THE DEVELOPMENT OF A DIAGNOSTIC APPROACH TO PREDICTING THE PROBABILITY OF ROAD PAVEMENT FAILURE

by

MEGAN ROSE SCHLOTJES

A Thesis submitted to The University of Auckland and University of Birmingham
for the degree of DOCTOR OF PHILOSOPHY

Department of Civil and Environmental Engineering
The University of Auckland

And

School of Civil Engineering
University of Birmingham

2013

PREFACE

The research reported in this thesis was jointly supervised by The University of Auckland and the University of Birmingham under the *Universitas 21* joint PhD scheme. The research, undertaken between March 2009 and November 2012, was primarily carried out at The University of Auckland. The University of Birmingham hosted the researcher for the period between May 2010 and August 2011 and for a subsequent research visit in May 2012 to June 2012.

The contents of this thesis are the original work of the researcher, except where appropriately acknowledged in the text, and no part of this thesis has been submitted for a degree at another institution.

Megan Schlotjes

ABSTRACT

Road maintenance planning, an essential component of road asset management, preserves the integrity of road networks. Current state of the art pavement management systems exercise optimisation tools, pavement deterioration models, and intervention criteria to forecast the future maintenance requirements of a road network. These tools have been utilised to forecast future maintenance requirements of road networks; however, with this current approach to pavement management, uncertainties associated with the failure of individual sections of road may not always be accounted for explicitly, and therefore the susceptibility of a road network to failure is unknown.

Predicting the probability of the end of life of a road pavement involves wholly understanding possible modes of failure and utilising suitable computational techniques, so that engineering knowledge can be well represented in data driven models. To this end, this thesis describes the development of a diagnostic approach that infers **engineering knowledge** into **computational models**, to quantify the probability of failure of road pavements and identify the most likely causes of failure.

To do so, this research developed a number of failure charts that capture engineering knowledge, such as citing influential failure factors of road pavements including the influence from external environments and internal pavement attributes. Engineering knowledge on road pavement failure was obtained from three sources: literature describing the fundamentals of pavement design and common causes of road failure, expert knowledge from the industry identifying relationships between failure mechanisms and causes, and a data analysis to obtain

site-specific causes such as road environments and material properties. Each chart presents a possible failure path, detailing a set of factors contributing to failure.

A comparative study evaluated the performance of five classification modelling approaches in order to determine the most suitable technique for this research. Based on performance and user interpretability criteria, the study identified one based on support vector machines as the most suitable.

The developed prototype system, consisting of a failure system for rutting, fatigue cracking, and shear, performed well in both the development phase and network testing of the system utilising data from the New Zealand Long-term Pavement Performance Programme. A case study focussing on rural New Zealand roads was carried out, which demonstrated the use of this tool in network and project level applications.

ACKNOWLEDGEMENTS

I, as the researcher, would like to sincerely thank my supervisors, Dr. Theunis F.P. Henning (The University of Auckland) and Dr. Michael P.N. Burrow (University of Birmingham), for whom without the success of this research would not have been possible. I am appreciative of their time, inspiration and support through both the highs and lows of the PhD journey. I further acknowledge the input from my co-supervisors, Dr. John St. George (The University of Auckland) and Dr. Harry Evdorides (University of Birmingham).

To my family, I thank you for your support applauding encouragement, despite the misconception over what exactly it is I do. To my Mum in particular, I thank you for your constant support, reassurance and love over the years. Without you, the journey would not have been enjoyable, nor as fulfilling as it has been.

The valuable input in various stages of the research from Noel Welsh, Fritz Jooste and Derek Roux is gratefully acknowledged. Their invaluable knowledge and experience got me out of some very sticky situations. I would also like to thank the asset managers from Southland District Council, New Zealand, for their time and input in this research.

Finally, without the financial support from the New Zealand Transport Agency and The University of Auckland, I would not have been fortunate to experience as many opportunities as I did throughout my PhD journey.

"Ehara taku toa i te toa takitahi Engari, he toa takitini"

"Success is not the work of one, but the work of many"

Maori Proverb

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

GLOSSARY OF TERMS

AADT	Annual average daily traffic.
Basecourse layer (base layer)	The aggregate layer beneath the wearing surface, which dissipates the induced forces from traffic loadings prior to reaching the subgrade.
Binary	The state of an output which can only be either zero (0) or one (1).
Bituminous layer (surface layer)	The wearing course of the pavement, consisting of a bitumen layer above the aggregate layer(s).
Chipseal (asphalt) surface	A cost effective surface dressing, consisting of one or more layers of asphalt and aggregates.
Classification	A modelling approach where the output is categorical.
Confusion (performance) matrix	A matrix where the failure predictions are categorised against the actual reported occurrence of failure in the dataset. These categories are further used in the calculations of the performance measures and evaluations of techniques.
Critical (most probable) failure path	The failure path associated with the maximum predicted probability from the prototype system, as per failure mechanism, if all failure paths and subsequently the causes of failure are considered equally critical. The terms *"critical"* and *"most probable"* are interchangeable.
Cross-validation	A sampling method employed in performance evaluations of computational models, when there is insufficient data provided in the dataset for an independent evaluation dataset.
Cum_ESA	Cumulative number of equivalent standard axles.

CumRain_ifCracked A factor involved in the modelling phase of this research that quantifies the cumulative amount of rainfall on a cracked pavement, intimating the amount of water that easily reaches the lower layers of the pavement through breaks in the surface.

Dependent variable(s) The output(s) from the model, such that the model is dependent on the given input data.

ESA Equivalent standard axles which, as a traffic loading factor, assesses the impacts from all traffic classes, including heavy vehicles.

Excessive traffic Traffic loadings which exceeds that of the design, or an over-loaded heavy vehicle.

Failure chart A collection of failure paths, presented in a tree-like form, depicting the causes of failure.

Failure path A path identified on the developed failure charts, where a combination of factors acts concurrently resulting in failure.

Fatigue cracking failure The analysis in this research uses the percentage of surface cracked (by any crack type), rate of cracking, and number of years continually cracked as the surface condition data in the State Highway LTPP dataset relating to fatigue cracking.

Flexible pavements Pavements constructed with granular or asphalt materials.

FMEA Failure modes and effects analysis.

FTA Fault tree analysis.

FWD Falling weight deflectometer, where the strength of the pavement can be inferred from the FWD measurements.

HCV Percentage of heavy commercial vehicles.

Independent variable(s) Input(s) of a model, such that these variables are used to calculate the output / result..

Input(s) Independent variables used in modelling equations to obtain the desired output(s).

Local Authority road network New Zealand roads which are managed by Local Authorities (road controlling authorities), such as local and district councils. These roads are not included in the main trunk road network.

Low volume road Defined in this research as roads that carry a traffic loading of less than 10,000 vehicles per day.

LTPP programme Long-term pavement performance programme.

NIWA National Institute of Water and Atmospheric Research, which is a Crown Research Institute focussed around environmental science and sustainable management in New Zealand.

NZTA New Zealand Transport Agency, who manages the State Highway road network in New Zealand.

Output(s) Dependent variable(s) calculated by the model using established input(s).

PMSs Pavement management systems.

RAMM Road assessment and maintenance management database where road controlling authorities retain road inventory and condition data.

RCAs Road controlling authorities, who manages the local roads in New Zealand.

Rutting failure The analysis in this research uses the rutting depth and rate of rutting as the surface condition data in the State Highway LTPP dataset relating to rutting.

Shear failure	The analysis in this research uses pothole information, shoving, structural patches, and mechanical damage as the surface condition data in the State Highway LTPP dataset relating to shear.
SNP	Modified structural number.
State Highway road network	The main trunk road network of New Zealand that runs the length of the country facilitating the movement of people and goods around the country. This network is managed by NZTA.
Sub-base layer	An additional pavement layer of aggregates, which lies beneath the basecourse layer. This sits above the subgrade.
Subgrade	The underlying ground / soil / foundation of the pavement.
Successful factor combinations	The combinations of failure factors identified on the failure charts as possible failure paths.
Support vector machines	A classification model which employs kernels to transform data into a feature space for a simplified separation of data classes.

Chapter One

INTRODUCTION

1.1 The Context of Predicting Road Failure

Maintenance planning is imperative for preserving the integrity of any road network and consequently an essential component of managing road pavements. Such planning relies heavily on a collaboration of historical network trends and inventory data, current condition reports and field inspections, predicted performance of road pavements, and sound engineering judgment and first principles.

Predicting the likelihood of road pavement failure and the associated causes is an important aspect of maintenance planning. However, the causes of pavement failure are often complex in nature and can be attributed to a number of factors including traffic loadings and environmental conditions exceeding design loading, deficiencies in pavement composition and strength, a weak subgrade, and poor construction practices. Such combinations, because of their complexity, are often overlooked in the diagnostic process despite the comprehensive road reporting process. Furthermore, where road condition surveys rely on visual inspections only, such as the case for many low volume roads that make up the majority of any country's road network, potential failures may not be recognised because, in these cases, visual inspections commonly report only minor deterioration and disregard any obscured structural problems. Although the data and knowledge associated with maintenance planning, is widely available, its significance is often undervalued.

New Zealand, alike to many countries, faces financial shortfalls in road maintenance funding and currently its rate of increase in shortfall is greater than the rate of network growth and inflation (Schlotjes et al., 2009). Given the financial pressures faced by road controlling authorities (RCAs), the attitude towards asset management is shifting to a *'do more for less'* approach, whereby decreasing the investment levels results in a reduced level of service across the network. Under such an approach, the maintenance of roads of lower importance and hierarchy in the network are often disregarded from the forecasted maintenance schedule through deferring required maintenance and ignoring unexpected failures (SATCC, 2003). However, effective management of road networks requires an evaluation of the consequences from deferring maintenance and of ignoring unexpected failures, such that the asset manager can appreciate the associated risks resulting from these actions.

The case study, which forms much of the basis of this thesis, focuses on the greater New Zealand road network, where the majority of the network consists of aged pavements carrying low volumes of traffic (less than 10,000 vehicles per day). These pavements display little visual deterioration over time, often failing unexpectedly (Arampamoorthy and Patrick, 2010; Bailey et al., 2006; Henning and Roux, 2008), and based on current practices, are generally excluded from the forecasted maintenance schedules (Schlotjes and Henning, 2012). Changes in the road environments, such as increasing weight and axle allowances of heavy vehicles, increase the potential occurrence of unexpected failures across the network (Schlotjes and Henning, 2012). From a practical viewpoint, unexpected failures have the following implications to the RCAs:

- Unplanned maintenance, excluding isolated events, is not cost effective, and

- An increase in the variability in the forecasted maintenance schedule and costs leads to inefficiencies in the current reporting practice and maintenance planning.

Currently, the forecasted maintenance schedules advise **what** maintenance treatments are required, **where** such maintenance is needed, and **when** it should be implemented. Such schedules are determined using a combination of optimisation processes, predictions of the future state, expert opinion(s), and intervention criteria (Henning, 2008; Schlotjes et al., 2013a). However, current practices do not:

- Fully assess the impact of unexpected failures on the performance of road networks with more accuracy and efficiency than current practices;
- Measure the shift in the failure probabilities and probability profiles over time;
- Quantify the consequential effects of maintenance deferral on the integrity of road networks as network data analysis, particularly data trends, do not necessarily assess this;
- Evaluate the vulnerability of networks to changes in loading and environmental conditions, and
- Automate the process of identifying the causes of failure to assist in the maintenance decision-making processes.

1.2 Problem Statement

A large amount of research has been carried out to develop models of road pavement deterioration. Using these models, asset managers are able to forecast the expected life of pavements and network maintenance requirements. However, a robust methodology for an evaluation and assessment of the probability of road pavement failure, given its associated environmental conditions, is lacking from pavement management systems (PMSs) (Salt et al., 2010).

A number, and combination, of causes and physical attributes of pavements contribute to the failure of road pavements. However, the interactions between these causes and environmental conditions are lacking in the literature. These interactions are difficult to represent with a computational model for the following reasons (Reigle, 2000):

- There are numerous factors affecting pavement failure;
- The uncertainty of the behaviour of pavements under loading and environmental conditions, and
- Errors associated with condition monitoring and variation included in the current reporting processes.

Such methodologies have been successfully established for predicting failure probabilities of other infrastructure assets, including water networks. Therefore, this research proposes a methodology to:

- Develop a prototype system to predict the probability (likelihood) of road pavement failure;
- Infer engineering knowledge into the development of the computational model;
- Compare and contrast the effectiveness of several computational modelling techniques for evaluating the probability of road pavement failure from road datasets, and
- Identify the most probable causes of failure to facilitate appropriate road maintenance recommendations.

1.3 Aim, Objectives, and Scope of the Research

1.3.1 Aim of the Research

The aim of this research is to develop a framework that develops a prototype system capable of quantifying the probability (likelihood) of road pavement failure and diagnosing the most probable causes of failure.

1.3.2 Objectives of the Research

To achieve the above aim, this research has the following objectives:

1. Understand the mechanisms of structural road pavement failure by capturing engineering judgment in failure charts, and define the factors contributing to failure;
2. Identify modelling techniques that can be used to predict the likelihood of failure from road datasets and, from a comparative study, select the most appropriate for the task at hand;
3. Develop a prototype system based on the outcomes from the previous objectives, which will assess the likelihood of failure for a given road site, and
4. Demonstrate the effectiveness of the methodology and modelling technique on an independent New Zealand road dataset.

1.3.3 Scope of the Research

This research seeks to develop a robust and universal methodology that can be used to assess the probability of road pavement failure, which is applicable to other pavement types and road networks. However, the case study for this research focuses on New Zealand roads and, as a result, Figure 1-1 identifies the road types addressed in this research. At least 95 % of the New Zealand road network consists of flexible unbound pavements (Hayward, 2006), as it is uneconomical for New Zealand to construct stronger, rigid pavements. The majority of these roads are thinly designed based on the layer composition shown in Figure 1-2. The typical design life for these pavement types is 25 years (Bailey et al., 2006), although Arampamoorthy and Patrick (2010) reported the actual life of such pavements can be upwards of 180 % of the design life.

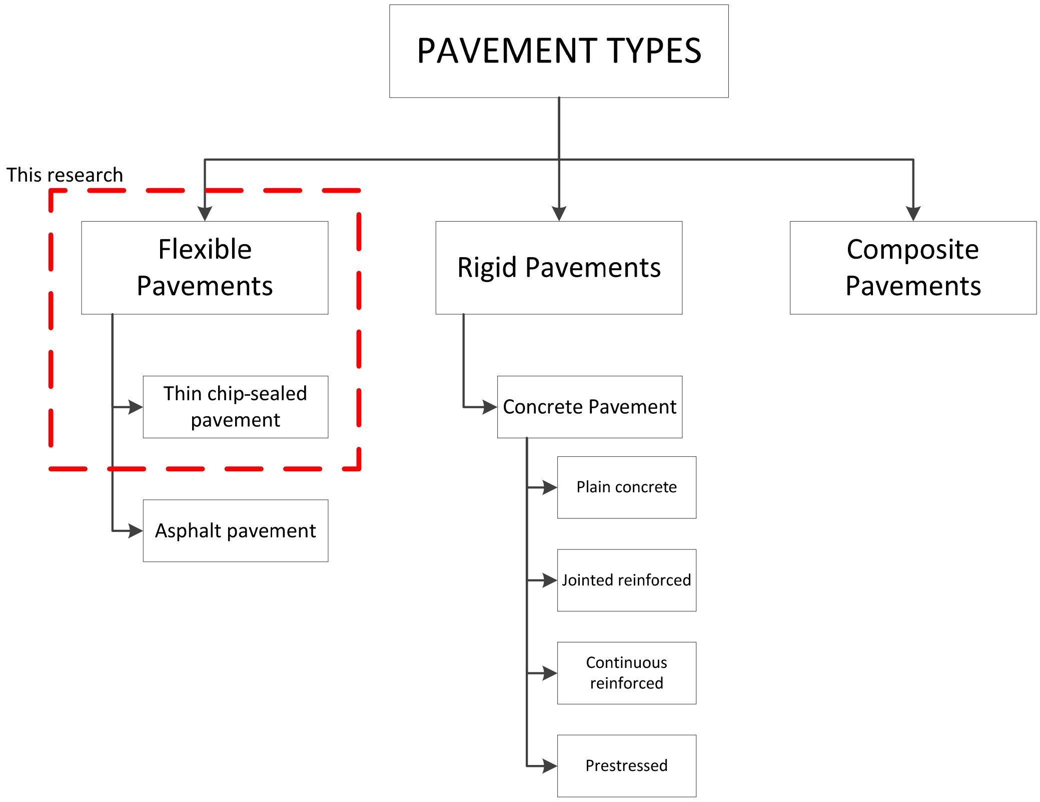


Figure 1-1: Overview of pavement types

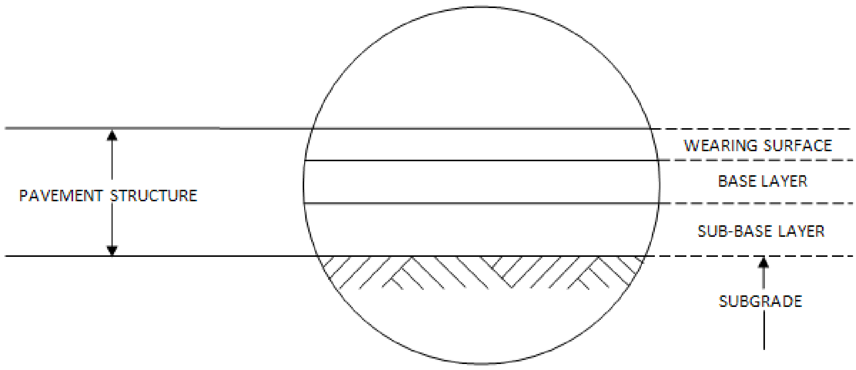


Figure 1-2: Flexible pavement road structure

Managed by the New Zealand Transport Agency (NZTA), close to 100 % of the national strategic road network of New Zealand, the State Highway network, is sealed. Only 61 % of the local roads, managed by the Local Authorities, are sealed. Chipseal surfaces for this research include single coat seals, two coat seals, void fill seals, slurry seals, racked in seals, and texturising seals.

The geography of New Zealand has resulted in variable climatic conditions and road environments, typically experiencing warm dry summer months and cold wet winters. The variation in temperature between the seasons is not as severe as other countries, yet the regions experience variable rainfall and temperatures over the seasons.

The case study of this research focuses on New Zealand road networks; therefore, the scope of this research is as follows:

- **Pavement types:** Thin, flexible, unbound granular pavements, and / or lightly stabilised;
- **Surface types**: Chipseal surfaces;
- **Traffic volumes:** Up to 10,000 vehicles per day, and
- **Failure mechanisms:** Rutting, cracking, and shear.

1.4 Structure of the Thesis

To achieve the objectives listed above, the thesis is structured as shown in Figure 1-3.

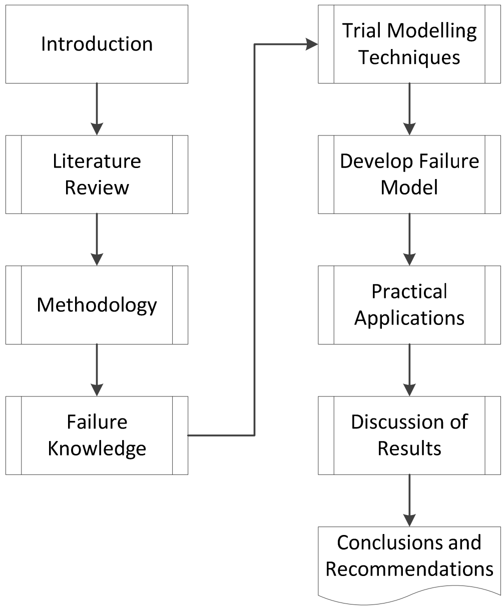


Figure 1-3: Overview of the thesis

1. **Chapter Two** reviews the predominant failure mechanisms on New Zealand roads, modelling approaches previously used in deterioration modelling and performance assessments of infrastructure assets including pavements, and further possible classification modelling techniques used in a range of industries;
2. **Chapter Three** presents the framework followed in this research to develop the prototype system. The generic nature of this chapter allows the transferability of this work to other studies;
3. An understanding of road pavement failure is developed in **Chapter Four**, which presents the causes of each failure mechanism with the use of failure charts. This engineering knowledge is used in the computational models developed in subsequent chapters;

4. **Chapter Five** presents a comparative study of a number of classification modelling techniques identified in Chapter Two, and in conclusion, establishes the most appropriate technique to further develop for the specificities of the research dataset;
5. The most appropriate modelling technique and developed failure charts are used in the development of the prototype system, as presented in **Chapter Six**;
6. An independent road network is analysed using the developed prototype system in **Chapter Seven**. The case study demonstrates the network and project level applications of the developed pavement performance tool;
7. A critical review of the research including the methodology adopted, assumptions of the research, and the effectiveness of the prototype system is presented in **Chapter Eight**, and
8. **Chapter Nine** concludes the thesis by presenting the accomplished work, the conclusions of the research, and recommendations for further work.

Chapter Two

LITERATURE REVIEW

2.1 Introduction

An outcome of this research, and consequently the research framework, is a computational model that is capable of predicting the probability of road pavement failure from road datasets. The computational model will make use of engineering knowledge of pavement failure to ensure that correct model independent variables are chosen. Therefore, to facilitate this research, this chapter reviews current knowledge in the following areas:

1. Failure mechanisms of interest to this research;
2. Methods of inferring failure knowledge into computational models;
3. Performance modelling of infrastructure assets, including pavement performance modelling, and
4. Classification modelling techniques available to predict the probability of pavement failure.

The pertinent information from the review is then utilised to develop a theoretical framework of the proposed system.

2.2 Background to Road Pavement Failure in New Zealand

A large number of defects are associated with road pavement failure, depending on the pavement types, the construction of road pavements, loading environments, and underlying subgrade properties. However, since the New Zealand case study focuses on flexible, unbound, granular pavements, this section reviews the predominant types of structural failures prevalent on New Zealand road networks comprising of these pavement compositions, namely rutting, cracking, and shear failure (Arampamoorthy and Patrick, 2010; Creagh, 2005; Gribble and Patrick, 2008). Although roughness is an observed failure type, it is predominantly regarded as a functional (and not structural) defect and, therefore, was not considered further herein (Arampamoorthy and Patrick, 2010; Creagh, 2005; Robinson, 2008); instead Austroads (2007a) explains in detail the roughness failure mechanism, which can be considered to be a measure of the combined effects of all of the other road surface failure types, including rutting, cracking and shear failure types. While the subject of optimising maintenance treatments is beyond the scope of this research, a comprehensive treatment of the following failure types is given by Thom (2008).

2.2.1 Rutting Failure

Ruts appear in the wheelpaths of the pavement as longitudinal depressions resulting from structural deformation of the layers below the wheel loadings (Papagiannakis and Masad, 2008). The occurrence of rutting can indicate the deterioration of the structural integrity of the pavement and pavement deficiency. If the ruts are wide and evenly-shaped, the depressions are caused by weakness in the lower layers of the pavement. If the ruts are narrow and sharply defined, the upper pavement layers are deficient in dissipating the induced traffic loadings through the pavement (CSRA, 1992). Austroads (2007b) defines ruts with a length-to-width ratio of at least 4:1.

