types of pavement are tested. The practical exam is specifically designed for the use and testing of the CROW system of the global inspection. The accuracy of the inspection results is checked during the practical exam. The 'Visual Inspector Roads' diploma can only be obtained successfully if the participating inspector has sufficient theoretical knowledge of visual inspection, has received sufficient education and training and has sufficient experience and insight. During the practical exam (aptitude test), the practical skills for the global inspection of various types of pavement are tested. The practical exam is specifically designed for the use and testing of the CROW system of the global inspection. The accuracy of the inspection results is checked during the practical exam. The 'Visual Inspector Roads' diploma can only be obtained successfully if the participating inspector has sufficient theoretical knowledge of visual inspection, has received sufficient education and training and has sufficient experience and insight. During the practical exam (aptitude test), the practical skills for the global inspection of various types of pavement are tested. The practical exam is specifically designed for the use and testing of the CROW system of the global inspection. The accuracy of the inspection results is checked during the practical exam. The 'Visual Inspector Roads' diploma can only be obtained successfully if the participating inspector has sufficient theoretical knowledge of visual inspection, has received sufficient education and training and has sufficient experience and insight.

Deployment of an inspector with the 'Visual Inspector Roads' diploma is a step towards an accurate objective inspection, but offers no guarantee that the inspection result will be beyond doubt. Imposed high inspection rate and daily production, but also traffic conditions can always have a negative effect on the inspection results. Therefore, sufficient attention must also be paid to these aspects before and during inspections.

Continuous attention to training also remains important for inspectors who hold the 'Visual Inspector Roads' diploma. Experienced inspectors will eventually develop their own interpretation of damage and an associated assessment. Every inspector must consult the damage catalog periodically. Consistency in the assessment is increased by following courses, but also by having the same road sections assessed by different inspectors and changing inspection teams *(training on the job).* For more information about the courses that CROW provides, the CROW website can be consulted.

2 Inspection method

The global inspection is described in more detail in this chapter. The following aspects are discussed in succession:

- damage groups and damage;
- explanation of forms;
- step-by-step assessment of damage claims;
- minor maintenance;
- example of the inspection form.

Damage groups and damage

Table 2 shows the **damage groups** relevant for the global inspection and the associated damage.
Edge damage, joint width and settlement are not standard part of the inspection, but can be involved by the manager as an option at the discretion of the manager. This exception does not apply to other claims. Drainage problems are noted as a comment during the inspection if they give rise to nuisance or dangerous situations.

Table 2. Relevant damage groups and damage during the global inspection

	Damage		
Damage group	*Asfaltbetonverhardingen*	*Elementenverhardingen*	*Cementbetonverhardingen*
Texture	Rafeling		
flatness	Dwarsonvlakheid	Dwarsonvlakheid	
	Bumps	Bumps	Bumps
Consistency	Cracking		Cracking
Watertightness			Joint fill

In addition, there is the possibility during the inspection to indicate the need for minor maintenance and to make comments, for example, the wish to have measurements taken.

Damages can occur in mutual relationships, for example, unevenness with transverse unevenness or cracking with transverse unevenness. In that case, every damage is assessed and noted separately.
In the damage catalog, the following aspects are described for each type of pavement per damage to be inspected:

- definition;
- severity and scope;
- explanation.

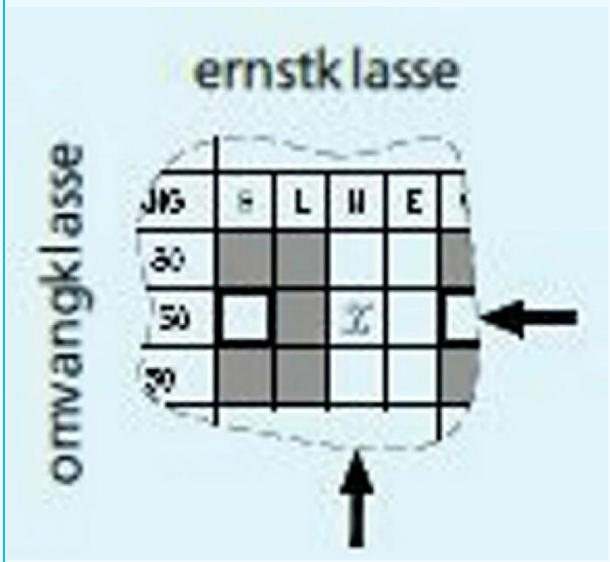
Explanation of the forms

Two forms are available for performing the global inspection (see Annex I):