These depressions pose a safety hazard to the road user allowing water ponding to occur in the surface ruts with the potential of aquaplaning and black ice[1] hazards. Figure 2-1 shows the appearance of rutting on a New Zealand road section. Recently, the noticeable increase in accelerated rutting[2] has become a problem on New Zealand roads (Henning et al., 2007). Furthermore, the presence of fatigue cracking within the surface ruts is also noticeable on New Zealand road networks, where it commonly occurs in pavements consisting of thin bituminous layers (CSRA, 1992).

Figure 2-1: Rutting in the left-hand wheelpath

Rutting is accounted for in the Austroads pavement design using Equation 2-1 (Austroads, 2012), where the allowable number of Equivalent Standard Axles (ESA) loadings (N) is calculated from the maximum compressive strain allowed at the top of the subgrade ($\mu\varepsilon$).

---

[1] Transparent ice on the wearing surface of the road resulting in a slick road surface, which poses an inconspicuous hazard to drivers.

[2] Defined by Henning et al. (2007) as the rapid deterioration phase of rut progression towards the end of its design life, often once failure has been initiated.

Equation 2-1

$$N = \left[\frac{9300}{\mu\varepsilon}\right]^7$$

Henning (2008) adapted the World Bank Highway Development and Management Model (HDM-IV) to replicate the performance of New Zealand road pavements, focussing on the progression of rutting. In this study, Henning (2008) identified in the proposed rut progression model a period of accelerated rutting on pavements and suggested the progression of rutting is swifter for thinner pavements than thicker pavements, under the same traffic loadings conditions.

2.2.2 Cracking Failure

Cracking is defined as breaks in the integrity of the pavement surface indicating vertical rupturing of the pavement, which may not necessarily extend through the entire thickness of the surface or pavement layer(s) (Austroads, 2006). Different types of pavement cracking are usually defined as follows (Henning, 2008; Thom, 2008):

- Fatigue (alligator) cracking, appearing in the wheelpaths of the pavement and resembles the texture of *'alligator skin'*. As it is associated with excessive traffic (wheel) loading, it is also often referred to as load associated cracking;
- Transverse cracking, is an environment associated phenomenon and is due to settlement, freeze thaw, or shrinkage consequently resulting from stabilising the lower pavement layers;
- Longitudinal cracking, occurs as a result of construction joints, settlement or movement in the subgrades;

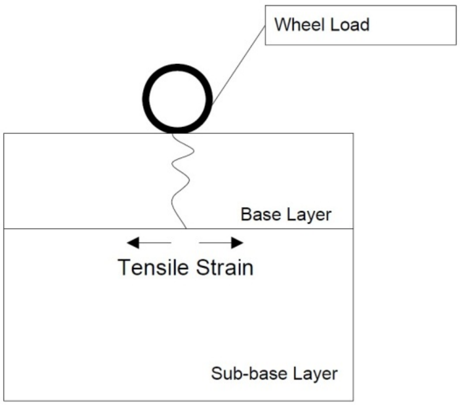
- Block cracking, appears as a block pattern anywhere on the pavement as a result of shrinkage of the stabilised layer;
- Thermal cracking, resulting from cyclic temperatures;
- Edge cracking, is associated with the edge of the sealed pavement cracks, and
- Joint cracking, where the joints of different pavement subsections meet, and it typically occurs in concrete (rigid) pavements.

Surface cracking permits the ingress of water into the lower layers of the pavement, which further exacerbates its progression (Austroads, 2006). As surface cracking progresses, the cracks become interlinked and the surface seal disintegrates under continual traffic loading.

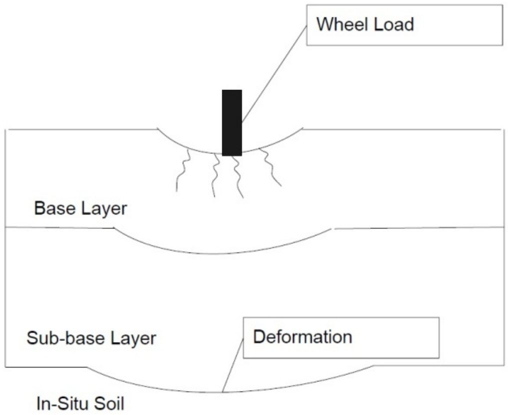
Fatigue cracking, shown in Figure 2-2, is the predominant cracking type associated with structural failure on New Zealand road pavements (Henning, 2008), and largely appears in the wheelpaths, indicating the association of this defect with the induced loading and strain repetitions (CSRA, 1992).

Figure 2-2: Fatigue cracking on sealed pavements in New Zealand

Two crack initiation mechanisms are common in New Zealand. Figure 2-3a presents cracking caused by large tensile strains at the bottom of the base layer where the cracks initiate form beneath the surface. In this case, the design (fatigue) life of the base layer has been exceeded by the current traffic loadings, resulting in a vertical divide in the base layer as the crack further develops and propagates to the surface. Cracking as a result of insufficient pavement support from below the base layer is shown in Figure 2-3b. The subsidence of the lower pavement layers cause cracks to form in the top of the base layer in the wheelpath as the material compresses under the wheel load (Austroads, 2006; CSRA, 1992; Henning, 2008).



(a) Cracking due to base layer performance



(b) Cracking due to inadequate pavement support

Figure 2-3: Cracking mechanisms (Henning, 2008)

Fatigue cracking of an (unstabilised) asphalt layer is designed for using Equation 2-2 (Austroads, 2012), taking into account the volume of binder in the asphalt mix (V_b), elastic (*Young's*) modulus of asphalt (*E*) and the maximum allowable compressive strain at the top of the subgrade ($\mu\varepsilon$) to calculate the design traffic loadings (*N*).

Equation 2-2

$$N = RF\left[\frac{6918 \times (0.856 \times V_b + 1.08)}{E^{0.36} \times \mu\varepsilon}\right]^5$$

$$where: RF = 1\ (desired\ project\ reliability\ 95\ \%)$$

Henning (2008) investigated the initiation of cracking to be used as a detection point in the forecasting of preventative road maintenance intervention for New Zealand. Henning (2008) concluded that previously cracked pavements had a higher chance of cracking again, and pavements with multiple surface layers were less stable than single surface layer pavements due to significant flexing of the surface layers. The developed crack initiation model identified the main drivers of cracking to be:

- Age of the pavement surface;
- Annual number of equivalent standard axles;
- Total surface thickness (including the number of surface layers), and
- Modified structural number.

2.2.3 Shear Failure

Shear failure appears as potholes or deformations on the surface of the pavement resulting from insufficient support of the underlying pavement structure. It is common for this type of failure to occur on or near the edge of the pavement where it is particularly susceptible to weakness from the ingress of seasonal moisture (Emery, 1992; Schlotjes et al., 2009; Thom, 2008). However, shear failures can also be seen across the entire pavement surface as a result of material shear between the surface and pavement layers. Figure 2-4 illustrates a shear failure in the form of a shove near the edge of the pavement.

To the road user, the ride quality diminishes with the presence of potholes and shoves on the pavement surface, as well as posing a safety hazard. These pavement defects allow water to enter the lower layers of the pavement, further adding to the variation of the seasonal moisture zones located under the shoulders and near the edge of the pavement (see Figure 4-2). New Zealand road networks are not normally susceptible to shear failures in periods of dry weather; however, after heavy periods of rainfall, it is common to see a large number of newly formed potholes.

Figure 2-4: Shear failure near the edge of the pavement

There are limited design criteria surrounding shear failures due to the inherent behaviour of material properties, the nature of the failure type, and poor performance in computational models (Schlotjes et al., 2011; Schlotjes et al., 2012b). However, a number of recent studies have been carried out to investigate the occurrence of shear failures on New Zealand roads. Hussain et al. (2011) investigated the effect of varying moisture contents on common New Zealand pavement aggregates, and concluded that the detrimental effect on the performance of the pavement was a result of water ingress in the basecourse layer. Patrick (2009) reported that shear failures were caused by the penetration of water through chipseal surfaces under high traffic volumes. The study concluded that the permeability of the surface layer increases especially in cases where water ponds on the surface after heavy rainfall periods, which explains the increased number of reported potholes after heavy rainfall events.

2.2.4 Multiple Failures

Flexible pavements are conventionally designed to fail by one failure mechanism (Austroads, 2012); however, the manifestation of multiple failures is a common occurrence on New

Zealand road networks, which creates difficulties in determining the primary cause(s) of failure. Multiple failures can result from:

- Two or more failure mechanisms occurring at the same time, or
- Secondary effects, where a secondary failure type is the result of the primary failure mechanism, such that rutting can cause cracking and vice versa.

2.3 Representation of Failure Knowledge

Failure Mode and Effect Analysis (FMEA) and Fault Tree Analysis (FTA) techniques are widely used to recognise and understand the causes of future occurrences of failure of a system (Isa et al., 2005; Lindhe et al., 2009; Patev et al., 2005; Pickard et al., 2005; Xiao et al., 2011).

FMEA, first developed in the 1960's, is an analytical tool utilised in reliability analysis studies, such as the management of engineering systems, to assess the reliability and safety of system processes (Seyed-Hosseini et al., 2006). As a result of the analysis, the possible failure cause affecting the system's functionality and the consequences of such failure are identified to minimise or eliminate failure in systems. It aims to answer the questions of: *'What might go wrong?'*, *'What could cause it to go wrong?'*, and *'What would the effect of it be?'* (Seyed-Hosseini et al., 2006). The consideration of subjective and expert knowledge results in a weighted ranking system that assigns a priority (risk) value to the possibility of a failure event. From this, the overall impact of an event can be calculated using Equation 2-3 below.

Equation 2-3

$$Risk = Probability\ of\ Event\ \times Consequence$$

FMEA has been successfully employed in a number of fields to combine the effect of multiple failure modes (Pickard et al., 2005; Xiao et al., 2011), prioritise failures based on an adjusted risk priority number (Seyed-Hosseini et al., 2006), and with the use of functional models automate the FMEA process of complex engineering systems (Hawkins and Woollons, 1998). Further development of the approach includes using fuzzy logic, causal reasoning, and Boolean representation of the inputs and outputs (Bell et al., 1992; Wang et al., 1995; Xu et al., 2002) to improve the assessment of the known and unknown variables.

The causes of failure can be presented in a graphical manner using FTA, such that concurrently occurring causes of failure can be subjectively included in the representation of failure (Patev et al., 2005; Yuhua and Datao, 2005) and presented as failure paths. The approach is utilised in safety and reliability assessments of engineering systems, in a similar manner to that of FMEA. This approach includes logic *AND* gates and *OR* gates, allowing for multiple causes of failure to be included in the binary assessment of the system and increasing the number of scenarios the approach can replicate (Lindhe et al., 2009). However, unlike the FMEA, no consequence of the event is included in the FTA approach. Despite this, the benefit of this approach includes the representation of multiple causes, and the interactions between such causes, of failure (Patev et al., 2005; Schlotjes et al., 2012b).

FTA has been adopted in manufacturing, chemical, and other engineering fields. Such studies include using fault trees to estimate the failure probability of gas pipelines (Yuhua and Datao, 2005), building binary decision diagrams from fault trees to calculate event probability (Reay and Andrews, 2002; Remenyte-Prescott and Andrews, 2008), and risk probabilities of drinking water systems (Lindhe et al., 2009). Development work to improve the FTA approach includes incorporating human logic into the system (Ortmeier and Schellhorn, 2007), modelling sequential failure logic in fault trees (Long et al., 2000), and using fuzzy logic to account for imprecise data (Yuhua and Datao, 2005).

However, although there appears to be little information in the literature about the use of the FMEA and FTA approaches in pavement performance modelling, the concepts of each of these approaches are applicable to the analysis of road pavement failure. Consequently, their application to assist with incorporating engineering knowledge in the development of the computational model for predicting road pavement failure is explored further in Chapters Four and Six.

2.4 Modelling Terminology

Generally, the approaches used to model pavement deterioration are deterministic or probabilistic (Haas et al., 1994; Lytton, 1987; Martin, 2008). Figure 2-5 graphically defines the different model types, which are further discussed below.

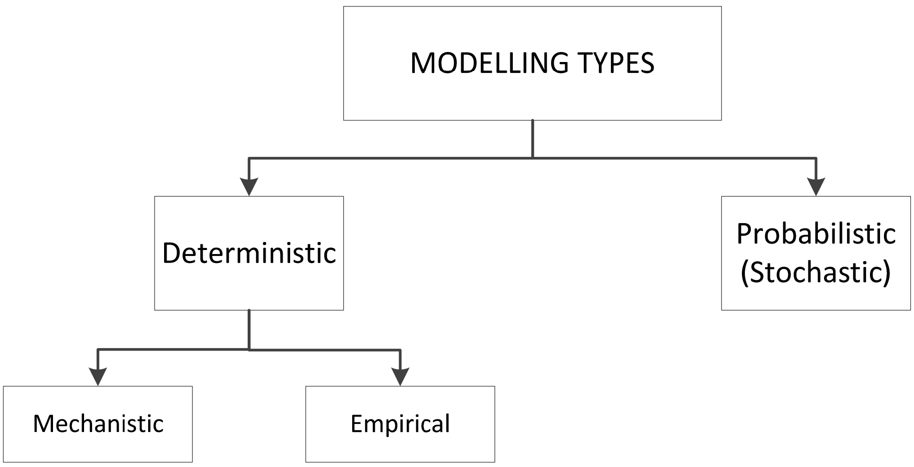


Figure 2-5: Overview of modelling types

2.4.1 Deterministic versus Probabilistic

Deterministic: The outputs of the model are determined through known statistical relationships and model parameters, without the presence of random variation. Randomness does not play a role in the future model predictions. Therefore, for a given initial state, the model will always produce the same result, such that the exact output is known once the system begins. Such models yield future values of the forecasted parameters, which in practice have been used widely for both network and project level applications (Austroads, 2009a).

Probabilistic: An element of chance is involved in a probabilistic (stochastic) model where the model predicts the variability and probability of the dependent variable. The outputs of the model are determined through variables described by probability distributions as opposed to unique (discrete) values. Therefore, the model can predict the likelihood of an event, but the timing or the occurrence of the event is unknown. The mean of the probability distribution defines the most likely output estimate. As these model types demand computing resources, they are predominantly used at a network level (Austroads, 2009a).

2.4.2 Mechanistic versus Empirical

As per Figure 2-5, deterministic models may be mechanistic or empirical as described below.

Mechanistic: An understanding of the behaviour of the system is used to develop the model. Fundamental knowledge, first principles and laws are referred to and used in describing the model relationships.

Empirical: Data from observations or experiments is used to develop the model. Statistical methods, such as regression analyses, are employed in obtaining the relationships between the inputs and output(s) and defining the model parameters. These methods are useful when the behaviour of pavement failure is not fully understood. However, the transferability of the developed models outside of the data domain is often not feasible (Martin, 2008).

Mechanistic-Empirical: A combination of the mechanical and empirical approaches where the model development is based on theoretical hypotheses and is calibrated using observational data. These models can be used outside of the data domain, only with appropriate calibration and sound engineering knowledge.

2.4.3 Linear versus Non-Linear

Linear: A linear function takes the form of Equation 2-4, resulting in a straight-line relationship between the dependent (output) and independent variables (inputs). This function form satisfies both superposition (addition) and homogeneity (scalar multiplication) properties (Bapat, 2000; Pruim, 2010), inferred by the constants c and m in Equation 2-4. Given the complexity of pavement performance data (refer Section 1.2) (Reigle, 2000), linear model forms are not expected to be successful in accurately modelling pavement performance.

Equation 2-4

$$y = mx + c$$

Non-Linear: In non-linear models, the output is not directly proportional to the independent variables and further does not satisfy the associated properties of linear functions above. Conventional pavement design typically follows a non-linear approach design criteria.

2.4.4 Classification versus Regression

Model outputs can take the form of classification or regression, further explained below.

Classification: The model outputs take a discrete class label, such as a binary value {0, 1}, or a group membership from a list of categorical expressions, such as material types {rock, sand, clay}. Pavement failure is represented in this research as a binary class label, where a failed pavement is assigned a one (1) and a sound pavement is assigned a zero (0).

Regression: The values of the outputs take continuous values, opposed to discrete values alike above, where the dependent variable of a regression model is numerical. Early pavement deterioration models employed simple regression analyses; but now regression analyses are commonly employed in calibration exercises of mechanistic-empirical models (Austroads, 2009b).

2.5 Performance Modelling of Infrastructure Assets

Performance models play an integral part in the management of infrastructure assets where forecasting future maintenance requirements heavily rely on the predictions from such models. Numerous studies have employed both deterministic and probabilistic approaches in the development of performance models, focussing on two phases of the asset's life:

- **Deterioration:** How the asset deteriorates over time, and

- **Asset failure:** The end of life of the asset, when replacement, reactive or remedial but not preventative, maintenance is applied.

2.5.1 Deterioration Modelling Techniques

The success in the development of deterioration curves has increased the popularity of predictive methods for infrastructure assets, including bridges (Basheer et al., 1996; Lounis and Madanat, 2002; Madanat et al., 1997; Pan et al., 2009), water, wastewater and stormwater networks (Abraham et al., 1998; Ariaratnam et al., 2001; Baik et al., 2006; Kleiner et al., 2001; Osman and Bainbridge, 2011; Wirahadikusumah et al., 2001; Younis and Knight, 2010), and others (Morcous et al., 2002; Sharabah et al., 2006).

Both stochastic and deterministic approaches have been explored in pavement deterioration studies, and include Markovian processes (Wang et al., 1994; Yang et al., 2005; Yang et al., 2006), Bayesian probabilities (Amador-Jiménez and Mrawira, 2011; Gao et al., 2012), and regression analysis to relate the independent and dependent variables (Arambula et al., 2011; Ben-Akiva and Gopinath, 1995; Gupta et al., 1997; Isa et al., 2005; Manik et al., 2007; Martin, 2008; Park et al., 2001; Prozzi and Madanat, 2002; Sadek et al., 1996). Markovian processes exhibit stochastic behaviours often used in Markov chains, where the next state in the chain is solely based on the current state and not the preceding events in the chain (Austroads, 2009b; Wang et al., 1994). Bayesian probabilities enable reasoning to be inferred in models mathematically through the use of probabilities and prior evidence (Amador-Jiménez and Mrawira, 2011; Gao et al., 2012).

Henning (2008) investigated the use of the generic mechanistic-empirical HDM models for New Zealand road pavements and identified limitations with the linear regression approach taken in the development of these models. Henning (2008) found that the generic HDM models were not able to accurately predict pavement deterioration and instead developed logistic regression models (see Section 2.6.1) for predicting crack initiation and accelerated rutting, using the development of the HDM models as a benchmark. The proposed logistic regression models considered additional influential factors, such as pavement thickness when calculating the probability of accelerated rutting, which was previously omitted from the generic HDM models.

2.5.2 Predicting End of Life Probabilities

Although an important consideration in the management of infrastructure assets, literature focussing on predicting failure is scarce compared to studies aimed on deterioration modelling. However, failure models developed for other infrastructure assets were discussed in the literature, such as for power systems (Halilčević et al., 2011; Hu et al., 2011; Rau et al., 1977), water networks (Anghel, 2009; Zhong et al., 2011), mechanical engineering systems (Pickard et al., 2005) and others (Cai, 1996). Previously, stochastic approaches were employed to generate probabilities of failure, such as failure or fault trees, fuzzy logic, and probability theory (Cai, 1996; Hu et al., 2011; Pickard et al., 2005). However, with the recent advances in computation efficiency and performance, sophisticated computational methods have become a popular choice in engineering fields.

Anghel (2009) investigated the performance of classification approaches to predict the failure probability of pipelines. The predicted output was successfully calculated with a binary support vector machine classifier. Included in the methodology was a definition of the failure modes in the model inputs. An assessment of the performance of the independent variables influenced the model parameters, including the choice of kernel transformation (see Section

2.6.3), such that industry knowledge is inferred in the computational model. The results from this study were comparative to traditional stochastic methods used in predicting pipeline failures.

A variety of complex factors, some uncertain, play a role in the failure of dams. Zhong et al. (2011) employed stochastic methods to calculate the probability of dam failure risk, based on the individual probabilities from the three selected failure types. Probability theory was employed in combining these into an overall failure assessment, assuming the factors contributing to dam failure were independent. To employ such theory, the authors discussed the importance of identifying the factor relationships prior to calculating the overall failure probability.

Cai (1996) defines failure in terms of a predetermined failure criterion; either the system has failed or not, and represents the outcome in a binary format {0, 1} where failure is represented by one (1) and zero (0) represents non-failure. This study investigated the use of fuzzy logic in safety, reliability, and risk assessments of system failure to improve other model formats, such as deterministic or probabilistic. Employing fuzzy logic in such models (see Section 2.6.7) accounted for the fuzzy undefined states of variables in engineering systems.

In the transportation sector, Arampamoorthy and Patrick (2010) based the predictions of pavement failure probability solely on the historical trends of four road networks, using regression analysis techniques, and deterministic design criteria. This simplified approach did not identify the interactions between factors causing failure; instead the study included only traffic as a determining factor in the predictive approach with no consideration of the pavement strength, composition, or subgrade properties. A further limitation of this study was the unsuccessful delivery of a robust computational model that could be geographically transferred to other road networks and other road datasets.

2.6 Application and Performance of Classification Methods

Classification techniques, both discriminative and generative methods, have been employed by other researchers to predict an event, such that the predicted output takes a discrete value. Discriminative classification techniques aim to explicitly define the classification problem and separate the classes of data using a decision boundary. On the other hand, generative classifiers define an event from the observed data using a probabilistic distribution of each variable. Therefore, the latter are employed to predict a classification from a probability density function, as opposed to the former that predicts to which class the new observation belongs (Rogers and Girolami, 2012).

The literature reports better performance from discriminative classifiers than generative methods. However, excluding generative classifiers from the literature search preconceives the notion that one technique is superior over another. Therefore, this section investigates both classification methods, including the following discriminative classifiers:

- Logistic regression;
- Neural networks;
- Support vector machines;
- Decision (probability) trees and random forests, and
- k-nearest neighbours.

2.6.1 Logistic Regression Studies

Logistic regression separates the data into classes with the use of a linear decision boundary (Eastaugh et al., 1997). The popularity of this technique is attributed to the transparency of the learning algorithm and the identification of causal relationships within the data (Tu, 1996). Despite the apparent simplicity of the technique, logistic regression has performed comparatively to other techniques using a range of datasets containing up to 60 input variables; although this technique reported a better performance on smaller datasets (Lim et al., 2000). The limited computational power required and quick training time makes this technique attractive to researchers and practitioners; however, the linearity of the decision boundary hinders the performance of this technique in solving complex classification problems (Lim et al., 2000; Razi and Athappilly, 2005). However, logistic regression has been applied to road pavement data, leading to the successful employment in the development of deterioration models for New Zealand road networks as described earlier in this chapter (Henning, 2008).

Perlich et al. (2003) investigated the performance of logistic regression models in classifying binary outcomes on several datasets, ranging from 1000 to several million datapoints, in various subject areas. As reported in this study, the logistic regression technique performed well on smaller datasets; however, other techniques tested, such as decision trees, outperformed logistic regression on larger datasets due to the clear class separations included in the larger datasets. The authors were apprehensive on drawing conclusions on superior algorithms, given the results differed significantly between datasets.

Peng et al. (2002) found logistic regression to be effective in predicting binary outcomes in educational analysis, specifically determining if remedial reading is necessary for a child.

Furthermore, Dreiseitl and Ohno-Machado (2002) canvassed studies employing logistic regression and reported on the success of the technique across a wide variety of datasets. In comparison to decision trees and k-nearest neighbour approaches, logistic regression generally outperformed when continuous data was used. Austin et al. (2010) also reported on the superior performance of logistic regression, against decision trees, in predicting mortality of hospitalised patients.

Many studies have compared the performance of logistic regression models to that of neural networks in classification problems (Dreiseitl and Ohno-Machado, 2002; Eastaugh et al., 1997; Eftekhar et al., 2005; Saghafi et al., 2009; Tu, 1996). In most cases, neural networks outperformed logistic regression, as the latter employs a linear decision boundary (Eastaugh et al., 1997). However, Eftekhar et al. (2005) reported a higher accuracy for logistic regression, although the authors employed alternative performance measures given the erroneous and bias performance assessments associated with accuracy (Ben-David, 2008; Dreiseitl and Ohno-Machado, 2002; Parker, 2011).

2.6.2 Neural Networks Studies

Neural networks, also known as multi-layer perceptrons, have become popular in machine learning applications, given their ability to determine non-linear relationships within the data, even in the presence of noisy and incomplete data (Jagielska et al., 1999) or without knowing the relationships between the independent and dependent variables (Razi and Athappilly, 2005). However, this technique is prone to over-fitting, requires greater computational resources, and is described as a *'black box'* approach (Tu, 1996). Neural networks are known to locate local minima[3], yet researchers are not adverse to employing neural networks given the infinitesimal difference between the local minima and global minimum[4] (Eastaugh et al., 1997). The complexity, in particular the non-linear decision boundary, of this technique has resulted in improved performance in comparison to other popular, yet simpler, classification techniques (Dreiseitl and Ohno-Machado, 2002; Eastaugh et al., 1997; Saghafi et al., 2009; Tu, 1996).

Kaseko and Ritchie (1993) employed neural networks to detect and classify pavement cracking. Using images of the road pavement surface, neural networks were employed to distinguish if the pavement was cracked and, furthermore, the type of cracking present on the pavement surface. Initial results showed the multi-layer feed-forward neural network model had difficulty (67 % accuracy) in distinguishing between two types of alligator cracking, yet the model performed well for transverse, longitudinal, and block cracking. However, when distinguishing only between the types of alligator cracking, the detection accuracy of the neural network model increased to 86 %.

In a multi-class problem, Bhattacharya and Solomatine (2006) employed neural networks to classify soil types using data from cone penetration tests. In the cases of binary classification, the neural network model accurately classified sandy soil samples but, in the case of classifying non-sandy soil types, was outperformed by other classification methods.

In the power industry, an index describing the transient stability (a binary output) of power systems was developed using neural networks (Tso et al., 1998). Using a training dataset of 200 samples, including a range of fault locations and loading conditions, the model structure

[3] A possible solution to a mathematical problem, which is located in the neighbourhood of the exact minimum solution, and is often referred to as a relative solution.
[4] The unique solution to a mathematical problem that is associated with the smallest error, and is often referred to as the absolute solution as there is only one optimal solution to the problem.

included one hidden layer and employed 11 input variables. On testing the applicability of the trained classifier, only 3 % of the test samples from an independent evaluation dataset were misclassified. The authors recognised the transferability of the neural network model to other power systems despite the varying complexity between these systems.

The medical industry has employed neural networks as an alternative classification method to linear approaches such as logistic regression; although both of these techniques remain popular (Dreiseitl and Ohno-Machado, 2002). Eastaugh et al. (1997) investigated the use of neural networks in predicting high risk pregnancies. Although this study identified the shortfall of neural networks converging to a local minimum, the results showed a slightly higher accuracy from the neural network model than logistic regression. However, Dreiseitl and Ohno-Machado (2002) reported on similarities in performance between the two approaches when used in binary classification tasks and concluded no single algorithm outperformed another. Instead, logistic regression has been popular in the medical industry given its interpretability and user-friendliness, whereas the complexity from the non-linear decision boundary has led to the successful employment and popularity of neural networks.

2.6.3 Support Vector Machines Studies

Recent advances in machine learning algorithms have resulted in the suitability of support vector machines as a classifier (Cortes and Vapnik, 1995). Despite the similarities with neural networks, support vector machines locate the global minimum and require less computational resources than neural networks resulting in its increased popularity. With the addition of kernel functions[5], engineering knowledge can be inferred in these non-linear models (Noble,

[5] Functions that transform the data into a higher level space where a simpler classification separation is capable. Knowledge of the data can be included in the model, based on the choice of kernel transformation.

2004) by an appropriate choice of the transformation function (for class separation), given the knowledge of the data.

Gualtieri et al. (1999) reported the early success of support vector machines as a binary classifier, while the algorithm was in the early stages of development, and suggested the transferability of this technique to other classification problems. However, as with neural networks, this technique is less transparent than linear models but is more interpretable than neural networks. Support vector machines require fewer input variables and a less complex network structure than neural networks, for generally better classification capabilities (Pal, 2006).

Bhattacharya and Solomatine (2006) reported that support vector machines outperformed neural networks in classifying non-sandy soil types, yet the overall accuracy of the two techniques differed by only 0.2 %. Pal (2006) further recommended support vector machines utilising the radial basis function kernel (see Table 5-2) instead of neural networks for soil classification, given the more accurate results produced and less training time required. The generalisation capabilities of this technique and the transferability of the technique to other datasets were discussed by Pal (2006), given the improved performance of the developed support vector machines model when tested with a dataset unfamiliar to the model.

The similar nature of support vector machines to neural networks has seen these models developed for the medical industry. For example, Gil and Johnsson (2011) reported accuracies greater than 90 % (see Table 2-1) in diagnosing medical dysfunctions using support vector machines with a five-fold cross-validation sampling method applied. The low precision of the results for the urological database was attributed to the quality of the data in comparison to the other two studies. The authors discussed the complications with binary classification, particularly when the data reflects skewed distributions, which could be reflected in the results; such distributions are evident in road pavement data.

Table 2-1: Summary of Results from Gil and Johnsson (2011)

Database	Training Dataset	Testing Dataset	Accuracy (%)
Breast Cancer	565	113	98
Parkinson	195	39	92
Urological	380	76	84

2.6.4 Decision (Probability) Trees and Random Forests Studies

Decision (probability) trees, described as a discriminative classifier, do not in fact construct a decision boundary; instead the algorithm splits the data optimally at each node, accordingly to the classes of the data (Dreiseitl and Ohno-Machado, 2002), and is a favoured choice for easily separable classes (Perlich et al., 2003). The interpretability of decision trees lends itself to studies in various research domains despite the performance of the technique lacking competitiveness against other modelling approaches (Chen et al., 2004). This technique has performed adequately with smaller datasets and models developed using this algorithm are readily scalable to large problems (Razi and Athappilly, 2005).

Chen et al. (2004) presented an automated system for the failure diagnosis of internet faults. This study employed decision trees to diagnose the cause of identified faults achieving a success rate of 93 %. Although the authors recognise the availability of other classifiers, generally with better failure prediction performance, they found that the interpretability and transparency of decision trees to the user exceeds that of alternative classifiers.

Bhattacharya and Solomatine (2006) reported comparative results for decision trees against more sophisticated models, such as neural networks and support vector machines. In fact, in this study the overall performance of the decision trees was similar to that of neural networks in classifying soil types. Razi and Athappilly (2005) compared the performance of decision trees to neural networks and linear regression models, such as logistic regression, in predicting the health of an individual. This study concluded that decision trees performed well against neural networks and both of these techniques were in fact substantially better performers than linear regression techniques. This was further supported by additional studies carried out, such that Perlich et al. (2003) reported a superior performance of decision trees against logistic regression on larger datasets.

A number of decision trees can be collated to create random forests through the method of bagging, where equal weightings are assigned to individual decision trees assuming each tree is equally important to the overall result (Roiger and Geatz, 2003). Through this process, the benefits and attributes of decision trees are inherited and furthermore, as a result of bagging, random forests improve on the stability of individual tree models. As an alternative classifier, random forests are becoming popular given the model simplicity, inherited from decision trees, in comparison to support vector machines and neural networks (Verikas et al., 2011), although the interpretability of decision trees is not conveyed in random forests. Random feature selection is also employed throughout the creation of a random forest (Prinzie and Van den Poel, 2008).

Chandra et al. (2009) reported on a comparative study between random forests, decision trees, support vector machines, a feed-forward neural network, and logistic regression models in predicting the failure of internet (dotcom) companies, given historical data. In this study, random forests outperformed the other four methods; however, the authors found that using two or more techniques to produce a series of classifiers exceeded the performance of any individual model.

Verikas et al. (2011) also investigated the performance of random forests against the performance of other popular techniques, such as support vector machines, neural networks, and decision trees. A variety of datasets were used to evaluate the classification accuracy of the methods, ranging in a number of classes, including binary, and sizes. Although this technique demonstrated superiority on large scale comparisons over the other trialled techniques, no statistically significant results were reported.

2.6.5 k-Nearest Neighbours Studies

The popularity of k-nearest neighbours is attributed to the simplicity, resulting from transparent modelling processes, of the technique (Rogers and Girolami, 2012). Only one parameter (k) is adjusted in the model, opposed to neural networks, for example, which requires the user to define many model parameters. However, the simplicity of the technique often leads to performance deficiency and inappropriate use in complex problems. On larger datasets, this technique can take a long time to train and becomes computationally expensive. As the user defines the distance between the nearest neighbours, this model can misrepresent the problem when the relationships between the data are unknown or difficult to define (Dreiseitl and Ohno-Machado, 2002).

In comparative studies, Caruana and Niculescu-Mizil (2006) reported a better performance from this technique than decision trees, logistic regression, and naïve Bayes. However, other discriminative classifiers, such as neural networks and support vector machines, were found to outperform this technique. Dreiseitl and Ohno-Machado (2002) reported on the poor performance of k-nearest neighbours favouring neural networks and logistic regression over this technique, especially on high-dimensional data, despite the simple and easy model creation. However, Wu et al. (2002) introduced methodology, such as incorporating data patterns and relative importance criteria of dimensional data into the model, to improve the training time and classification efficiency.

2.6.6 Alternative Generative Classifiers

Despite the reported success of discriminative classifiers over generative approaches (Long et al., 2007), it was necessary herein to investigate the following generative classifiers:

- Linear discriminant analysis;
- Naïve Bayes;
- Bayesian networks, and
- Hidden Markov model.

Table 2-2 summarises the advantages and disadvantages of the above techniques.

Table 2-2: Advantages and Disadvantages of Generative Classifiers

Method	Advantages	Disadvantages	Studies
Linear discriminant analysis	• A very simple classifier.	• Only linear relationships can be represented in this model. • This technique assumes the independent variables are normally distributed.	Verikas et al. (2011) reported poor results from linear discriminant analysis model in comparison to other techniques.
Naïve Bayes	• A simple probabilistic classifier.	• This technique assumes strong independence in the model. • As a classifier, this technique is a poor performer, with limited successes reported in classification studies.	This technique is optimal for binary classification, based on two binary independent variables and equal covariances between the two classes (Kuncheva, 2006). Caruana and Niculescu-Mizil (2006) reported naïve Bayes as the worst performer in an empirical study of eight techniques, both discriminative and generative.
Bayesian networks	• This technique can handle incomplete datasets. • A combination of expert knowledge and data can be facilitated in this model. • The technique identifies causal relationships in the data.	• The model can become subjective given the inclusion of prior information and knowledge. • It heavily exploits computational resources, especially with a large number of parameters.	Baesens et al. (2004) reported Bayesian networks performed better than naïve Bayes, linear discriminant analysis, and decision trees, although the accuracy of this model remained low (75 %). This study was outside of the engineering industry.
Hidden Markov models	• A non-linear modelling approach. • Is easy to understand.	• This technique requires a large amount of data to train the model. • A stochastic approach, which ultimately can lead to large unfounded assumptions about the data.	This technique poorly performs as a lone model and consequently is often used as part of a hybrid, resulting in the descriptive strengths attributed with this technique included in the model (Bicego et al., 2009).

2.6.7 Improving the Performance of Classifiers with Hybrids, Boosting, and Bagging

The performance of individual classification techniques can be improved with hybrids, boosting, and bagging methods. While such approaches were considered to be beyond the scope of this research, a brief description of each is given below for completeness.

Rough sets, fuzzy logic, and genetic algorithms, are often used to enhance the performance of the above discriminative classifier techniques in hybrid[6] approaches to classification problems (Hu, 2010; Jagielska et al., 1999; Örkcü and Bal, 2011). Rough sets can account for uncertainty based on exact rules and definitions of the upper and lower boundaries attaining to the specific data (Jagielska et al., 1999). Fuzzy logic accounts for uncertainty and imprecisions in the data, based on approximation reasoning, and characterises the *'grey'* area between data classes (Pan et al., 2011; Yuhua and Datao, 2005). Studies employ genetic algorithms to solve engineering optimisation problems (Bernard et al., 1998; Morcous and Lounis, 2005).

Boosting refers to the methodology of compounding a stronger model from a set of weaker performing classifiers, in a successive order, based on the results from the individual antecedent models. The later models concentrate on correctly classifying the datapoints that have been incorrectly classified in the earlier phases of the model sequence. On completion of the training phase, a weighting factor is assigned to each model in the sequence, based on the performance of the individual model on the training data (Roiger and Geatz, 2003).

Bagging, a simpler approach to compounding methods, assigns equal weightings to the models involved assuming equal importance of each model to the overall result (Roiger and

[6] A collection of different techniques used concurrently as one model.

Geatz, 2003). Random forests are an example of this methodology and, since only multiple decision trees (one model type) are combined, are considered in this research (refer Section 2.6.4 above). Composite bagging methods compound various and assorted modelling techniques and, therefore, were beyond the scope of this research.

2.6.8 Advantages versus Disadvantages of Classifiers

Table 2-3 summarises the key points from the literature review on classification techniques.

Table 2-3: Advantages versus Disadvantages of the Classification Techniques

Technique	Advantages	Disadvantages
Logistic Regression	• Quick to train and run. • A linear decision boundary easily understood by the user. • Better than decision trees, k-nearest neighbours, and linear methods.	• Weaker performance in comparison to neural networks and support vector machines.
Neural Networks	• Non-linear decision boundary so complex problems can be solved. • Reported to perform well. • Can handle large datasets and many inputs. • Transferable to other classification problems.	• Can be computationally exhaustive, associated with long running times. • A *'black box'* approach, which is not very transparent and interpretable for the user.
Support Vector Machines	• More interpretable than neural networks. • Faster to train and run than neural networks. • Proven to be a good performer. • Kernels provide an element of human knowledge incorporated into the model. • Transferable to other domains.	• Less interpretable than decision trees and logistic regression, and still considered slightly a *'black box'* technique.
Decision (Probability) Trees	• Very interpretable. • Good results from smaller datasets. • Comparable against neural networks. • A better classifier than linear methods.	• Works well on easily separable classes. • Less successful on complex problems.
Random Forests	• Good performer against neural networks. • Improved results on single decision trees. • Handles a variety of dataset sizes. • More interpretable than neural networks or support vector machines.	• Less interpretable than decision trees.
k-Nearest Neighbour	• A very simple and user-friendly technique, with little model building. • Generally better than the linear classification methods, generative classifiers and decision trees.	• Deficient in complex problems. • With larger datasets, takes a long time to train and is computationally exhaustive. • Poor performer in comparison to neural networks and support vector machines.
Generative Methods	• Linear discriminant analysis is a very simple model. • Naïve Bayes is a simple probabilistic classifier. • Bayesian networks facilitates human knowledge in the model development. • Hidden Markov models are easy to understand.	• Linear discriminant analysis has very limited scope. • Strong assumptions are associated with naïve Bayes, resulting in poor performance against discriminative classifiers. • Bayesian networks are a subjective approach to classification. • Unfounded assumptions accompanying the stochastic nature of hidden Markov models. More suited as a hybrid, as opposed to a lone modelling technique. • Generally, these models perform poorly against the discriminative classifiers.

In addition to the above, key elements from the literature utilising these binary classifiers are summarised below:

- There is not one superior technique for all domains and datasets (Perlich et al., 2003). The performance of the classifier greatly depends on the specific dataset; therefore, an assessment of the classifiers against each other using the research dataset will be required to identify the classifier most suitable for the task at hand;
- Accuracy is not an adequate measure of the performance of a classifier. Instead, other performance measures should be used to account for the biasness associated with accuracy. Alternative performance measures employed in the studies included phi coefficient, receiver operating characteristic curve, and Cohen's kappa curve (Ben-David, 2008; Dreiseitl and Ohno-Machado, 2002; Parker, 2011);
- Poor quality and incomplete datasets result in low precision and accuracy of the classifier performance, thus addressing missing data in the research methodology will improve the comparative results of the classifiers;
- When the data class distributions are unequal, problems with binary classification occur. During the training phase, a model will always try to minimise the overall error rate when trained on an unbalanced dataset, which results in a developed model that will always predict the larger of the two binary classes (Prinzie and Van den Poel, 2008). Given that unequal class distributions are a reality in network datasets[7], a weighting factor should be applied in the training phase of the models to ensure equal preconception of the two classes (failure and non-failure) by the modelling technique;

[7] Road failure, as a result of preventative maintenance, is less frequent on the network than the number of sound pavements. Therefore, this is reported and represented in the data.

- Using cross-validation sampling methods is a sufficient approach to ensure the variability in the predictions are accounted for in the assessment of the performance of the classifiers (Gil and Johnnson, 2011);
- The majority of the studies, across all classification techniques, tested the model on a reserved (unfamiliar) portion of the dataset. This portion varied from 10 % to one-third of the entire dataset. Larger training datasets will result in a better generalisation of the model, whereas a larger testing dataset will present a better error estimate of the technique (Dreiseitl and Ohno-Machado, 2002), and
- Son et al. (2012) identified the importance of identifying relationships, both linear and non-linear, prior to modelling in large datasets to reduce the number of predictive factors included in the model.

2.7 Summary of the Literature Review

This chapter has reviewed literature associated with:

- Road pavement failure in New Zealand;
- Methodologies for representing failure knowledge in computational models;
- Performance modelling studies, including deterioration models and failure models of infrastructure assets, and
- The performance of binary classifiers for various datasets.

Through the review of the literature, the approaches considered by other researchers were scrutinised. Firstly, from a fundamental perspective, FMEA and FTA have been used to represent knowledge of failure in computational models. FTA has the ability to represent failure as a binary outcome and can represent multiple causes of failure and interactions between failure factors. Therefore, this research adopted the FTA approach to infer engineering knowledge into the computational model, further described in Chapter Four.

Researchers have developed deterioration models for road pavements using a variety of methods, including both deterministic and stochastic approaches. Deterioration models have also been employed in the management of other infrastructure assets.

Although limited studies focussed on predicting the probability of pavement failure, other industries have successfully developed methodologies to address such aims. One study employed a binary classifier to calculate the failure probability of water pipelines.

Comparative studies identified the advantages and disadvantages of classification techniques, as listed in Table 2-3. Classification techniques have been used successfully in a number of areas, ranging from medical diagnosis, power systems, geomechanics, and other engineering fields. From the literature, neural networks and logistic regression emerged as popular choices of binary classifiers; however, Table 2-3 describes the benefits for the employment of the other classification techniques. Since the performance of these techniques was shown to be greatly dependent on the datasets used in the studies, a comparative study (see Chapter Five) further investigates the appropriateness and effectiveness of the classification techniques identified in this chapter for the task of predicting the failure of road pavements from road datasets.

Chapter Three

METHODOLOGY

3.1 Introduction

The literature review demonstrated there has been little research to develop robust computational models, which integrate engineering knowledge, to predict the failure probability of road pavements. To address this, this chapter presents the systematic approach followed to quantify the probability of pavement failure, addressing the aim of the research. Although this research concentrates on the structural failure of chipsealed flexible road pavements, the approach outlined can be adapted for other road pavement types and failure mechanisms.

3.1.1 Conceptual Design

The conceptual design of the system consists of two integral items:

1. A knowledge based approach to aid in diagnosing the cause(s) of failure from road datasets, and
2. A computational model to calculate the probability of failure.

Failure charts for each of the three failure mechanisms, outlined in the scope of this thesis (refer Section 1.3.3), were developed to form the diagnostic element of the system and were a direct output from the first research objective. The computational model formed the probability element of the system and was developed using a technique selected from an analysis, using objective criteria, of a number of techniques described in the literature. With this and the failure charts, the system can determine the probability of failure and identify the most probable failure path, deemed to be the critical failure path in this research.

3.1.2 Research Methodology

The research methodology, shown in Figure 3-1, consisted of the following:

- Understanding the failure mechanisms, capturing engineering knowledge associated with road pavement failure, and utilising this to develop failure charts;
- Selecting a number of classification techniques, from the literature, to evaluate for the task at hand;
- Investigating the performance of the selected techniques to accept one as the most appropriate technique for further development;
- Developing and testing of the system using research data, and
- Testing of the prototype model on an independent road dataset.