- asfaltbeton / elementenverhardingen;
- asfaltbeton / elementenverhardingen

The forms have a general information part and an inspection part. The interpretation of the inspection part is described in a step-by-step plan (see table 3).

Table 1. Global inspection step-by-step plan

Step	Action	Explanation
1	For each damage, determine the normative severity class (E> M> L) or establish that no significant damage or no significant damage is detected (G).	This is the highest common severity class. If serious damage (E) is noted as minor maintenance (see Table 4), the normative severity class becomes the next present lighter severity class (ie M or L): E = serious, M = moderate, L = light. In the tables in which the severity of the damage is defined, different criteria are included for a number of damage per severity class, separated by dashes of thought. In those cases, the rule applies that if one of the criteria is met, the damage then falls into the relevant class. If criteria of different severity classes are met, the highest severity class applies. G: no damage or no significant damage.
2	Determine for each damage the extent in area, length or number of the normative severity class.	Bear in mind that serious damage sometimes has to be noted as minor maintenance (see step 5 and table 4).
3	Determine the size class of the determining damage per damage.	The extent of fraying and joint width is determined based on the total surface area of the road section. The extent of all other claims is expressed per 100 m section of road section.
4	Mark the box on the inspection form that corresponds to the determined severity and scope class. If no damage or no significant damage is found, the leftmost box must be ticked (G).	
5	Indicate separately for each damage group whether small maintenance is necessary, by placing an 'X' in the relevant box.	Only the need for minor maintenance is indicated; not the type of minor maintenance or size. For a description of minor maintenance, see table 4. Serious damage noted under minor maintenance may no longer be noted as indicated in step 4.
6	Write down any comments under comments.	Enter additional information or add text on the form.

The following must be entered in the general information part of the forms:

a. road name;
b. route number;
c. road section number;
d. from ... to ... 'information';
e. length of road section;
f. road section;
g. paving type;
h. surface area of road section;
i. length of road section;
j. date;
k. observers (inspectors);
l. weather conditions.

Parts a up to and including i come from the database. The parts j, k and l speak for themselves. The following abbreviations are used as an example for the format of the road section:

- RB = roadway;
- FP = cycle path;
- VP = footpath.

Space has been left on the form for any deviating names (for example: PS = parking lane, US = escape lane, WE = residential area).

The following abbreviations are used as an example for the position of the road section part in the cross section of the road section:

- L = links;
- R = right.

Here too there is some room to indicate deviating indications (for example: M = middle). For a further explanation of the coding of the road section, reference is made to Figure 2.

Step-by-step plan for damage assessment
Table 3 indicates the steps by which the assessment of each damage must be carried out and noted during the global inspection. If the road section consists of several road section components (roadway on the left, roadway on the right, footpath, cycle path, etc.), then the steps in table 3 must be completed separately for each road section.

Minor maintenance
When encountering serious damage in a size that falls in the smallest size class (very small), minor maintenance is ticked (see Table 3). The indicated amount in table 4 is a maximum per road section and can be distributed over a few places.

Table 4. Relationship repair measure - damage (quantities based on a road section length of 500 m)

Repair measure	Quantity	Unit	Damage / Damage group
Asphaltic			
emulsion treatment	50	m	
area surface treatment	50	^{2}	
last corpse Slemmen	50		
holes fill	all		
Deposits	25	m	
Micro coating / track pad	15		