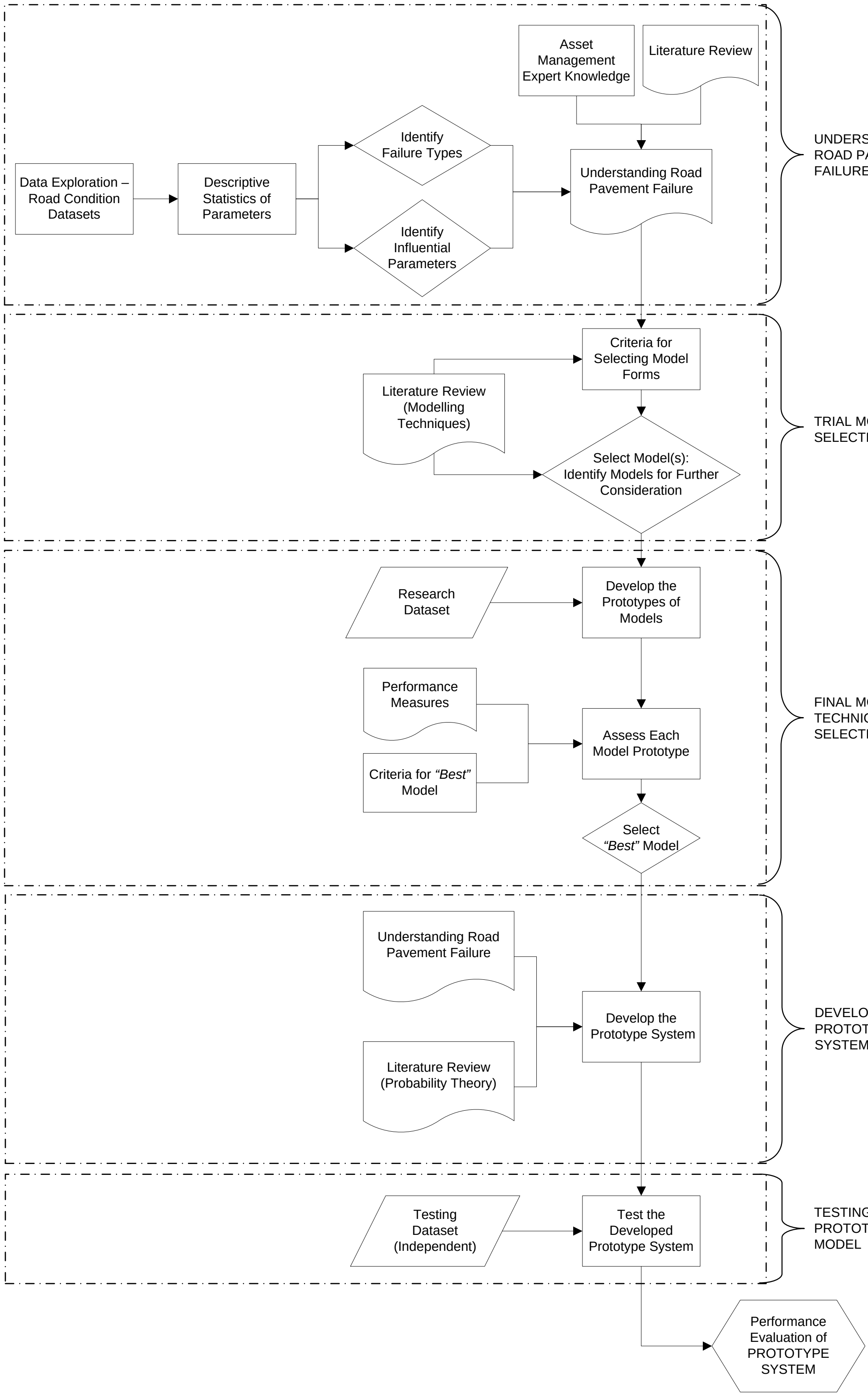


Figure 3-1: Methodology for quantifying the probability of failure

3.2 Research Modules

3.2.1 Expert Engineering Knowledge and Failure Understanding

Consultation with experts in conjunction with a preliminary analysis of road network datasets, comprising of descriptive statistics and correlation analyses of the available road datasets including a typical RCA network dataset, and a review of the literature identified the most important factors that contribute to road pavement failure. Incorporating engineering knowledge into the modelling process resulted in a more accurate and knowledgeable model (Schlotjes et al., 2011; Schlotjes et al., 2012a). In order to facilitate the conceptual design of the research and infer engineering knowledge within the computational model, failure charts were used to present the key factors and combinations contributing to road pavement failure.

The approach to the development of the failure charts was based on FTA (refer Chapter Two), which recognises the importance of identifying the causes of failure and understanding the interactions between the possible failure causes (Patev et al., 2005). The graphical format of the failure charts was found to be a convenient method of presenting the complex interaction between factors and identifies the failure paths.

Figure 3-2 outlines the process of developing the failure charts. The first task was to define failure for each of the failure mechanisms (e.g. rutting, cracking, and shear). The definition of failure for each mechanism assisted in identifying the causes of failure, as well as the interactions between these causes. From this, the combinations of multiple failure causes represent the failure paths for each failure mechanism. Chapter Four further describes how the failure charts were developed using information from the literature and expert opinion, together with specific methodology for the development of an understanding for flexible pavement failure.

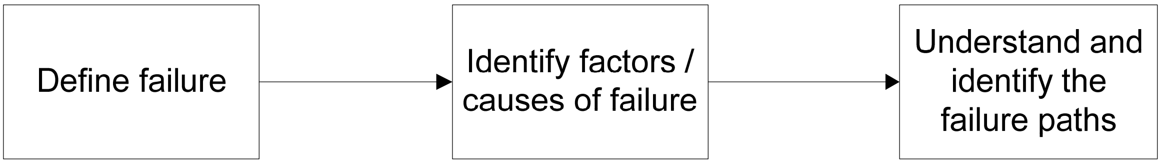


Figure 3-2: Understanding failure of road pavements

3.2.2 Selecting Suitable Classification Techniques

Chapter Two has reviewed the use of computational techniques in a number of fields, including the transportation sector, which may be suitable for modelling road pavement performance. As pavement failure is defined in the research dataset as a binary output {0, 1} (see Section 3.3.1.7), the research problem becomes a binary classification problem. Table 2-3 identified a number of classification techniques for further consideration and the suitable classification techniques were selected using the following criteria:

1. The performance of the modelling technique must be adequately demonstrated in the literature;
2. Road pavement datasets are potentially extensive and may consist of large amounts of data relating to many variables, thus the modelling technique needs to be able to handle such data;
3. Complex interactions exist between the independent and dependent variables of road pavement data (Reigle, 2000). Complex problems are better represented with non-linear model functions, therefore non-linear techniques should be favoured in this study, with the provision of using linear methods with caution;
4. The time required to run the selected model should be kept to a minimum to improve computational efficiency;

5. The training phase of the model should be simple, easily inferred, and based on the occurrences in the data;
6. The overall process should be simple and easily understood by asset managers and pavement engineers for successful implementation in the current road reporting processes, and
7. The model can infer engineering (human) knowledge.

The detailed investigation of the suitability of the techniques is discussed in Chapter Five.

3.2.3 Assessing the Performance of Binary Classifiers

The selected techniques were compared using a research dataset (see Section 3.3) to select the most appropriate technique for the development of the prototype system, as the literature reviewed in Chapter Two concluded there is not one superior technique for all domains and datasets. This was achieved by assessing the output from each technique according to four performance measures (accuracy, misclassification error, F-score, and phi coefficient), as accuracy alone is not acceptable as an assessment of the model performance (Ben-David, 2007; Parker, 2011). However, this research extended the assessment criteria to include interpretative qualities and usability characteristics of the techniques in reference to the future implementation of the computational technique in the industry. Therefore, the following criteria were adopted:

1. The performance of the model;
2. The speed of the model, simplicity of use, and the overall process is easily interpreted, and
3. The generalisation of the model, such that over-fitting is avoided.

Chapter Five further describes and discusses the performance assessment of the five trialled modelling techniques.

3.2.4 Prototype System Development

As described earlier in Section 3.1.1, the conceptual design of the system consisted of two elements:

1. Failure charts, from the knowledge of failure, to diagnose the cause(s) of failure, and
2. A computational model to calculate the probability of failure.

Therefore, the development of such a system consisted of failure charts and a computational model, as described in Sections 3.2.1, 3.2.2, and 3.2.3. The prototype system was developed from the failure paths presented in the failure charts, with each path being modelled by the computational technique. The resultant system predicted the probability of failure of a road section for each failure mechanism, which was calculated as the maximum (most probable) probability of each failure path, per failure mechanism. The statistical package *R*[8] (Everitt and Hothorn, 2006) constructed the prototype system, as well as the classification techniques trialled in the comparative study; although the purpose of this research was not to assess this statistical package.

Two approaches were considered to determine the overall probability of failure of a pavement section taking into account all failure types. The first was a simple approach that assumed that each failure mechanism acts independently, following the approach adopted in conventional pavement design (Austroads, 2012):

[8] A number of packages are readily available to be used for the task at hand; however, given the access to a large library of model files, *R* is capable of modelling each of the techniques outlined in Chapter Two.

Equation 3-1

$$P_{FAILURE} = max[P(A), P(B), \ldots, P(N)]$$

$$where: A = Failure\ type\ A, B = Failure\ type\ B, and\ N = Failure\ type\ N$$

However, a more complex model considered to be more realistic was also developed that considered the interdependence of each failure type using the *'Additive Law of Probability'* (probability unions), by which multiple failure mechanisms can be combined (Ayyub and McCuen, 2003; Mendenhall and Beaver, 1991) as follows:

Equation 3-2

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

$$where: P(A \cap B) = P(A) \times P(B)$$

$$A = Failure\ type\ A, B = Failure\ type\ B, C = Failure\ type\ C$$

$$P(A\ or\ B\ or\ C) = P(A) + P(B) + P(C) - P(A) \times P(B) - P(A) \times P(C) - P(B) \times P(C) + P(A) \times P(B) \times P(C)$$

$$where: 0 \leq P(A \cup B \cup C) \leq 1$$

$$A = Failure\ type\ A, B = Failure\ type\ B, C = Failure\ type\ C$$

Equation 3-2 calculates the probability of either failure mechanism occurring using the maximum probabilities from each failure mechanism, as described previously (refer Section 3.2.4). The description of the development of the system and interactions between the failure mechanisms is given in Chapter Six, as well as an assessment of the performance of the prototype system.

3.2.5 System Testing and Applications

The prototype system was tested using an independent dataset which had not been used to train the computational model. The test consisted of comparing the probabilities of failure computed using the prototype system with the number of actual failed sections in the dataset.

The practical application of the system includes:

- The probability of failure of a road section for each failure mechanism;
- An overall failure probability for a given road section, and
- The performance of the road network, based on the system's predictions.

Chapter Seven presents a detailed analysis of the prototype system testing, along with a discussion on the practical applications of the system.

3.3 Research Datasets

The New Zealand Long-Term Pavement Performance (LTPP) programme was established in 2000 and consists of a number of sites where data is gathered from New Zealand road networks. This research used two datasets compiled from the LTPP database; one of the road sites on the State Highway network, and the other from sites across the Local Authority road networks. The primary difference between the two datasets is the external environment of the road pavement sections.

The benefit of the LTPP programme is the frequency of data collection and the high level of detail of the data, which has resulted in a sufficient amount of data that may be considered appropriate for such research studies. Traffic data, inventory data, condition data, detailed site inspections, and maintenance reports are included in the LTPP programme database (Henning et al., 2004). Each site is surveyed annually to record the condition of the pavement using highly accurate instruments and detailed road profile equipment.

Figure 3-3 illustrates the layout of a typical LTPP inspection site. Data is collected from the first 300 metres of each site with the length of the site divided into six 50 metre sub-sections.

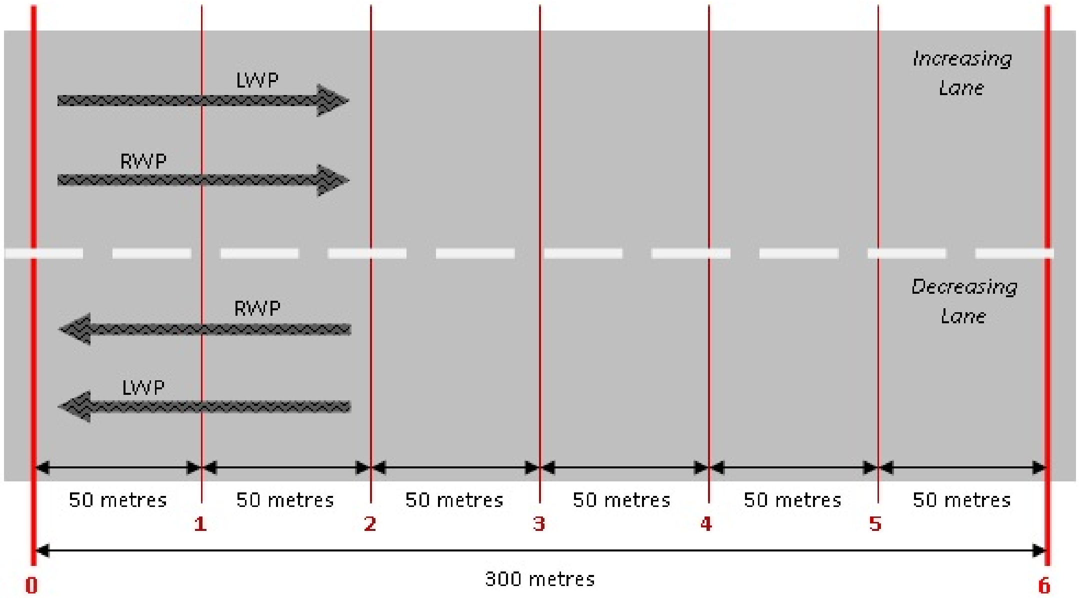


Figure 3-3: LTPP road pavement section layout

3.3.1 State Highway LTPP Dataset

From the State Highway LTPP dataset the following data types were obtained:

- Traffic;
- Pavement composition;
- Pavement strength;
- Environment;
- Subgrade sensitivity;
- Surface condition, and
- Failure data.

3.3.1.1 Traffic Data

The dataset contains the Annual Average Daily Traffic (AADT) and Percentage of Heavy Commercial Vehicles (HCV). However, it was felt necessary to convert these into an alternative measure of traffic, the Cumulative Number of Equivalent Standard Axles (Cum_ESA), which takes into account the impact of heavy vehicles on road pavements over time. This was achieved using Equation 3-3 (Austroads, 2012) as follows:

Equation 3-3

$$Cum\_ESA = \sum_{age} N_E \times G_F \times 365$$

$$= \sum_{age} \frac{AADT \times F \times D \times HCV}{100} \times G_F \times 365$$

$$where: age = base\ layer\ or\ surface\ layer\ age$$

$$AADT = Annual\ average\ daily\ traffic$$

$$F = Number\ of\ ESAs\ per\ heavy\ commercial\ vehicle$$

$$D = \ Direction\ factor\ (vehicles\ travelling\ in\ one\ direction) = \frac{1}{Number\ of\ Lanes}$$

$$HCV = Percntage\ of\ heavy\ commercial\ vehicles$$

$$G_F = Growth\ factor = \frac{(1 + 0.01r)^n}{0.01r}$$

$$r = Percentage\ of\ annual\ traffic\ growth$$

$$n = number\ of\ years\ (in\ this\ case, n = 1)$$

The State Highway LTPP dataset used historical traffic data to calculate the annual traffic growth rate r and the breakdown of the HCV classes to calculate the ESA per heavy vehicle. Furthermore, the dataset obtained two Cum_ESA parameters from using both the ages of the surface layer and the base layer in the calculation of cumulative number of ESAs.

3.3.1.2 Pavement Composition Data

Road Assessment and Maintenance Management (RAMM) databases provided the composition data of the LTPP pavements. The information available included the material properties of the pavement layers (surface layer, base layer, and sub-base layer, if applicable), the layer ages, layer thicknesses, number of lanes, and pavement widths; however, some of this data was limited and incomplete for all sites.

3.3.1.3 Pavement Strength Data

The State Highway LTPP dataset contained four factors associated with the strength of a road pavement section. The modified structural number (SNP) is back-analysed from deflection readings and its average value, along with a binary strength parameter which indicates whether the pavement is weak or strong, was utilised in this research dataset. A pavement was considered strong in this research if (i) SNP ≥ 2.5, or (ii) the thickness of a single layer exceeded 250 mm. Individual deflections from the falling weight deflectometer (FWD) were available, from which the following parameters were calculated for comparative measures of strength (Horak, 2008; Thom, 2008):

Equation 3-4

$$Radius\ of\ Curvature = \frac{L^2}{2 \times D_0 \times \left(1 - {}^{D_{200}}/_{D_0}\right)}$$

Equation 3-5

$$Base\ Layer\ Index\ (BLI) = D_0 - D_{300}$$

Equation 3-6

$$Middle\ Layer\ Index\ (MLI) = D_{300} - D_{600}$$

Equation 3-7

$$Lower\ Layer\ Index\ (LLI) = D_{600} - D_{900}$$

$$where: L = 200mm\ for\ FWD$$

$$D_0 = Maximum\ deflection\ at\ point\ of\ loading$$

$$D_{200} = Deflection\ measured\ 200mm\ from\ point\ of\ loading$$

$$D_{300,600,900} = Deflection\ measured\ at\ edge\ of\ zone\ 1\ (positive\ curvature),$$

$$zone\ 2\ (curvature\ inflection), zone\ 3\ (reverse\ curvature)$$

Finally, a concurrent New Zealand study developed structural indices algorithms which, by definition, measure the pavement's ability to withstand failure in relation to a particular failure mode (rutting, flexure, shear, and roughness) (Salt et al., 2010). Each LTPP site was analysed using the structural indices and were included in the State Highway LTPP dataset.

3.3.1.4 Environmental Data

The LTPP programme obtains rainfall data from the National Institute of Water and Atmospheric Research (NIWA) and, accordingly, the State Highway LTPP dataset contains the annual rainfall values (in millimetres). Water entering the pavement once the surface seal has cracked is a key factor in the deterioration of the pavement structure, which is not a factor included in the LTPP programme. Therefore, it was felt to develop a new variable that is a measure of the cumulative rainfall once the pavement is cracked (CumRain_ifCracked), using the rainfall data in conjunction with site condition reports.

3.3.1.5 Subgrade Sensitivity

A design matrix included in the LTPP analysis (Henning, 2008) categorises the pavement sites into sensitivity classes based on the properties of the underlying subgrades. This accounts for the subgrade material types, drainage, and environmental conditions. This factor classifies the sites into low, medium, and high sensitivity zones.

3.3.1.6 Condition Data

This research defined three failure mechanisms that have been found to be prevalent on low volume roads in New Zealand, namely rutting, cracking, and shear (Arampamoorthy and Patrick, 2010; Creagh, 2005; Gribble and Patrick, 2008).

Rutting Data:

The rutting depths on the LTPP pavement sites were evaluated using a transverse profile beam. For each wheelpath, the rutting depth (in millimetres) is calculated by averaging two measurements taken at the same point on the road surface. Following this, the method is repeated every 10 metres from the start (0) point, resulting in an average rut depth for each wheelpath every 10 metres. Along each 50 metre subsection (1 through to 6) (refer Figure 3-3), the rut depths are further averaged for a resultant rut depth per wheelpath per subsection. A combination of the results from the two wheelpaths provides an overall estimate of the rut depth for the lane as a whole. For completion, the State Highway LTPP dataset included the rut depths for the both of the wheelpaths and the lane (increasing or decreasing). Further manipulation of the rut depths produced a rut rate (per year) for each subsection.

Cracking Data:

The total amount of cracking was included in the dataset in terms of a percentage of the subsection area. A number of cracking types were defined in Chapter Two and, although structural failure is largely associated with alligator cracking, all information regarding cracking types (e.g. alligator, transverse, longitudinal, and parabolic cracks) reported in the State Highway LTPP database were included in the research dataset. Further manipulation of the total cracking produced a crack rate (per year) for each subsection, as well as the number of years of continual cracking (i.e. years with no natural closure of the cracks or maintenance to remedy the cracks).

Shear Data:

Shear failure often manifests in the form of shoving on the pavement surface or, secondarily, by potholes (refer Chapter Two). Therefore, pothole information (number, depth, and diameter), shoving data, structural patches, and mechanical damage were obtained from the State Highway LTPP database. To ensure an unbiased comparison, the latter three data elements were converted into a percentage of the total subsection area. Pothole depth and diameter remained in millimetres and the number of potholes remained as a number, as this condition data type is more meaningful in the raw data format, as per the LTPP programme collection process.

3.3.1.7 Failure Data

In practice, pavement failure is defined by:

- Fundamental limits and definitions, defining failure as the design thresholds of conventional pavement design practices. For rutting, this was defined as a rut depth ≥ 20 mm and for cracking, an affected surface area ≥ 15 % or a crack width ≥ 5 mm (Austroads, 2006; Austroads, 2007b), and
- Opinions of practitioners, where the timing of maintenance intervention defines failure.

The objective of the LTPP programme restricts the use of maintenance until such time that the pavement becomes unsafe and intervention is necessary to ensure the functionality and safety of the road pavement. As a result of this, this research defined failure in the State Highway

LTPP dataset as the time of maintenance intervention. Maintenance records and inspection reports provided the information to establish the type of failure, the year of failure, and the maintenance treatment applied to the pavement. In the State Highway LTPP dataset, binary classifiers represent the failure condition of the road pavement sections, with zero (0) indicating a sound pavement (a positive outcome) and one (1) indicating a failure (a negative outcome). Given the information available, this dataset defined failed sections for each of the rutting, cracking, and shear failure mechanisms.

3.3.2 Local Authority LTPP Dataset

Much like the State Highway network, the LTPP programme includes sites on the New Zealand Local Authority network. A total of 82 sites in rural and urban areas of the network have been monitored since 2003. The data available in the database is identical to that of the State Highway LTPP network, meaning the format of the data is much the same to the State Highway dataset. However, Local Authorities manage these sites and consequently the inventory records are less complete and reliable compared to those included in the State Highway dataset.

This dataset was utilised in the final module of the methodology (testing the prototype system) because of the similarities in the presentation and format of the data to that of the State Highway dataset, meaning the transferability of the computational model to an independent dataset was simplified. However, some factors included in the prototype system were absent in this dataset because of the differences in the road structure. For example, the road pavements in the Local Authority dataset lack a sub-base layer because the traffic demand is much less than that of the State Highway network. These differences were considered minor so to not affect the verification and testing of the prototype system.

3.3.3 Summary of the LTPP Datasets

Based on the descriptions of the LTPP data presented above, Table 3-1 summarises the variables included in the computational models for each of the three failure modes. The data available on material properties was considered to be incomplete and therefore was not included in this research dataset.

Table 3-1: Variables included in the LTPP Datasets

Factor Group	Variables from the Research Dataset
Traffic	• Average Annual Daily Traffic (AADT)[a,b,c] • Total percentage of heavy vehicles[a,b,c] • Cumulative number of Equivalent Standard Axles (ESA), given the base layer age[a,b,c] • Cumulative number of Equivalent Standard Axles (ESA), given the surface layer age[b]
Composition	• Base layer age[a,b,c] • Subbase layer age[a,b,c] • Surface age[b] • Total pavement thickness, excluding surface thickness[a,b,c] • Total pavement thickness, including surface thickness[b] • Pavement width[a,b,c] • Number of lanes[a,b,c]
Strength	• Strength of pavement (weak or strong)[a,b,c] • Structural number (SNP)[a,b,c] • Structural indices (SI) for rutting, flexure, shear and roughness[a,b,c] • Falling weight deflectometer (FWD) parameters (Radius of Curvature, Base Layer Index, Middle Layer Index, Lower Layer Index)[a,b,c]
Environment	• Cumulative rainfall once the pavement is cracked[a,b,c]
Subgrade Sensitivity	• Sensitivity of pavement (low, medium, or high)[a,b,c]
Surface Condition	• Rut depths for left-hand wheelpath, right-hand wheelpath, and lane[a,c] • Rut rate for left-hand wheelpath, right-hand wheelpath, and lane[a] • Total cracking (all cracking types)[b] • Crack rate[b] • Number of years of continual cracking[b] • Mechanical damage[c] • Structural patch[c] • Pothole diameter, depth, and number[c] • Shoving[c]

[a] Included in the rutting datasets; [b] Included in the fatigue cracking datasets; [c] Included in the shear datasets

3.4 Summary of the Methodology

In order to quantify the probability of road pavement failure, the development of the prototype system comprised of three elements; **engineering knowledge** was incorporated into a **computational model**, and **probability theory** was referenced to combine the individual failure probabilities into one overall assessment of the failure probability. To achieve the research aim, this chapter has described a systematic approach comprising of five stages. These sequential modules were required to produce knowledge of road pavement failure and failure charts, explore and assess a number of modelling techniques in order to select an appropriate technique, further develop the overall system, and test the prototype system on an independent dataset. While a large number of techniques are available to represent road pavement failure, it was necessary to assess the performance of each and the requirements of the overall system in order to select the most appropriate technique. The research dataset from the LTPP programme was used in developing the prototype system (Chapters Four to Six), with the testing of the system using an independent LTPP dataset (Chapter Seven). Detailed descriptions of each module are presented in Chapters Four to Eight.

Chapter Four

UNDERSTANDING ROAD PAVEMENT FAILURE

Complementing the Subject Knowledge

4.1 Introduction

This chapter addresses the first objective of the research by developing a comprehensive understanding of road pavement failure that will form the foundation of the model development (as detailed in Chapter Five). The knowledge obtained will improve the success of the computational model, as it will ensure the model reflects the knowledge of road failure as opposed to coincidental relationships in the data that may result from statistical processes (see Section 4.1.1). Furthermore, it establishes the relationship between the factors involved in each failure mechanism and determines the causes of failure, thereby facilitating the identification of the most probable (critical) failure path assuming all failure paths are considered equally critical in the analysis.

This chapter addresses Objective One of this research by:

- Developing a methodology to enable knowledge to be captured in a format that can be used to aid in and inform the development of the computation model;
- Identifying the factor which contribute to failure, and
- Developing comprehensive failure charts for flexible pavement failure.

4.1.1 Importance of Understanding Road Pavement Failure

The conceptual design (refer Section 3.1.1) of the proposed system includes a diagnostic element, provided by engineering knowledge that is inferred into computational models. Since this element is fundamental in the development of the prototype system, understanding the failure mechanisms was seen as a major part of the research. As road failure is complicated and difficult to artificially replicate with computational models (Reigle, 2000), understanding the mechanisms surrounding road pavement failure is imperative in developing such models.

In practice, diagnosing road failure is challenging for asset managers due to the variable behaviour of road pavements, the complex interactions of factors that can contribute towards failure, and the occurrence of multiple failure modes. An incorrect failure diagnosis leads to ill-chosen and ill-timed maintenance decisions, as well as variable outcomes of the forecasted maintenance programme and budget. However, a better understanding of the causes and the identification of failure paths, associated with the underlying causes of failure, can facilitate the application of appropriate maintenance. Therefore, the diagnostic framework not only contributes towards the outcome of this research, but also has a direct practical application (see Chapter Seven). Accordingly, the importance of developing a comprehensive understanding of pavement failure can:

- Assist in identifying the cause of failure and subsequently the failure mechanism(s);
- Provide information that can aid in making informed decisions regarding the forecasting of maintenance, including the treatments;
- Inform the computational modelling process by assisting with the choice of appropriate variables to be included in the development of such models;

- Assist with the interpretation of the results from the computational model, and
- Validate the outputs of the statistical (data driven) computational model based on the relationships included in the model.

4.1.2 Methodology for Failure Knowledge and Charts

Developing such an understanding for the three failure mechanisms (refer Section 1.3.3) of this research utilised three sources of knowledge:

- A literature review identifying the first principles and fundamental factors of failure;
- Expert knowledge from the industry, and
- A preliminary analysis of road datasets to identify causal relationships between failure and it's causes.

The knowledge obtained from these sources was presented in the form of failure charts as described below. Initially, from first principles surrounding pavement design, a generic pavement failure chart was developed and, based on five factor groups identified by experts, additional failure charts were developed for each of the three failure mechanisms.

4.2 Fault Tree Analysis

The development of the failure charts was based on the FTA approach, whereby a tree-like chart is used to represent the multiple causes of failure and failure paths are used to show the collection of factors contributing to failure (Yuhua and Datao, 2005). The breakdown of each failure mechanism in this way enables the model definition to represent concurrently occurring failure factors, competently and effortlessly (Lindhe et al., 2009; Patev et al. 2005;

Remenyte-Prescott and Andrews, 2008). The FTA method recognises the importance of identifying the causes of failure to accurately classify model predictions.

Among other applications, FTA is commonly used in research studies as a diagnostic tool (Volkanovski et al., 2009) to identify the causes of a specified event. Because of its prioritisation attributes in the design of fault trees, the placement of highly influencing factors earlier in the tree gives these factors more control on the outcome (Ortmeier and Schellhorn, 2007). The use of combinatorial logic, such as *AND* and *OR* functions, enables the interactions between the factors to be easily represented (Contini and Matuzas, 2011; Lindhe et al., 2009) and, as a result, the model can represent a number of scenarios (Remenyte-Prescott and Andrews, 2008; Yuhua and Datao, 2005). The development of such fault trees for the research reported here is described further below.

4.3 Generic Pavement Failure

The development of a generic failure chart provided:

1. An understanding, derived from first principles, of pavement failure which aided in identifying the mechanisms of failure;
2. Five factor groups contributing to failure to assist in developing specific failure understanding charts for the pavement types (thin, flexible, unbound, granular road pavements) considered in this research, and
3. A methodology for other researchers to develop failure paths and understanding of other pavement structures and types.

From engineering fundamental design principles, pavement failure can occur through either the failure of the structure to support the design load (bearing capacity) and / or because the environmental and traffic load exceeds the design loading (loading demand). Figure 4-1 presents the failure chart developed for any pavement structure, based on the FTA approach described above. As pavement failure can be attributed to either a single factor or a concession of factors, such as excessive traffic loadings alone or combined failure of excessive traffic loadings on a poorly designed pavement, the FTA approach accounts for both of these failure situations. Table 4-1 lists five main groups of factors, which are considered to contribute towards failure, that were obtained through engineering knowledge and the use of Figure 4-1. This table successfully summarises the numerous independent factors presented in Figure 4-1.

The failure chart presented in Figure 4-1 demonstrates the causative nature of the factors contributing to generic pavement failure. The chart is presented in such a way that the causes contributing to the failure are sequential; for example, a pavement with insufficient layer thickness results in a poor pavement composition, which in turn negatively impacts on the design of the structure and results in a bearing capacity failure.

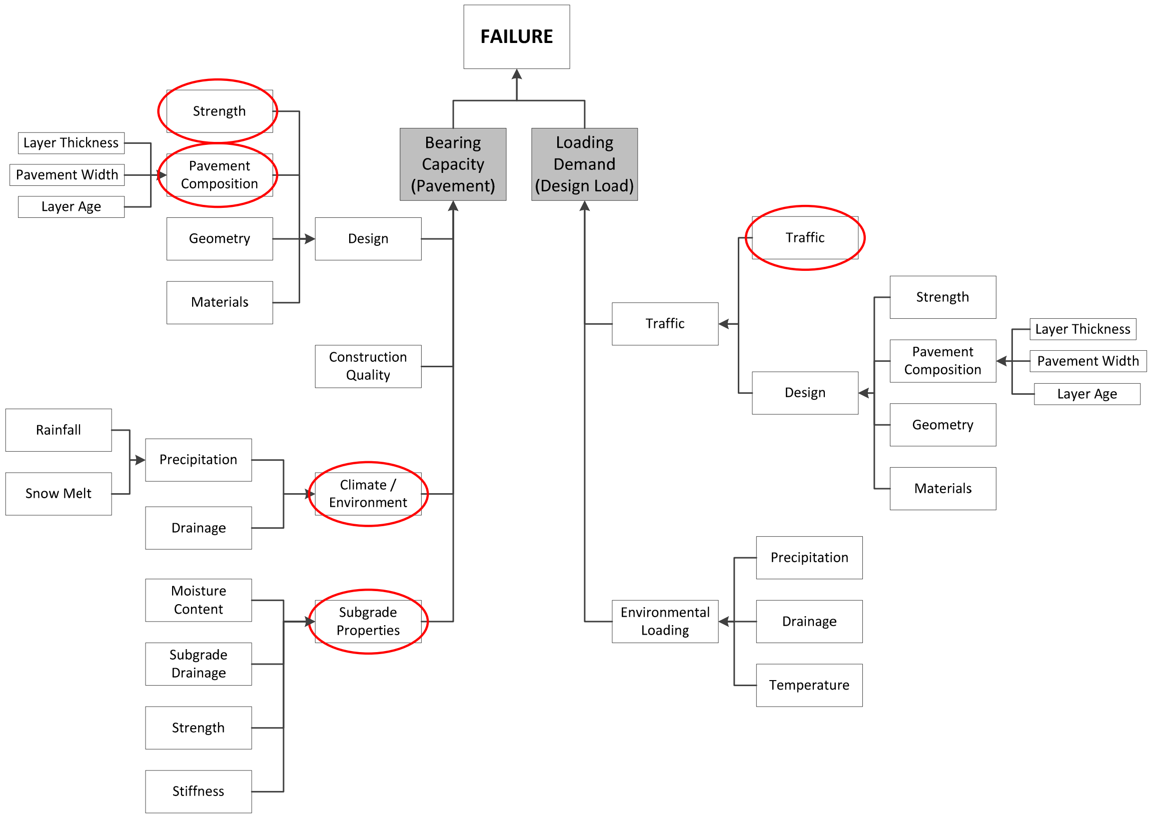


Figure 4-1: Generic failure paths for road pavements

The loading demand branch in Figure 4-1 consists of two paths; one associated with traffic and the other with environmental loading. In the case of the traffic branch, solely traffic exceeding the design loading can cause failure. Alternatively, failure can be caused by a combination of excessive traffic and poor pavement design. The environmental loading branch shows that excessive environmental loading can be caused by cyclic temperatures, precipitation, or poor drainage.

Bearing capacity failure is a result of poor design, the overall quality of the construction of the pavement, climate or environment factors, or issues associated with the properties of the subgrade. These causes can be further broken down into contributing causes as presented in Figure 4-1, for example poor pavement design occurs because of inadequate strength, poor pavement composition, insufficient geometry layout, and / or poor choice of materials.

4.3.1 Failure Factors

Table 4-1 describes the contributing role each of the five identified failure factors (from Figure 4-1) has on pavement failure. Furthermore, surface condition, although primarily considered a measure of failure and not a contributing factor, may also be in fact associated with the occurrence of a particular defect that can induce the development of another defect. For example, severe rutting can also induce pavement surface cracking, yet the primary cause of failure in this case would be rutting (Henning, 2008; Martin, 2008); or vice versa. Therefore, it was considered appropriate to include this factor in Table 4-1.

Table 4-1: Fundamental Factors Associated with Pavement Failure

Factor Groups	Description
Traffic	Pavements are designed and constructed to withstand traffic loadings for a predetermined period of life (design life), such that the pavement dissipates the forces from the induced traffic loadings. However, overloading exceeding that of the design contributes to early failure.
Composition	As part of the design process, the composition of a road pavement is designed to carry an expected traffic loading for the design life. However, under-designed pavements, thin pavements, older pavements, and those that have exceeded their design life fail to dissipate the induced forces prior to reaching the subgrade, thus possibly resulting in failure. Furthermore, narrow pavements and number of lanes contributes to failure.
Strength	The bearing strength of the pavement is an important measure of road pavement performance. A weak pavement will perform insufficiently if under-designed for the given traffic loadings. It can also become susceptible to early failure and environmental changes.
Environment	The climate can damage a pavement significantly. Precipitation, weathering, and temperature have detrimental effects on the performance of the pavement. Water entering the pavement compromises the structural integrity of the lower layers of the pavement. High temperatures affect the performance of the bituminous layer(s) and low temperatures can result in freeze-thaw. The change in the temperature gradient reduces the sealant function of the bituminous layer (e.g. providing a water-tight layer).
Subgrade Sensitivity	Since pavement design aims to minimise the impact of traffic loadings on the subgrade, the susceptibility of the subgrade to deform under either traffic or environmental loadings is a factor involved in the failure of pavements. Such susceptibility is influenced by the strength, stiffness, moisture content, and subsoil drainage of the subgrade.
Surface Condition	The current condition of the pavement surface can give an indication on the type of failure, how advanced the failure is, and the rate of progression of the failure; therefore, surface condition is encompassed in previous pavement performance models. However, there are some cases where the condition data is a secondary defect to the primary cause of failure, such as severe rutting can also result in pavement surface cracking, or vice versa, yet the primary cause of failure in the former case is the rutting.

4.4 Development of the Failure Knowledge and Charts

From hereon, using the concepts of the generic pavement failure chart demonstrated above, the remainder of this chapter focuses on three predominant failure mechanisms seen on New Zealand roads networks, which consist of predominantly flexible, unbound, granular, chipsealed pavements. Accordingly, in order to develop the associated failure charts, the following was carried out (refer Section 4.1.2):

- A literature review of the predominant types of pavement failure on New Zealand roads, namely rutting, cracking, and shear failure;
- The canvassing of expert opinions to identify the causes of multiple failures and interacting factors, and
- An analysis of New Zealand network datasets.

4.4.1 Review of the Fundamental Failure Factors

Rutting Failure:

Rutting is an indication of the deterioration of the structural integrity of the pavement; in other words, the pavement is no longer able to dissipate the forces induced by the traffic adequately. Rutting appears on pavements as depressions in the wheelpaths. There are two main types of rutting; one related to the surface layer and one related to the structural layers of the pavement. This research focuses on the latter, where the pavement presents wide shallow ruts (CSRA, 1992). The impact from traffic loadings causes the lower layers of the pavement to compact under the wheel loads (Papagiannakis and Masad, 2008), although rutting can also be attributed to poor pavement support (CSRA, 1992). Several problems with the underlying pavement layers, drainage, and aggregates contribute to poor pavement support and densification of the materials. Therefore, excessive traffic loadings or problems with the underlying layers of the pavement are the main causes of rutting failure.

Cracking Failure:

There are a number of different types of cracking (refer Section 2.2.2). However, the main concern structurally for flexible pavements with all types of cracking is that the formation of cracks permits the ingress of water into the lower layers of the pavement. Further concerns include the impact on road user costs given the disintegration of the pavement surface, resulting from interconnected cracking, which brings about an increase in the roughness profile of the pavement.

Fatigue cracking is the prevalent type of cracking on New Zealand roads, and is caused by excessive traffic loadings (CSRA, 1992), but it can also be a result of unbalanced pavement layers (e.g. the stiffness of the individual pavement layers is unbalanced[9] across the pavement), poor pavement support, and / or the use of brittle surface materials (e.g. asphalt which has aged excessively) (Henning et al., 2006; Martin, 2008). Narrow carriageways or those without shoulders, common in rural areas of New Zealand, are also prone to edge cracking, resulting from a lack of edge or shoulder support from the pavement, given the shoulder is generally designed weaker than the pavement itself (Thom, 2008).

Joint or imbalanced cracking does not often occur on the roads in New Zealand; however, most of this type of cracking is caused by a stiff upper layer (Henning, 2008), and usually occurs in stabilised pavements. Other causes include a brittle surface material, thermal cycles, and moisture changes.

[9] Where there is an excessive accumulation of stress between successive pavement layers.

Thermal cracking results from temperature changes in the environment but can also be caused by surface shrinkage, reflective cracks originating beneath the surface, and top down cracking (Thom, 2008).

The other common type of cracking in New Zealand is longitudinal, transverse, or block cracking. The causes of this type of cracking is similar to other types of cracking, such as top down cracking, brittle materials, reflective cracking from underlying layers, and shrinkage. In addition, longitudinal segregation and construction techniques, such as joints or widening projects, can cause longitudinal, transverse, or block cracking (Henning, 2008).

Shear Failure:

Shear failure commonly manifests itself on the pavement surface as potholes or deformations and, unlike rutting, can occur anywhere on the pavement section. Generally, as a result of seasonally changing zones of moisture in the base layers, shear failures are prevalent on or near the edge of the pavement (Emery, 1992; Schlotjes et al., 2009; Thom, 2008) (see Figure 4-2 below). Like cracking, such defects can allow water to enter the pavement.

A weak basecourse or underlying layers cause shear failures, due to layer compression, causing depressions throughout the pavement. Inadequate material properties, insufficient shoulder support, and material shear can also cause shear failures (Schlotjes et al., 2011). This type of failure is usually not a direct result of excessive traffic on low volume roads (Schlotjes et al., 2011); however, traffic loadings can further exacerbate it. Insufficient shoulder support permits shoulder failure and subsequently results in failure near the edge of the pavement.

4.4.2 Expert Knowledge from the Industry

Canvassing the opinions of experts provided this study with additional knowledge of road pavement failure. Asset managers from New Zealand RCAs and a pavement specialist highly regarded in the industry were consulted in this stage to assist with defining the interactions between factors, infrequent and site-specific failure causes, the occurrence of multiple failure mechanisms, and maintenance treatments. It is common for multiple factors and / or failure mechanisms to occur simultaneously, thus making the correct failure mechanism(s) difficult to diagnose (Schlotjes et al., 2011). Consulting with experts allowed such elements of failure to be better understood and defined, and therefore the knowledge from the experts was incorporated into the failure charts (Figures 4-7 to 4-11).

4.4.3 Preliminary Data Analysis

This section summarises the results of the preliminary data analysis (see Appendix B). The objectives of the preliminary data analysis were:

- To identify the additional relationships between failure and potential causes of road pavement failure;
- To identify apparent trends on New Zealand road networks, and
- To identify site-specific factors.

The above relationships, trends, and factors were not immediately identifiable in the literature. However, the results of the descriptive data analysis (see Appendix B) suggest the importance of defining such associations. Given the variation in the geography and environments of New Zealand, the inclusion of site-specific factors and apparent trends on New Zealand road networks ensured this research produced comprehensive failure charts, including the factors influencing pavement failure and their associated interactions for flexible pavement failure.

4.4.3.1 Data Analysis Methodology

The data analysis used simple descriptive statistical tools, included in the *R* statistical package (Everitt and Hothorn, 2006), to assess the relationships between the variables in the datasets considered (refer Table 3-1). Smoothing the trend line when examining the relationships between the variables was required to remove the noise associated with some variables and provide clarity to the exercise. In these cases, to avoid confusion, the smoothed trend line is presented in the figures below as a red line, unless otherwise stated.

Two independent datasets were analysed:

- Southland District Council (SDC) road network data, and
- Long-term Pavement Performance (LTPP) programme data.

The SDC is a local authority at the southern base of the South Island, New Zealand. As the RCA for the Southland district, the SDC manages a network of approximately 6,450 kilometres of pavements with 2,770 kilometres of these pavements being sealed (Land Transport New Zealand, 2008). These roads are typical of network hierarchical level and omit any State Highway roads located in the Southland District. The SDC dataset includes network data typical of all RCAs in New Zealand and, although the data was not as detailed in comparison to the research-based LTPP datasets, the SDC dataset is one of New Zealand's more suitable local authority datasets, with a completeness of inspection data and an extensive and inclusive history of maintenance and inspections. The available RAMM data was able to be collated and manipulated accordingly to obtain variables similar to those described in Table 3-1 and included in the LTPP datasets.

Unfortunately, the SDC dataset lacked information required to determine advanced relationships, such as those involving detailed pavement inventory data. Therefore, the LTPP datasets, described in Section 3.3, were analysed to ensure such detailed relationships were included in the development of the failure charts. The two LTPP datasets (State Highway and Local Authority) were analysed separately.

4.4.3.2 Rutting Relationships

Strong correlations emerged between traffic loadings and the base layer properties of the pavement with rutting failure respectively, such as greater rut depths were reported as the traffic (AADT and HCV) increased and on older pavements. Although the SDC dataset showed less robust relationships between rutting and traffic than the LTPP dataset (see Appendix B), traffic loadings and base layer properties were concluded as predominant factors in the rutting failure mechanism.

However, negative straight-line correlations were found to exist between high speed rutting condition data and pavement width in the LTPP dataset, particularly concerning the left-hand wheelpaths. This suggests the width of the pavement influences the rut depths on the pavement, as narrower pavements concentrate the traffic into distinct wheelpaths resulting in less traffic wander across the pavement and a greater amount of deterioration in the wheelpaths. In addition, the State Highway sites showed greater rutting deterioration than the Local Authority sites, which is understandable, considering the difference in design life, traffic loadings, and compositions between these two networks. Conclusive relationships concerning rutting in the right-hand wheelpath were evident in this analysis, in comparison to the left-hand wheelpath, although it should be noted that the reported rutting depths were greater in the left-hand wheelpath. Figure 4-2 demonstrates how the seasonal variation in moisture directly underneath the left-hand wheelpath impacts on the robustness of the correlations relating to rutting in the left-hand wheelpath (Emery, 1992).

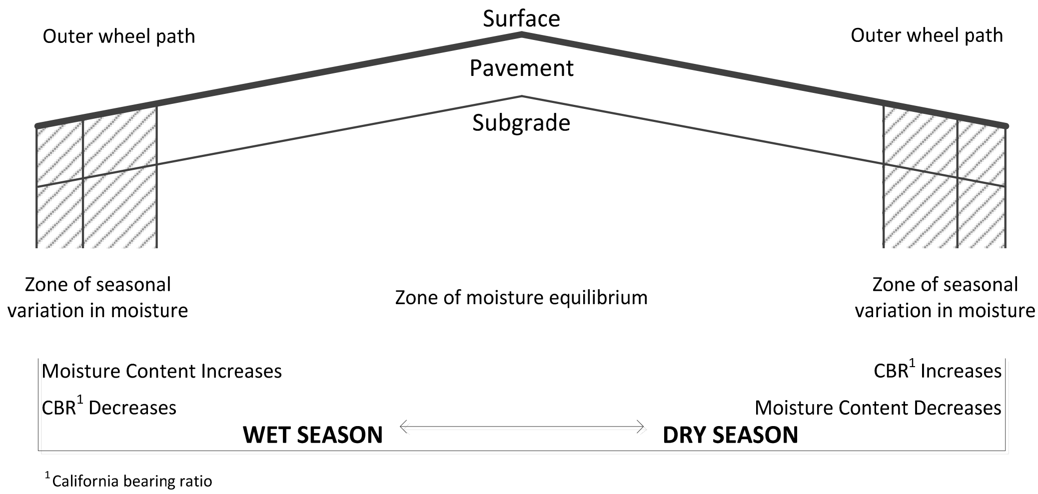


Figure 4-2: Outer zones of moisture variation below the pavement (adapted from Emery, 1992)

Furthermore, by visual inspection of the smoothed relationship (refer red line in Figure 4-3), pavements with cracked surfaces, predominantly alligator cracking, showed greater rut depths than those without cracked surfaces; however when other cracking types were explored, there was no distinct differences in rutting depths between these two surface conditions. From this, alligator cracking can be seen to aggravate the progression of rutting, most probably as a result of water ingress into the lower pavement layers further deteriorating the pavement's structural integrity (Thom, 2008).