Justify hot or cold	all	m^2	fraying and / or cracking transverse
Crack Filling	all		unevenness transverse unevenness / irregularities
Edge Damage repair	15	m	cracking / welding
		m^2	cohesion
		pcs	
		m	
		m^2	
		m	
		stuks	
		m	
		m	
Elements			
Resurrected	50	m	
locally	25	m^2	dwarsonvlakheid / irregularities
		m	
		m^2	
Cement Concrete			
Partial plate replaced	3	pieces plates	
Replacing Entire sheet	1	pieces plates	
repair erosion	2	pieces plates	unevenness
to splashing repair	incidentally	pieces	
plate angle repair	2	pieces	
Staples / verdeuvelen	2	pieces	cracks
Grouting application	incidentally	m	
Crack filling application	all	m	joint fill
Cracks looked up / filling	2	pieces plates	

When checking minor maintenance, the inspector assumes that the repair or conservation will be carried out in the current financial year. The serious damage is therefore eliminated. The next damage present (moderate or light) is noted as damage in addition to minor maintenance. If there is no other damage in the road section, no damage is entered in the box for the damage in question.

Appendix B

Cartesius: the Dutch supercomputer

Cartesius is a bullx system extended with one Bull sequana cell. It is a clustered SMP (Symmetric Multiprocessing) system built by Atos/Bull. Cartesius consists of a large number of batch nodes and a small number of special-purpose nodes. For the batch nodes we differentiate between so-called thin nodes, fat nodes, GPU nodes, and Xeon Phi nodes. Thin nodes constitute the majority of the available batch nodes. GPU nodes use GPGPUs to accelerate the computations. The Xeon Phi nodes have Intel Many-Core CPUs. Fat nodes have more memory (256 GB) than the thin nodes and more physical cores (32) than most thin nodes.

Special nodes are the two interactive front end nodes and five so-called service nodes. The interactive front end nodes provide the ssh login service and are meant for interactive work. The service nodes are meant for data staging (transfers to and/or from external contexts, such as the SURFsara archive facility or a remote site.).

Currently, Cartesius consists of the following compute nodes:

Node Type	Number	Cores	CPU	Clock	Memory
broadwell	177	32	E5-2697A v4	2.6 GHz	64 GB
thin	1080	24	E5-2690 v3	2.6 GHz	64 GB
thin	540	24	E5-2695 v2	2.4 GHz	64 GB
fat	32	32	E5-4650	2.7 GHz	256 GB
gpu	64	16	E5-2450 v2	2.5 GHz	96 GB
knl	18	64	7230	1.3 GHz	96 GB

Nodes Summary

This totals up to:

47,776 cores + 132 GPUs: 1.843 Pflop/s (theoretical peak performance):

25,920 Haswell cores (thin nodes): 1.078 Pflop/s

12,960 Ivy Bridge cores (thin nodes): 249 Tflop/s

5,664 Broadwell cores (thin nodes): 236 Tflop/s

1,056 Ivy Bridge cores + 132 K40m GPUs (GPGPU nodes): 210 Tflop/s

1,152 Knights Landing cores: 48 Tflop/s

1,024 Sandy Bridge cores (fat nodes): 22 Tflop/s

130 TB memory (CPU + GPGPU + HBM)

Interconnect

For the batch nodes we have introduced the concept of islands. Inside the bullx islands, the interconnect has a fully non-blocking fat-tree topology. Inside the sequana cell, the interconnect has a pruning factor of 2 : 1. Between islands, there is a pruning factor of 3.33 : 1 and the inter-island latency is 3 µs.

Every bullx node has a Mellanox ConnectX-3 or Connect-IB (Haswell thin nodes) InfiniBand adapter providing 4 × FDR (Fourteen Data Rate) resulting in 56 Gbit/s inter-node bandwidth, The GPGPU nodes have two ConnectX-3 InfiniBand adapters: one per GPGPU.

Every sequana node has a Mellanox ConnectX-4 InfiniBand adapter providing 4 × EDR (Enhanced Data Rate) resulting in 100 Gbit/s inter-node bandwidth.

Disk Space

We have 180 TB for home file systems and we have some 7.7 PB Lustre (parallel) file systems for scratch and project space.

Operating System

Each node runs under one operating system (bullx Linux, which is a Linux distribution compatible with Red Hat Enterprise Linux) and shows a single memory image to the user programs