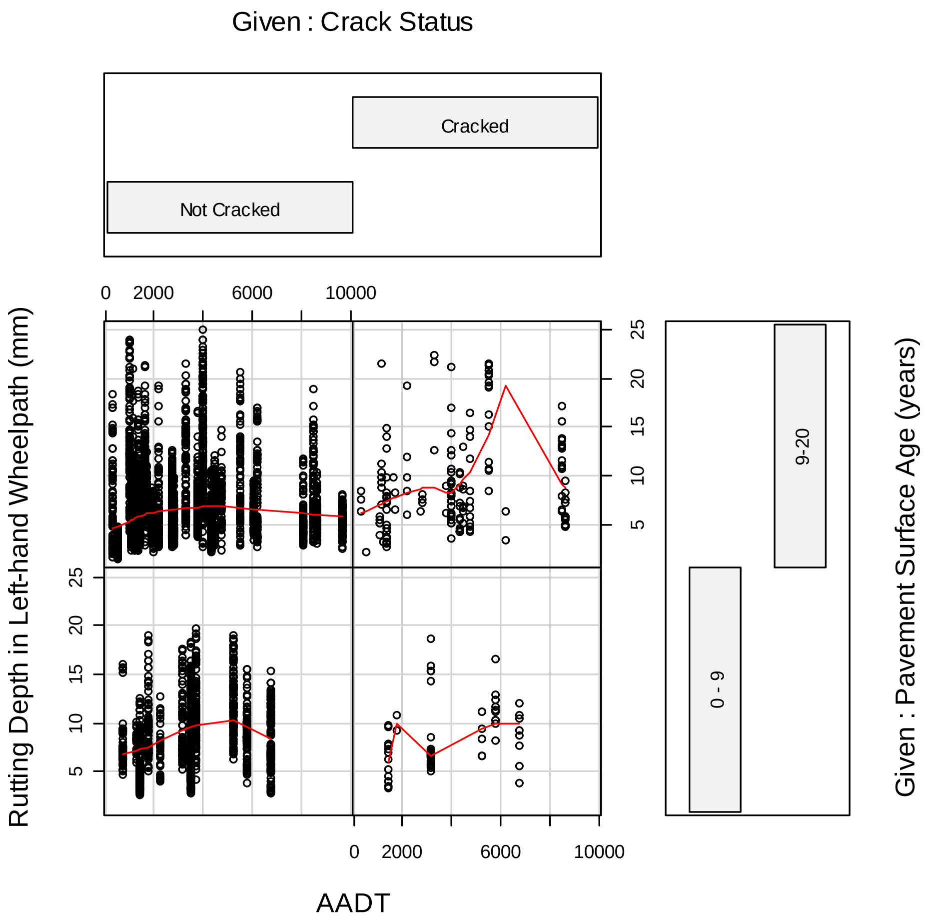


Figure 4-3: The influence of cracked surfaces on rutting depths, based on alligator cracking only, for the State Highway LTPP road network

From correlation plots, greater rut depths were visible on older pavements compared to younger pavements, and a rapid increase (shown by the red trend line associated with the positive rut progressions in Figure 4-4) in the rut depths were observed as older pavements aged further. The blue trend line in Figure 4-4 presented little information on the relationship between negative rut progressions and the age of the pavement. However, on young, weak pavements that carried increased traffic volumes, the pavement was found to be more susceptible to rapid rutting deterioration, yet older pavements did not. Interestingly and unexpectedly, thicker pavements (100 – 1000 mm deep) in the LTPP datasets displayed greater rut depths than thin pavements (0 – 100 mm thick) (see the smoothed red trend lines in Figure 4-5), although this may be a direct result of the pavement design. Capturing this anomaly in the failure charts poses difficulty; instead, this research included pavement layer thickness as an influential rutting failure factor.

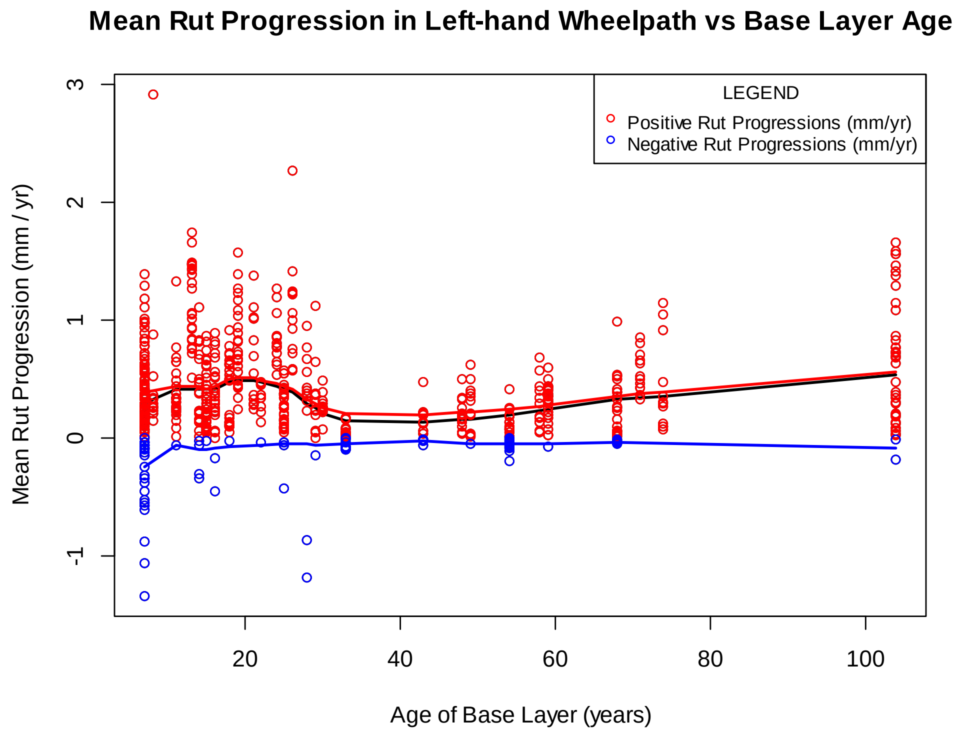
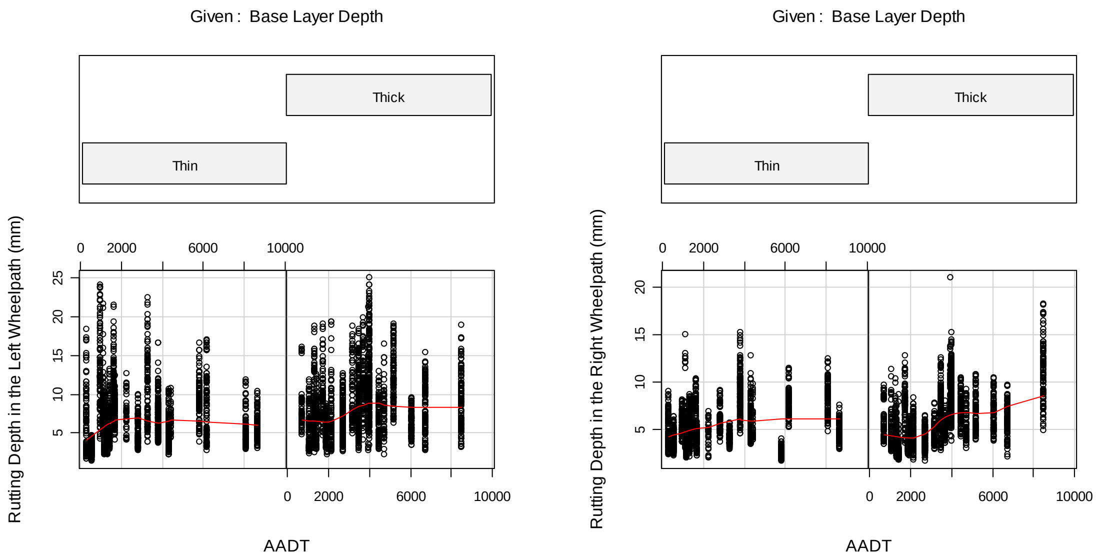


Figure 4-4: Aged pavements in the State Highway LTPP road network showing greater rut progressions



(a) Left-hand wheelpath rutting **(b) Right-hand wheelpath rutting**

Figure 4-5: Greater rutting depths on thicker pavements in the State Highway LTPP road network

From the above, it may be seen that the following factors influence rutting failure:

- Pavement width;
- Traffic volumes and vehicle classifications;
- Seasonal moisture zones which occur below the left-hand wheelpath;
- Cracked surfaces;
- Water ingress due to breaks (cracks) in the pavement surfaces, and
- Base layer properties, namely the age and thickness of this layer.

4.4.3.3 Cracking Relationships

The correlation plots from the analysis on the SDC dataset suggested that the cracking reported on the pavement was predominantly caused by traffic related factors, including traffic loadings. The SDC dataset provided limited information on the different types of surface cracking; however, a visually apparent relationship was shown between transverse cracking and the age of the pavement surface in the LTPP dataset, as shown in Figure 4-6.

The analysis of the LTPP data did not show all of the typical relationships expected between the occurrence of alligator cracking and the factors that the literature suggested may cause such defects. For example, no conclusive relationship could be found between the age of the pavement's surface layer and alligator (fatigue) cracking. Furthermore, alligator cracking was observed to be more severe across the whole pavement than that observed to occur specifically in the wheelpaths, which contradicts the expectation that repetitive traffic loadings can cause alligator cracking. Therefore, it may be concluded that in low volume roads the predominant causes of alligator cracking is environmental loading or the properties of the binder of the bituminous layer.

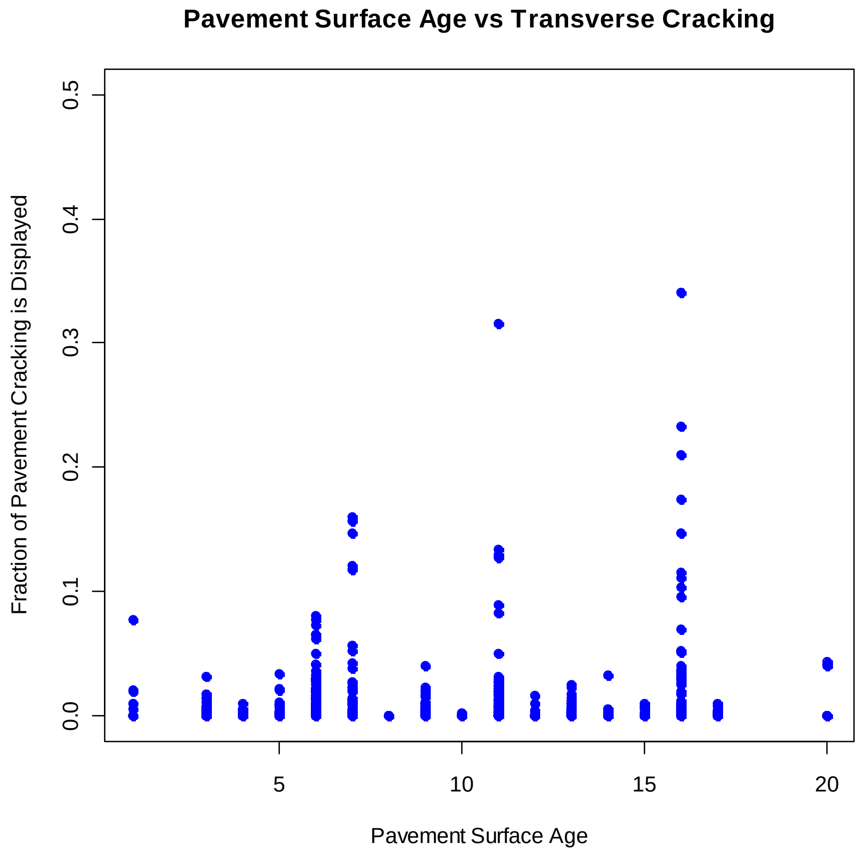


Figure 4-6: Relationship between transverse cracking and surface age on pavements in the Local Authority LTPP road network

4.4.3.4 Shear Relationships

Conclusive relationships were difficult to establish between the manifestations of shear failure and parameters that the literature would suggest are influential, due to a lack of categorical material properties in the datasets, such as information on aggregates used in construction of the road pavement. However, the SDC data confirmed that the narrower pavements reported a greater percentage of shoving and number of potholes on pavements, which further substantiated the literature detailing the relationship between shear failures and pavement width (Emery, 1992; Thom 2008).

4.4.3.5 Summary of Preliminary Data Analysis

The findings from the preliminary analysis reported above were used in the development of the failure charts and are summarised below:

- Traffic factors (loading volumes and vehicle classes) were main contributors to rutting and cracking failure;
- Pavement width influenced both rutting and shear failures;
- Clear and conclusive relationships concerning rutting in the right-hand wheelpaths, as a result of less variation in the data, were presented, suggesting a cause of rutting in the left-hand wheelpath is moisture ingression from surrounding soils and seasonal variations in the moisture zones;
- Cracked surfaces permitted water ingress and resulted in greater rut depths;
- The age and thickness of the base layer impacted on the depth and progression of rutting, and
- Cracking was not directly related with the age of the pavement, yet showed potential correlations with the surface age (refer Figure 4-6).

4.4.3.6 Additional Notable Findings

Other notable findings from the analysis of the datasets are summarised below:

- State Highway pavements in the LTPP data were of a higher strength than the Local Authority pavements, which was expected given the higher traffic volumes on the State Highway network (see Appendix B), and
- The design of the State Highway and Local Authority pavements differed as a result of the diverse road environments and traffic loadings of these pavements.

Given this, opposing correlations were drawn from the analysis of the two LTPP (State Highway and Local Authority) datasets. Although, in these cases, this created difficulties in defining conclusive relationships between failure and their respective causes, the research datasets provided valuable knowledge to the failure charts, particularly relationships involving pavement composition, traffic, and environmental (water ingress) factors.

4.5 Developed Failure Charts

The failure charts presented in Figures 4-7 to 4-11 have been developed, following the principles of typical FTA[10] (Lazzaroni et al., 2011) for creating fault trees, for rutting, cracking, and shear failure using the engineering knowledge ascertained from the above analysis, as per Section 4.4, and included:

- Common causes of failure reported in the literature;
- Opinions from pavement experts, and
- The results from the analysis of the LTPP datasets and SDC dataset, where additional causes of failure were established.

The presentation of the failure charts follows that of Figure 4-1 (refer Section 4.3) where the factors involved in pavement failure are sequential, based on their causative nature. Figures 4-8 to 4-10 show the five predominant types of cracking (fatigue, edge, joint or imbalanced, thermal, longitudinal or transverse or block cracking). Based on the FTA approach (refer Section 4.2), the failure charts present a combination of factors, indicating that failure is not

[10] Such principles for the development of fault trees include defining the problem, setting rules for the constituents of the failure on whether the causes of failure can occur simultaneously or distinctly, and event ordering.

always a result of a single factor as shown through the use of multiple and successive branches. These combinations represent the failure paths, which in turn identifies the predominant causes of failure. The individual failure paths consider the interactions between the failure factors, such that the interdependent nature of the factors involved in pavement failure can be represented in the computational model (see further Section 4.6).

For example, in Figure 4-7, rutting failure can be solely attributed to excessive traffic loadings, or attributed to a larger number of factors stemming from poor pavement support under the induced (design) traffic loadings. However, poor pavement support is a product of a weak basecourse, arising from water ingress as a result of inadequate drainage for example. Therefore, the cause of rutting in this case is inadequate drainage which, defined by the critical failure path, is a result of water ingress weakening the basecourse and ultimately resulting in rutting.

Examples of rutting not directly caused by traffic loadings are shown under the deformation branch in Figure 4-7, which can be further split into defects contained in the structural layers of the pavement. These defects can either be associated with the base layer or the subgrade. The base layer problems result from inadequate pavement design, quality of the construction and materials, excessive traffic loadings, or narrow carriageways. The problems with the subgrade result from inadequate pavement design, the design life of the pavement above the subgrade has been exceeded, excessive traffic loadings directly affecting the subgrade, or excessive moisture content in the subgrade to cause deformations to the pavement. These defects are further defined, as shown in Figure 4-7. Similar approaches to failure for cracking and shear failure are presented in Figures 4-8 to 4-11.

It is recognised that failure could occur with failure modes occurring simultaneously (for example rutting could occur at the same time as cracking). Such interdependencies have not been represented directly on the failure charts to preserve the clarity of the two-dimensional failure paths. Nevertheless, the developed computational model is capable of dealing with such interdependencies, as discussed in Section 6.5.

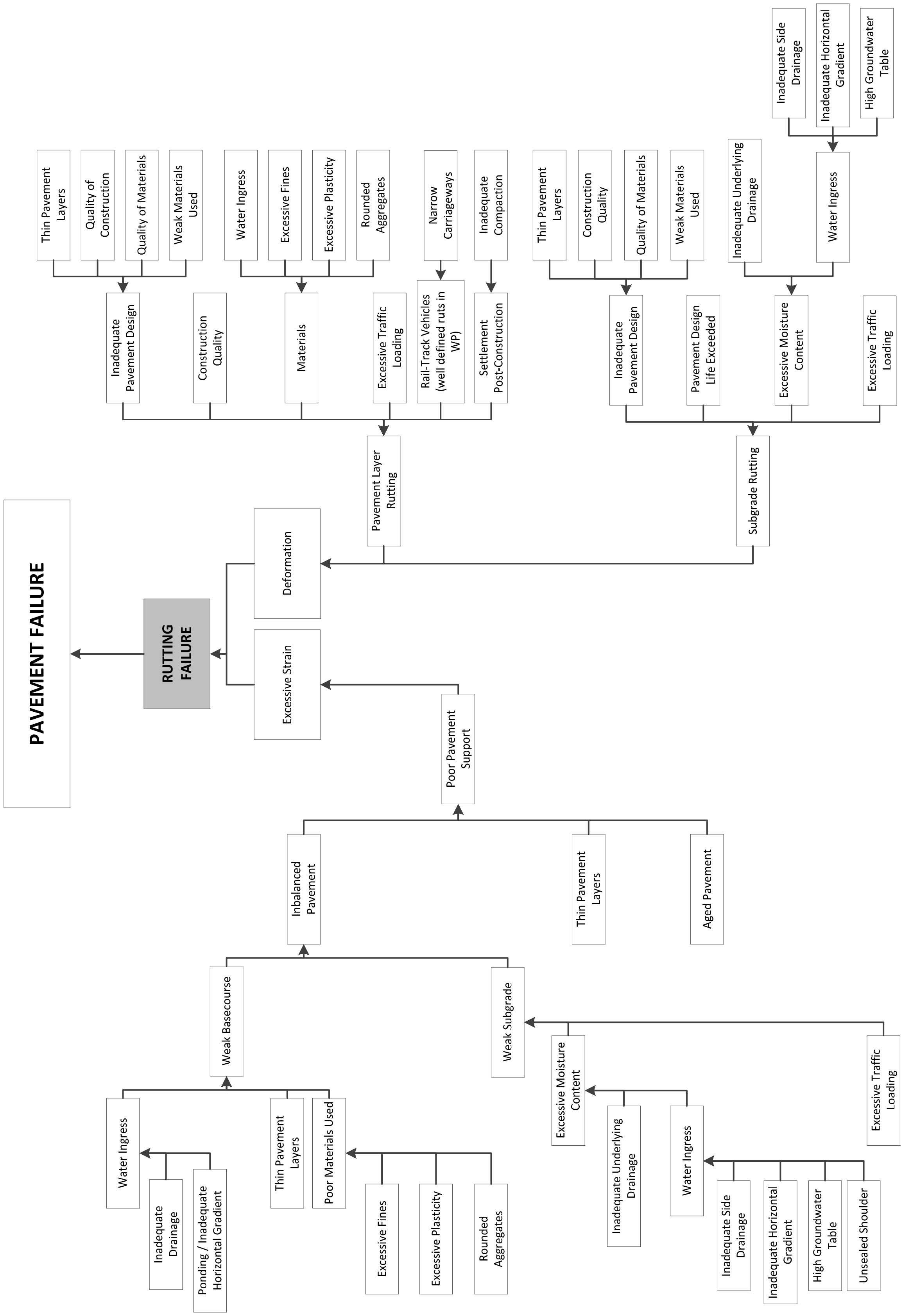


Figure 4-7: Failure chart of rutting failure

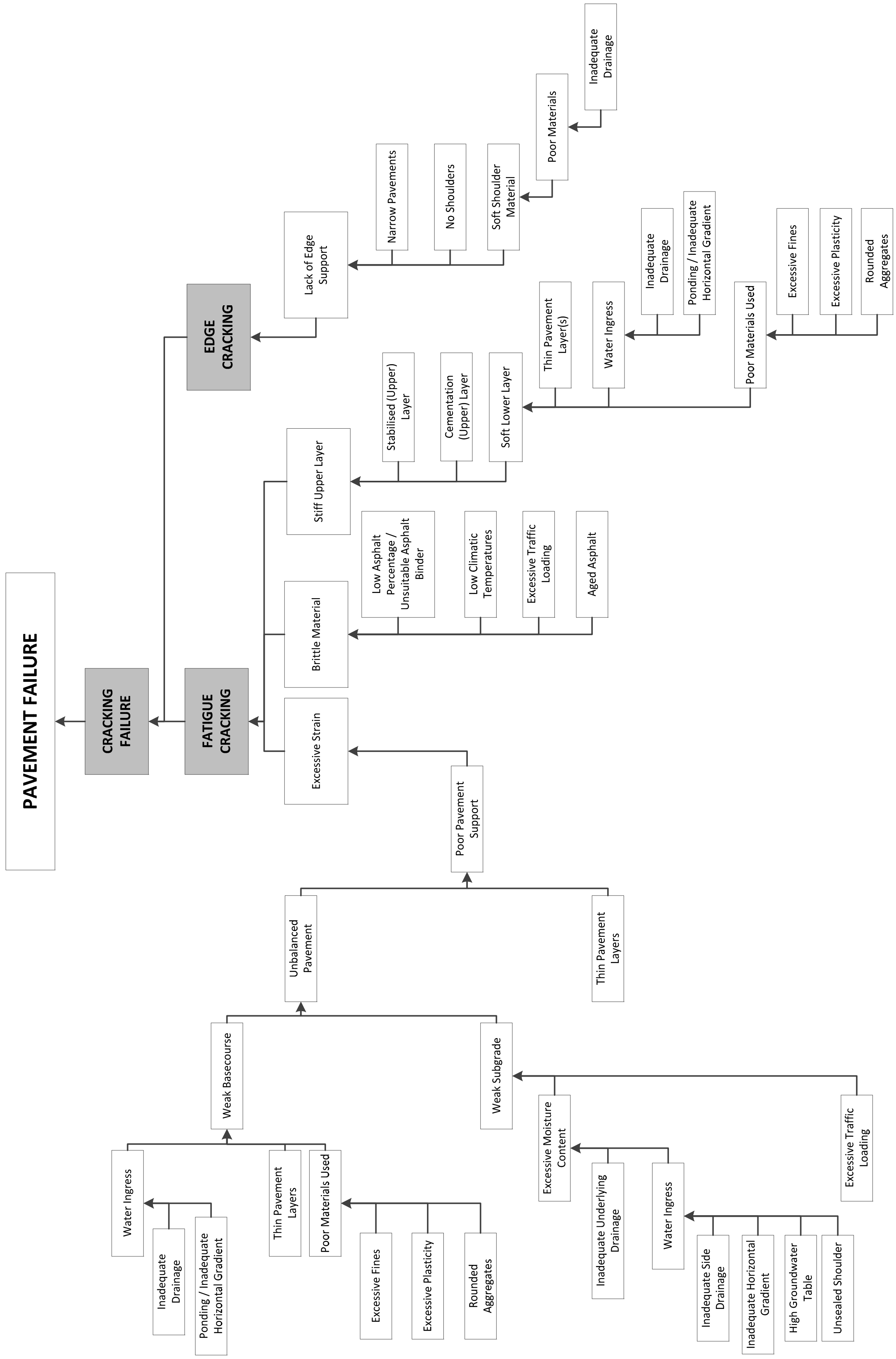


Figure 4-8: Failure chart for cracking failure - part 1

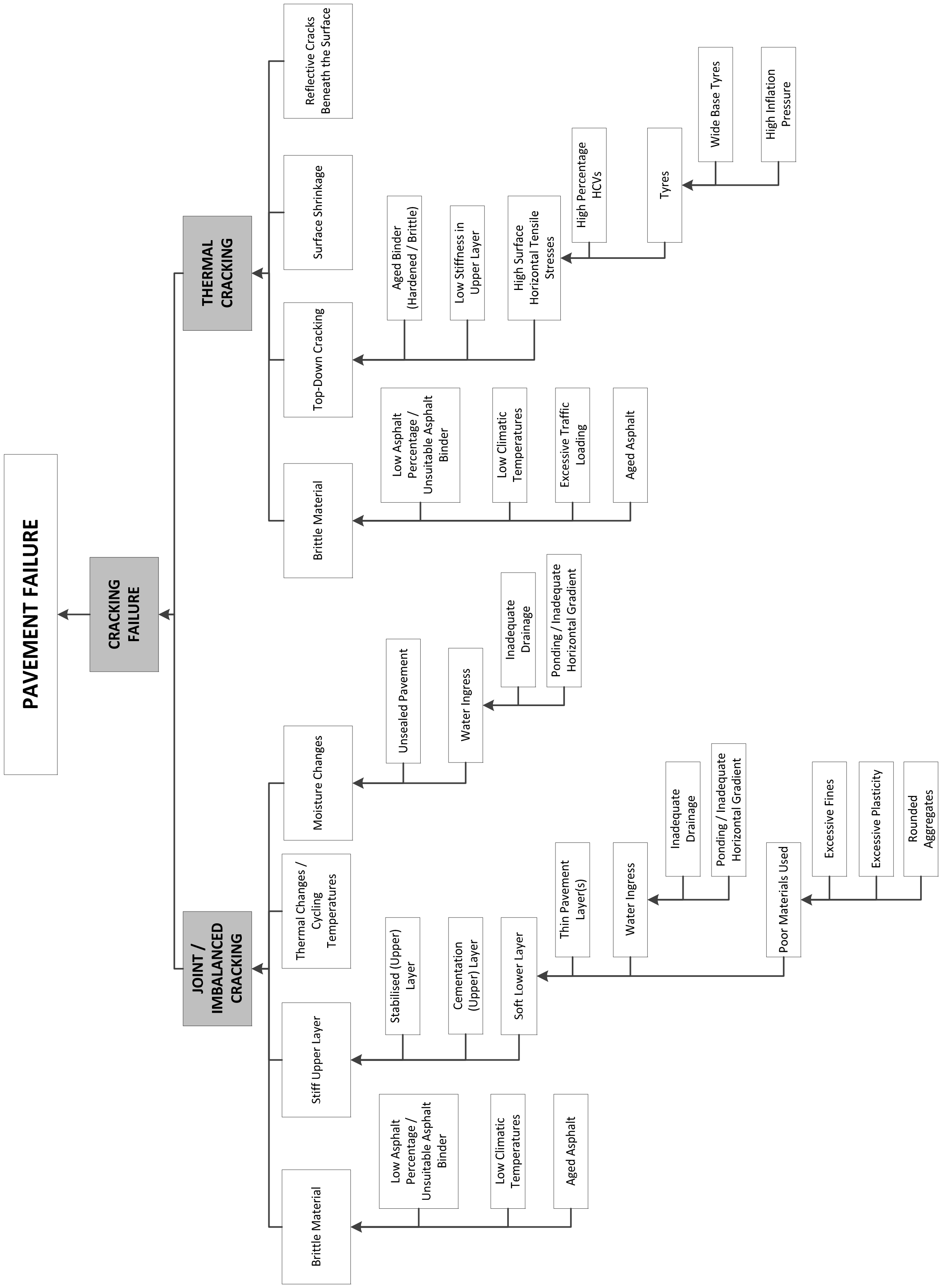


Figure 4-9: Failure chart for cracking failure - part 2

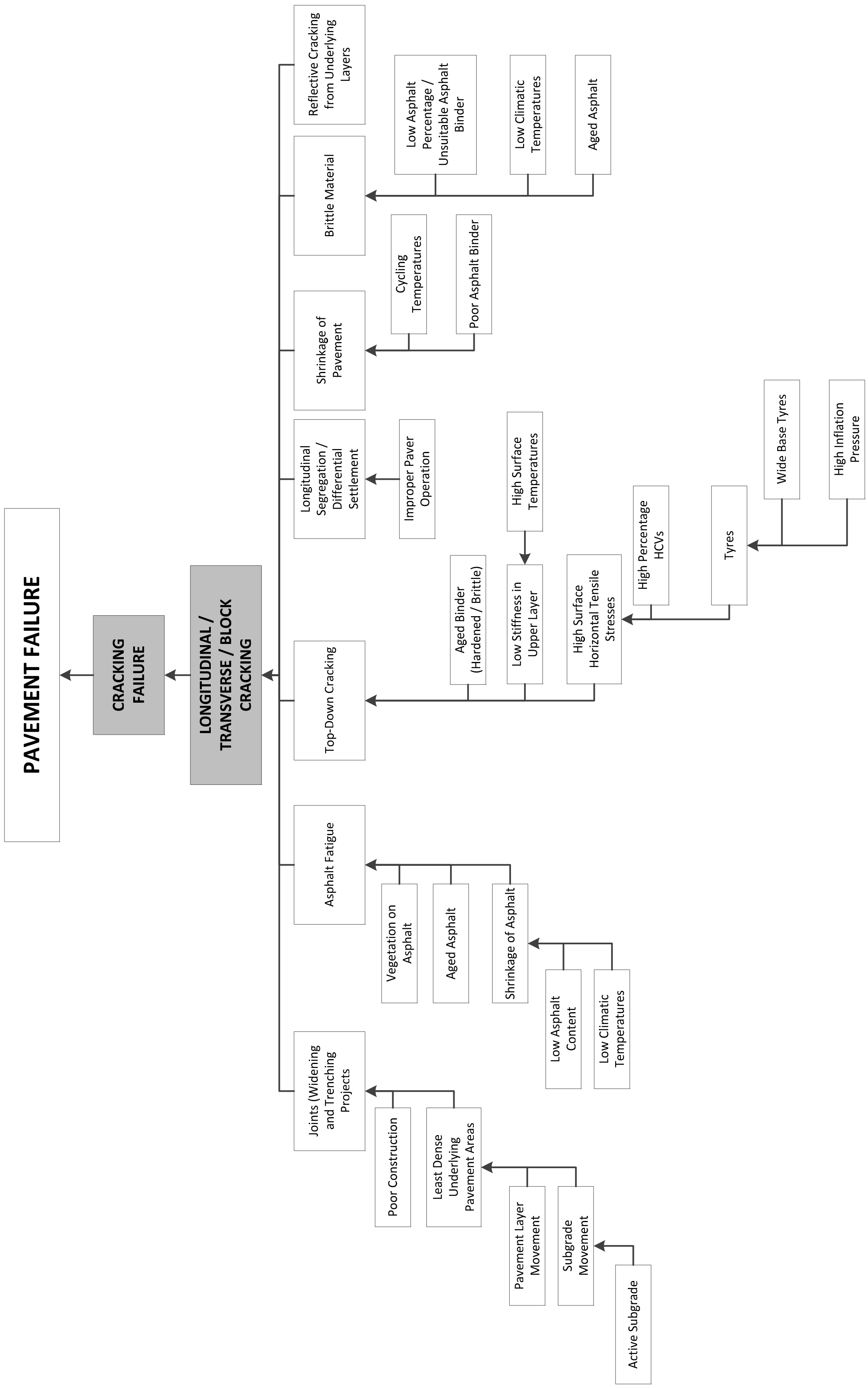


Figure 4-10: Failure chart for cracking failure - part 3

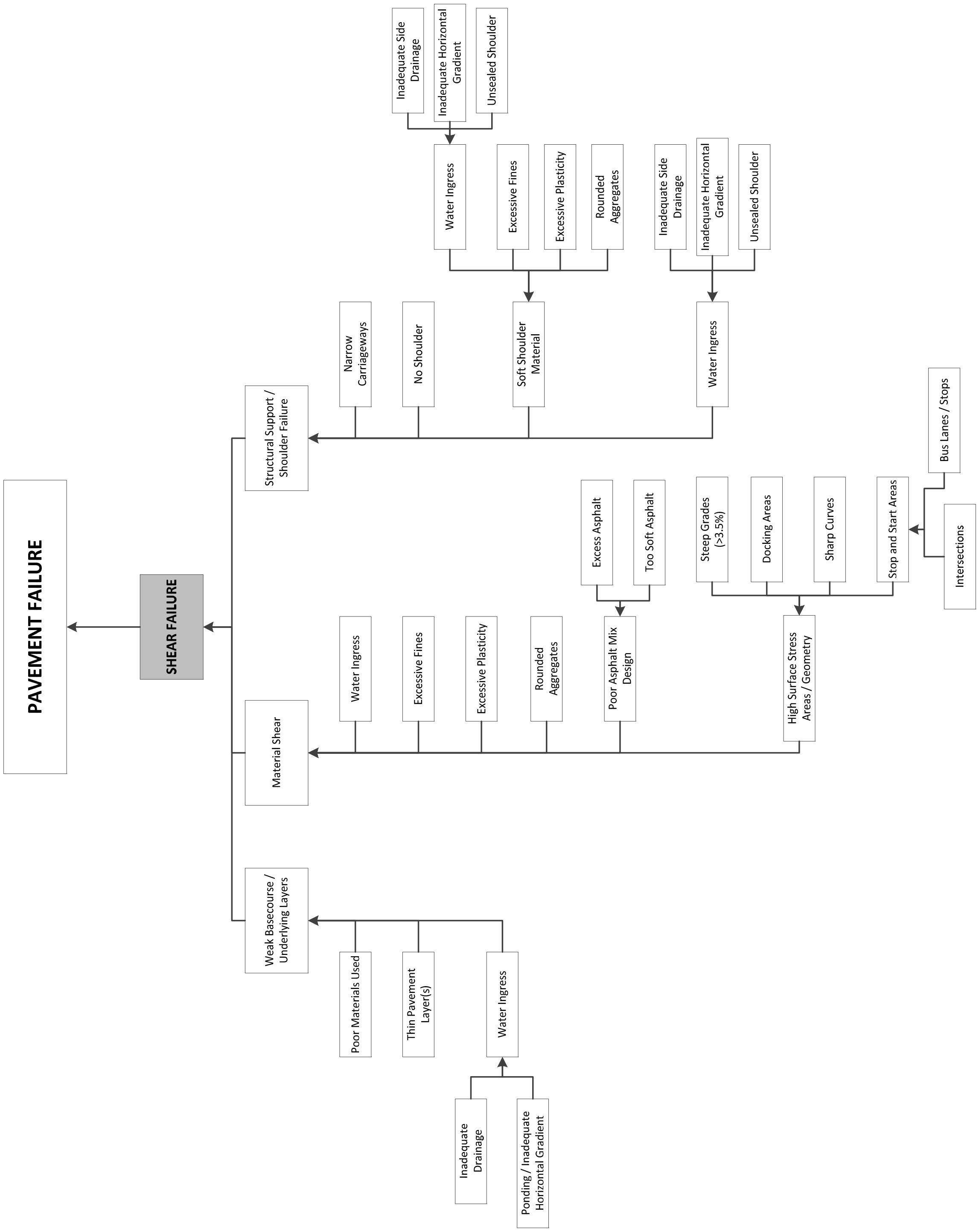


Figure 4-11: Failure chart for shear failure

4.6 Future Use of the Understanding in this Research

Figure 4-12 illustrates conceptually how the three sources of knowledge captured in the form of failure charts is used in this research to diagnose the cause of pavement failure and subsequently to assist in determining the probability (likelihood) of failure. Given that the five failure factors (traffic, composition, strength, environment, and subgrade sensitivity) are embedded in each failure chart, the combinations of these factor groups are presented as the failure paths involved in each failure mechanism, per failure chart, and therefore represent the interactions between the causes of failure for each failure mechanism. Each of these factor combinations will be modelled (see Chapter Five), and the successful factor combinations will be compared against the failure charts to ensure the inference of engineering knowledge in the computational model (refer Section 3.1.1). Only the successful factor combinations associated with the respective failure mechanism will be included in the development of the prototype system (see Chapter Six). With this approach, the combinations taken forward in this research will represent each failure path presented in the developed failure charts, dependent on the data present in the research dataset(s).

As the most probable failure path is predicted by the system (refer Section 3.3.4 and further Section 6.2.1), the knowledge captured in the failure charts is used to further diagnose the causes of failure. This process is further discussed in Chapter Seven.

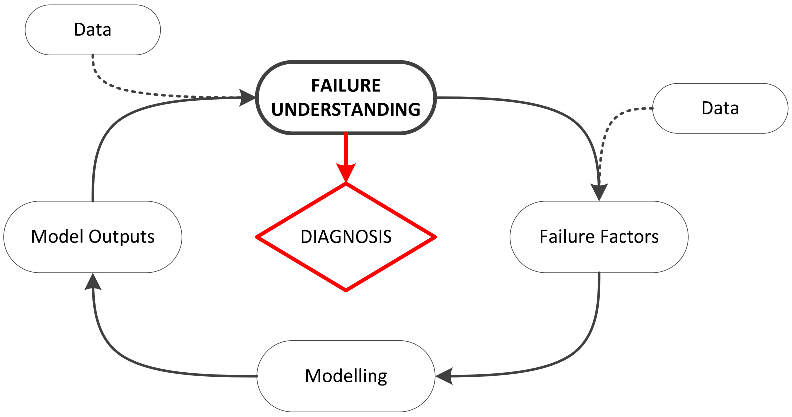


Figure 4-12: Employment of the understanding

4.7 Summary of Understanding Road Pavement Failure

This chapter discussed the importance of, and the process associated with, fully understanding road pavement failure. To this end, this research developed a methodology of collating engineering knowledge to predict the likelihood of pavement failure. Such knowledge was obtained:

- From the literature;
- Through canvassing the opinions of experts, and
- Via a preliminary analysis of two road datasets.

Five groups of factors were identified in this chapter that influence pavement failure, which could be considered generic for sealed road pavements. This chart was then expanded and further developed using the FTA approach to describe rutting, cracking, and shear failures (Figures 4-7 to 4-11). This process was further informed by the analysis of two datasets obtained on New Zealand roads, ensuring that site-specific failure knowledge was captured in the failure charts.

Although the focus of these charts is on flexible pavement failure, the methodology developed (Section 4.1.2) can be followed for other pavement types.

The failure charts developed, and the engineering knowledge which they capture, will be used to inform the development of the computational model which will ultimately be used to determine the probability of pavement failure from road datasets. The choice and development of these models is the focus of the remainder of this thesis.

Chapter Five

A COMPARATIVE STUDY OF CLASSIFICATION TECHNIQUES

Which Technique is Best?

5.1 Introduction

To address Objective Two of the research, this chapter presents a comparative study of a number of classification techniques suitable for the prototype system together with an objective process for selecting the most appropriate modelling technique for the prototype system. The literature review (Chapter Two) identified a number of modelling techniques that have been previously used in pavement performance studies primarily in deterioration modelling. However, given the definition of failure in the research dataset (refer Section 3.3.1.7), a range of studies successfully employed classification techniques to solve binary problems, and these studies were considered in the development of this chapter.

The purpose of the comparative study is therefore to comparatively assess the performance of the classification modelling techniques identified in Section 2.6. To do so, objective-based criteria, which included interpretability, inference of engineering knowledge, and performance elements, are used to select a number of suitable classification techniques for the task at hand from the methods reviewed in Section 2.6. These techniques are discussed in detail and their performance using the research dataset is evaluated.

Utilising a second criteria, the techniques are ranked based on their performance, interpretability and usability, and model generalisation to determine the most appropriate method for use in the prototype system described in Chapter Six.

5.1.1 Methodology for Selecting the Appropriate Technique

The methodology presented in this chapter was divided into three sections, as shown in Figure 5-1:

1. Select modelling techniques from the available techniques identified in the literature review using the first criteria, which may be considered initially suitable for the use in the prototype system;
2. Evaluate the performance of each technique using appropriate performance measures, and
3. Rank each technique, using an additional set of criteria, to determine the most appropriate modelling technique to further develop.

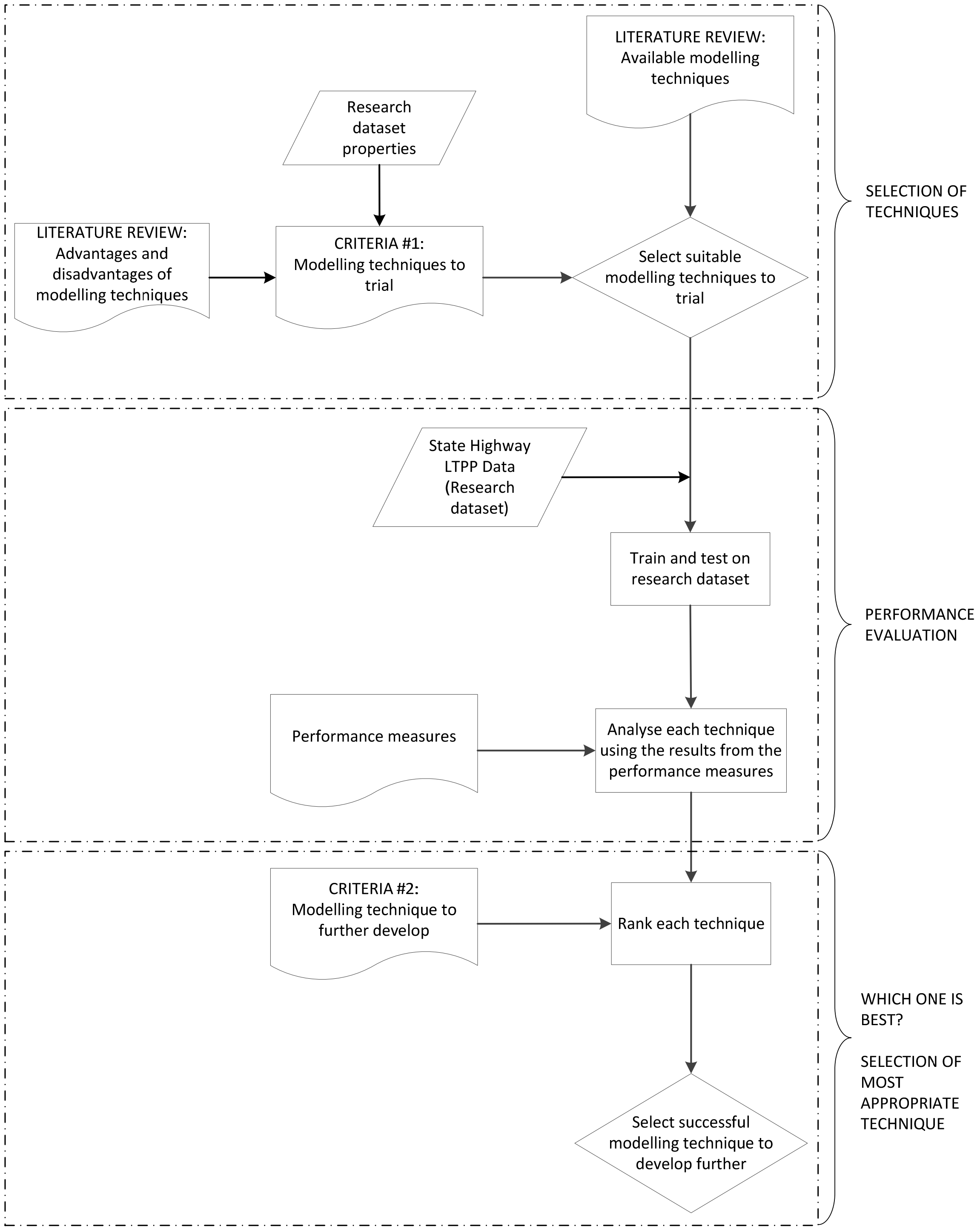


Figure 5-1: Methodology to select most suitable modelling technique

The first criteria referenced the advantages and disadvantages of each classification technique reviewed in Section 2.6 (refer Table 2-3). The criteria included the performance of the technique, user interpretability, short running time, and specific criteria to meet the needs of road network asset managers, such as the inference of engineering knowledge into the model, non-linearity properties, the use of the model with a large number of input variables, and non-stochastic model properties.

To evaluate the performance of each technique, performance measures were employed. Accuracy may be seen as a logical measure of the performance of a modelling technique; however, it is often characterised as a poor performance measure when used alone, due to its biasness towards the results (Ben-David, 2008; Parker, 2011). Much of the literature still used accuracy to evaluate the techniques, given its simple calculations, often alongside other performance measures. Therefore, this research employed accuracy alongside other point performance measures, which are further described in Section 5.3.2, and listed below:

- Accuracy;
- Misclassification error;
- F-score, and
- Phi coefficient.

The second criteria ranked each of the suitable techniques according to their performance in three categories:

- Model performance based on the research dataset;
- The speed, usability and interpretability of the technique, and
- Generalisation of the model to ensure the technique is not overly complex to the detriment of the predictions and does not over-fit.

This criterion above is weighted based on the importance of each category to the implementation of the overall prototype system in the asset management industry of road pavements.

R statistical software (Everitt and Hothorn, 2006), previously used in Chapter Four, was employed to construct each classification model.

5.1.2 Dataset for the Study

The State Highway LTPP dataset described in Section 3.3.1 was used in this chapter. Forecasting the future probability states of the LTPP sites was not addressed in this study; instead, to increase the amount of data available to train the modelling technique with, the historical data collected in the LTPP programme was used inherently as single entries resulting in a total of 4512 datapoints in the research dataset. Table 3-1 presented earlier details the variables from the State Highway LTPP dataset to be included in each of the models, per factor group identified in Chapter Four (refer Table 4-1). It should be further noted that the comparative study herein focuses only on the rutting, fatigue cracking, and shear failure mechanisms.

5.1.2.1 Data Manipulation

For success in modelling and to remove any discrepancies in the research dataset, the State Highway LTPP dataset was manipulated consistently across all variable fields prior to the assessment, using the following approaches:

1. **Missing Values:** Where missing values occurred in the dataset, a nominal value of zero (0) was assumed. This represents the worst possible scenario for the road pavement consistently across the majority of the missing fields in the State

Highway LTPP dataset. For this study, this assumption was considered sufficient; however, other methods of assigning values to missing fields are discussed in Chapter Eight (see Section 8.4.2.1);

2. **Normalisation:** The maximum and minimum values of the independent variables in the dataset differ significantly. For example, the traffic data ranges from between 100 and 10,000 vehicles per day, while the pavement age variable has values which are seldom above 100. Therefore, to avoid biasness between the independent variables and minimise data redundancy within the model, each independent variable was normalised using a straight-line transformation. Adopting this assumption often carries concerns given road pavement data seldom follows a normal distribution (further discussed in Section 8.4.2.2); however, the preliminary analysis of the State Highway LTPP dataset did not indicate a distortion in the distributions of the independent variables (see Appendix B) and, therefore, this assumption will not impact negatively on the purpose of this comparative study, and

3. **Weighting Factor:** Failure is avoided on road networks by the implementation of proactive maintenance strategies to maintain the integrity of the network. As a result, data representing sound road sections in road network datasets occur far more commonly than data representing failed sections, resulting in limited failure data with which to train a computational model. To address the issue of imbalance between reported failures and sound pavement sites, this research applied a weighting factor based on a ratio of the occurrence of failures to non-failures in the

State Highway LTPP dataset. This is further discussed in Chapter Eight (see Section 8.4.2.3).

5.2 Selecting the Classification Techniques

For the comparative study, suitable classification techniques were selected using the first criteria, based on the merits and limitations of both discriminative and generative techniques (refer Table 2-3). Section 2.6 reviewed the literature where these techniques have been used as a binary classifier. The first criteria inferred non-technical aspects of the methods with the reported performance of each, as follows:

1. The overall prototype system must accurately classify the occurrence of failure for road pavement data. The performance of the technique is subsequently involved in the choice of the most suitable technique for further development and, therefore, this comparative study aims to trial the better performing classification techniques. Thus the success and performance of the classification technique in binary studies must be well documented in the literature despite the research domain;
2. A number of factors can be involved in the failure of road pavements and, as a result, an extensive amount of data is associated with road networks. The size of the dataset is greatly influenced by the size of the road network; thus the modelling technique must be able to cope with all sizes of datasets with an extensive number of external variables;
3. Chapter Four discussed the complexity of road pavement failure, particularly the complex interactions between failure factors (Reigle, 2000) and the difficulty of replicating such interactions with computational models. It is not expected a linear

model function will accurately replicate such behaviour, thus pavement failure will be better represented with non-linear model forms. Non-linear modelling techniques should be favoured in this study; however, if required, the use of linear techniques should be with caution;

4. The computational efficiency of the model is imperative for the implementation in RCAs. The time required for the modelling technique to output predictions should be minimal (e.g. quick to run);
5. Furthermore, the training process of the model should be simple, as well as easily inferred and based on the occurrences in the data (e.g. no randomness in the model) to ensure accurate replication of pavement performance. Therefore, the modelling technique should avoid randomness in the predictions, *'black box'* approaches, and stochastic processes in the model, as well as being simple and easily inferred;
6. The overall process of this model, from training and application, should be easily understood and interpretable by road asset managers, and
7. The novelty of this research infers engineering knowledge into the model development and, therefore, the modelling technique should be able to make use of human input (engineering knowledge).

Table 5-1 evaluates each of the techniques identified in the literature review, based on the criteria outlined above.

Table 5-1: Evaluation of Modelling Approaches

Modelling Approaches	Performance	Extensive data and inputs	Non-linearity of the model	Short running time	Learning phase of the model	Interpretability	Human Input	'YES' TOTAL
Bayesian networks	NO	**YES**	**YES**	NO	NO[1]	**YES**	**YES**[2]	**4**
Decision (probability) trees	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**7**
Hidden Markov model	NO	**YES**	**YES**	**YES**	NO[3]	**YES**	NO[4]	**4**
k-nearest neighbours	NO	NO	**YES**	NO	NO	**YES**	**YES**[5]	**3**
Linear discriminant analysis	NO	NO	NO	**YES**	**YES**	**YES**	NO	**3**
Logistic regression	**YES**	NO	NO	**YES**	**YES**	**YES**	NO	**4**
Naïve Bayes	NO	**YES**[6]	**YES**	**YES**[7]	NO[8]	**YES**	NO[9]	**4**
Neural Networks	**YES**	**YES**	**YES**	NO	NO	NO	**YES**[10]	**4**
Random Forests	**YES**	**YES**	**YES**	NO	**YES**	NO	**YES**	**5**
Support Vector Machines	**YES**	**YES**	**YES**	NO	**YES**	NO	**YES**[11]	**5**

[1] Probabilistic or stochastic processes involved
[2] Subjective approach
[3] Requires a large amount of training data
[4] Unfounded assumptions included in the model
[5] User defines the distances between neighbours
[6] Inferred from Bayesian networks
[7] He et al. (2012)
[8] Probabilistic or stochastic processes involved
[9] Strong independence in the model
[10] Supervised learning
[11] Through the use of kernels

From the results above, the following five techniques with the highest score were selected to further evaluate:

- **Decision (probability) trees,** the best performer across all criteria;
- **Random forests,** second equal performer;
- **Support vector machines,** second equal performer;
- **Logistic regression,** out of all the techniques with a score = 4, the performance criteria dominated, and
- **Neural networks,** out of all the techniques with a score = 4, the performance criteria dominated.

To summarise, logistic regression is the simplest discriminative, linear classifier (Eastaugh et al., 1997). However, the nature of road pavement data is not always linearly separable; therefore, to address this, non-linear techniques were considered. Support vector machines met five out of seven criteria above and neural networks was reported a good performer in the literature. Both methods infer human knowledge into the model and are non-linear; however, neural networks are less interpretable than support vector machines, although the way in which both of these techniques work can be difficult to understand (refer Table 2-3). Decision (probability) trees outperformed the other classifiers across all the criteria, yet the simple format of this modelling technique suggests a limited performance, through a simple model curve (decision boundary), associated with the optimal solution. However, random forests are more expressive but very difficult to interpret visually in comparison to trees.

5.3 Further Scrutiny of Classification Techniques Using Road Pavement Data

As previous comparative studies have focussed on the performance of classifiers given data outside of the transportation sector, the five selected techniques are further scrutinised using road pavement data, including:

- A review of the model formulation and definition;
- Definitions of the performance measures used in the assessment of the techniques, and
- Evaluation of the performance of each technique in the transportation sector.

5.3.1 Review of the Selected Classification Techniques

The previous section selected five techniques for further investigation from those identified in the literature review. This section reviews each of these five techniques with respect to the model formulation and definition.

5.3.1.1 Logistic Regression

Logistic regression is a linear classifier and notably reported as one of the simplest classifiers, where a dichotomous[11] dependent variable can be predicted as a probability (Menard, 2010; Tu, 1996). Linear relationships between the independent and dependent variables are easily explained with this method; however, given the general non-linearity of road pavement data (refer Chapter Four), transforming the data by a natural logarithm of the odds, or logit of the dependent variable (Equation 5-1), can illustrate non-linear relationships (Everitt and Hothorn, 2006; Menard, 2010; Peng et al., 2002).

Equation 5-1

$$logit(Y) = \ln(odds) = \ln\frac{\pi}{1-\pi} = \alpha + \beta_1 X_1 + \cdots + \beta_n X_n$$

The outputs of the logit(Y) would be of a natural logarithm and therefore it is possible to convert the predicted outputs into a feasible solution, the odds, by the use of exponentials (Menard, 2010; Peng et al., 2002), shown below:

Equation 5-2

$$odds(Y) = e^{\ln\frac{\pi}{1-\pi}} = e^{\alpha+\beta_1 X_1+\cdots+\beta_n X_n}$$

Converting the output for odds (Equation 5-2) into practical probabilities, Equation 5-3 is used (Everitt and Hothorn, 2006; Menard, 2002):

[11] Where the variable can be divided into two parts or classes. Binary variables are characterised by dichotomy.

Equation 5-3

$$Probability(Y) = P(Y = 1) = \frac{odds(Y)}{1 + odds(Y)} = \frac{e^{\alpha+\beta_1 X_1+\cdots+\beta_n X_n}}{1 + e^{\alpha+\beta_1 X_1+\cdots+\beta_n X_n}}$$

Equations 5-1, 5-2, and 5-3 express the exact same logistic regression output in three different formats, depending on what is desirable by the user (Menard, 2010). The easiest method to analyse a binary dependent variable, given the non-linearity, is using the logit form of the probability. Mathematically, the logit format allows the output to be $\pm\infty$ which, given the theory surrounding probabilities, is unattainable for a probability (Menard, 2010). However, the transformation given in Equation 5-3 ensures the predicted probabilities adhere to probability theory ($0 \leq P(X = 1) \leq 1$).

The optimal model curve (the computational solution to the mathematical problem) is found by an iterative process, which revises the initial provisional solution until the errors of the model are no longer improving or are negligible. The parameters associated with the converged (best) solution characterise the trained logistic regression model (Menard, 2010).

5.3.1.2 Neural Networks

Neural networks are a non-linear classifier that have performed well in solving classification and prediction problems, yet they notably locate local minima as opposed to the global minimum (Eastaugh et al., 1997; Faraway, 2006). Unfortunately, neural networks can be difficult to interpret and are considered a *'black box'*, particularly as the connections from the inputs to the outputs through the hidden layer(s) are unknown to the user (Tu, 1996). Despite the limitations of this method, it is a valuable computational tool and is popular for solving resource intensive complex problems (Saghafi et al., 2009).

A neural network model typically comprises of three parameters:

1. The layout of the network;
2. The learning process of the network for updating the weights, and
3. The activation function for converting the weighted input into the predicted output.

Neural networks, often referred to as feed-forward multi-layer perceptrons, consist of a minimum of three layers; input(s), output, and at least one hidden layer. Each node in one layer is connected to every node in the succeeding layer by a weight, as shown in Figure 5-2. With similarities to the human brain, the inputs are processed through a series of nodes arranged through hidden layers, organising themselves unbeknown to the user in such a way that the desired output is predicted. Although the presence of a hidden layer infers a *'black box'* approach, the inclusion of this layer allows a neural network to model non-linear relationships (Tu, 1996). Multiple hidden layers are possible in neural networks; however, overly complicated and large networks may result in over-fitting of the data, such that there is no advantage of having more than one hidden layer (Tu, 1996), or more than 10 nodes in one hidden layer.

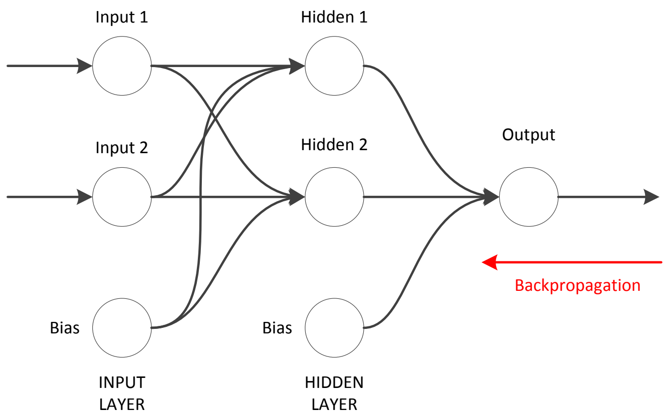


Figure 5-2: Multi-layer perceptron neural network (adapted from Tu, 1996)

The training of this network is done by the backpropagation learning process, popular with neural network applications to converge the best solution. Initially, the network error ε is fed backwards through the network (from output to input) to adjust the weights of each connection line between the nodes (Aitkin and Foxall, 2003; Tu, 1996).The network error ε (Equation 5-4) is calculated at the output node j by:

Equation 5-4

$$\varepsilon = \frac{1}{2}\sum e_j{}^2$$

$$where\text{: } e_j \text{ } is \text{ } the \text{ } difference \text{ } between \text{ } the \text{ } actual \text{ } and \text{ } predicted \text{ } values$$

This error is then sent back through the network and each hidden node (towards the input node *i)* to adjust the connection weights ω (Tu, 1996), using a gradient descent (Equation 5-5), where η defines the learning rate (always positive) to ensure the convergence of the results (Gupta and Lam, 1998; Reed and Marks II, 1999).

Equation 5-5

$$\Delta\omega_{ji} = -\eta\frac{\delta\varepsilon}{\delta\omega_{ji}}$$

$$where\text{: } -\frac{\delta\varepsilon}{\delta v_j} = \phi'(v_j)\sum_k -\frac{\delta\varepsilon}{\delta v_j}\omega_{kj}$$

The desired output of a node is defined by the activation function (ϕ). A linear activation function would remove the hidden layer of the network resulting in a linear system, thus such networks employ non-linear activation functions to solve non-linear classification problems. A number of functions are possible for neural networks; however, for this type of multilayer perceptron network the sigmoid activation function is appropriate (Equation 5-6) and can be represented as a hyperbolic tangent with the range -1 to 1, or as a logistic function with a range of 0 to 1, respectively (Eastaugh et al., 1997; Faraway, 2006; Tu, 1996):

Equation 5-6

$$\phi = \tanh(\nu_i) \quad and \quad \phi = (1 + e^{-\nu_i})^{-1}$$

$$where: \nu_i \text{ is the weighted sum of the input synapses}$$[12]

The normalisation and differentiable rules associated around Equation 5-5 are directly transferable to the activation function; the sigmoid function satisfies these rules and is therefore applied to all the nodes (Aitkin and Foxwall, 2003). After the error (Equation 5-4) converges, the resulting neural network model is defined by the connection weights (shown in Figure 5-2).

5.3.1.3 Support Vector Machines

A large number of studies have employed support vector machines as a non-linear classifier (Rogers and Girolami, 2012) since they were first developed by Cortes and Vapnik (1995) and may be considered to be similar to neural networks. Unlike logistic regression, the outputs are non-probabilistic; however, a distribution can be fitted around the outputs to obtain probabilities by employing scaling methods (Karatzoglou et al., 2006). Support vector machines create a non-linear decision boundary to separate the data with minimal errors.

To do so, the method uses a higher dimensional feature space to map the input vector, defined by the input variables, and it is hoped that in this feature space the transformed data can be linearly separated, such as in Figure 5-3. To transform the input data into the higher dimensional space, kernels are employed. Table 5-2 presents the kernels available and their associated decision boundary shape. A linear kernel will produce a linear separator and a polynomial kernel will result in a curved decision boundary; thus confirming the polynomial, radial basis function, and sigmoid kernel are more flexible than the linear kernel. Each of

[12] Analogous to the human brain, where the synapses are the junctions and connections between the nodes.

these requires additional input parameters that are set by the user, thus inferring human knowledge into the computational model. The radial basis function is a popular kernel to use because the decision boundary associated with this kernel is easily manipulated through the transformed data, opposed to the others. When used, although not popular with support vector machines, the sigmoid kernel replicates a multilayer perceptron neural network (Rogers and Girolami, 2012).

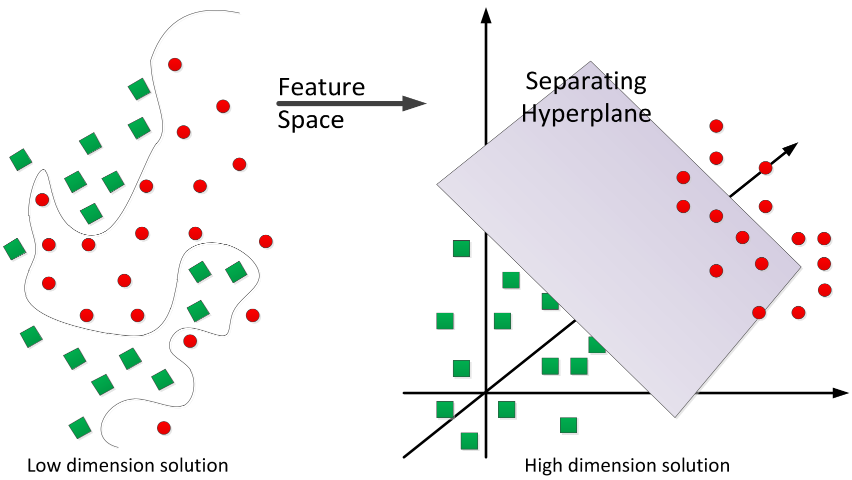


Figure 5-3: Transformed data using the support vector machines classifier (adapted from Van Looy et al., 2007)

Table 5-2: Kernel Methods for Support Vector Machines (adapted from Kecman, 2005; Rogers and Girolami, 2012)

Kernel Method	**Description/Decision Boundary**	**Equation**
Linear	Straight line, through the data	$k(\boldsymbol{x}_n, \boldsymbol{x}_m) = \boldsymbol{x}_n{}^T \boldsymbol{x}_m$
Polynomial (of degree *d*)	Curve, through the data	$k(\boldsymbol{x}_n, \boldsymbol{x}_m) = (1 + \boldsymbol{x}_n{}^T \boldsymbol{x}_m)^d$
Radial Basis Function (or Gaussian)	Circular, surrounding part of the data	$k(\boldsymbol{x}_n, \boldsymbol{x}_m) = e^{-\gamma(x_n - x_m)^T(x_n - x_m)}$
Sigmoid	S-curve, through the data	$k(\boldsymbol{x}_n, \boldsymbol{x}_m) = \tanh(\gamma\, \boldsymbol{x}_n{}^T \boldsymbol{x}_m + c)$

The decision boundary is used to classify the data and, as a linear separator, can be defined as (Kecman, 2005):

Equation 5-7

$$y(x) = \pm(\boldsymbol{w}^T x + b)$$

The $\pm$ sign references the data either side of the decision boundary and $\boldsymbol{w}$ and b are calculated from the data in the learning task of the model. The parameters relate to the optimal model where the margin (γ) defined in Figure 5-4 is greatest, or the loss is minimised.

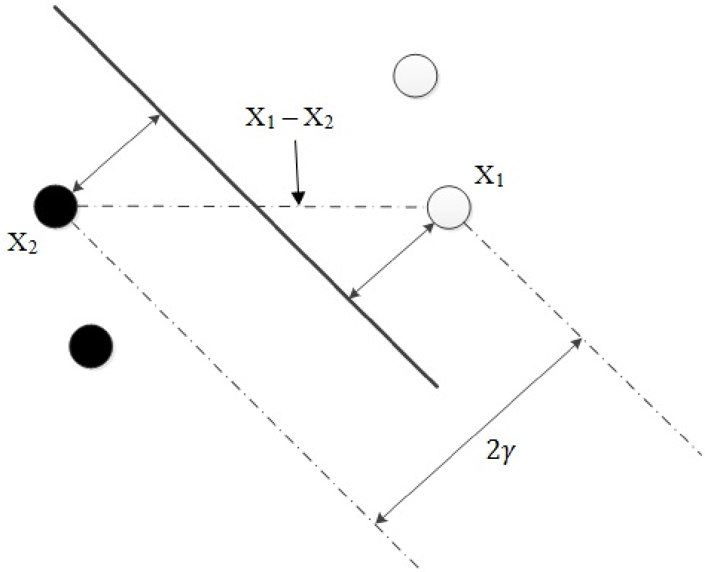


Figure 5-4: Margins surrounding the decision boundary (adapted from Rogers and Girolami, 2012)

The decision boundary is optimally placed to maximise the margin between the closest points either side to the boundary. This margin is calculated as the perpendicular distance from the decision boundary to the closest points on either side (Gil and Johnsson, 2011). A smaller margin would allow for a greater error (loss) of the model with a smaller confidence between class separators. Therefore, with reference to Figure 5-4, the margin can be written mathematically as (Rogers and Girolami, 2012):

Equation 5-8

$$\gamma = \frac{1}{2||\boldsymbol{w}||}\boldsymbol{w}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)$$

$where: \boldsymbol{x}_1\ and\ \boldsymbol{x}_2\ are\ the\ closest\ points (vectors) to\ the\ decision\ boundary$

Since Equation 5-7 can be rewritten as $(\boldsymbol{w}^T x + b) = \pm 1$, Equation 5-8 becomes (Rogers and Girolami, 2012):

Equation 5-9

$$
\begin{aligned}
2\gamma &= \frac{1}{||\boldsymbol{w}||}\boldsymbol{w}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2) \\
&= \frac{1}{||\boldsymbol{w}||}(\boldsymbol{w}^T\boldsymbol{x}_1 - \boldsymbol{w}^T\boldsymbol{x}_2) \\
&= \frac{1}{||\boldsymbol{w}||}((\boldsymbol{w}^T\boldsymbol{x}_1 + b) - (\boldsymbol{w}^T\boldsymbol{x}_2 + b)) \\
&= \frac{1}{||\boldsymbol{w}||}(1 + 1) \\
\gamma &= \frac{1}{||\boldsymbol{w}||}
\end{aligned}
$$

Therefore, to solve the classification problem, Equation 5-9 must be maximised. Since $\boldsymbol{x_1}$and $\boldsymbol{x_2}$ are the closest points to the decision boundary, these are known as the support vectors and define the decision boundary; thus they can be seen as the model definition. To predict a new observation, instead of the model using the entire dataset to relate the new observation in the feature space as some techniques do, support vector machines use only the known support vectors to define the boundary (Rogers and Girolami, 2012). The new observation is plotted with just this information, making this technique more computer efficient and faster than others.

5.3.1.4 Decision (Probability) Trees

Decision trees, referred to in the remainder of the thesis as probability trees, are a hierarchical classification method to solve both classification and regression problems. This technique is very interpretable and easily understood, yet their performance is expected to be inferior to the previous techniques, given the limited expressiveness of the model format and associated decision boundary in comparison to the other methods. However, given the advantage of yielding human-interpretable results, the inferior competitiveness regarding the model performance of this technique can be overlooked (Chen et al., 2004).

A sequence of questions is answered at each *'node'* and when there are no longer any questions to ask, the tree ends with a terminal node or *'leaf'*. Foote (1994) described the three parts involved in the construction of a probability tree as:

1. **Splitting rule,** which determines the decision threshold of the node (in other words, how the nodular question is answered);
2. **Stopping rule,** to determine when the recursion ends (when the tree no longer expands and a node becomes a terminal node), and
3. **Labelling rule,** assigning a class label to each terminal node.

The data is partitioned between each node by answering each nodular question; however, the possible positions of these partitions are endless. An optimal partition may not be attainable and instead the position of the partition is constrained by the use of hyperplanes. The learning phase of the model involves establishing this data separation and, subsequently, the optimal position of the hyperplanes. Like most other methods, the splits are chosen to minimise the loss of the model, or in other words maximise the gain (Foote, 1994), defined by Chen et al. (2004) as the change in information entropy between the original $H(t)$ and predicted $H(x_i, t)$ states given the inferred information from the model, such that:

Equation 5-10

$$Gain(x_i, t) = H(t) - H(x_i, t)$$

The tree continues to grow until no further split is possible or the node contains only one class. To avoid over-fitting, some branches with low gain or contribution to the overall model can be pruned (Everitt and Hothorn, 2006). This may mean that some internal nodes are converted to terminal nodes, ignoring any subsequent splits beyond that node. As a result, an optimum solution to the classification problem can be delivered.

On completion of the learning phase, nodes are assigned a respective class label and the statistics associated with each terminal node, such as probabilities, are computed (Everitt and Hothorn, 2006). For a new prediction, the constructed tree assigns the new observation with the statistics associated with the predicted terminal node, once the observation has passed through the tree.

5.3.1.5 Random Forests

Random forests are an ensemble[13] technique, analogous to bagging[14] except this technique compounds only the same type of classification model, where a collection of decision (probability) trees are constructed to create a forest. Since they are constructed from probability trees, the concepts of trees and recursive partitioning are applicable to random forests to the extent that the learning mechanism of random forests is that of probability trees. Although random forests are computationally effective given their substantial expressiveness (Robnik-Šikonja, 2004), the complication of constructing a forest results in the model becoming less interpretable than probability trees and difficult for the user to understand.

A number of specific rules are followed to construct a forest in addition to the learning phase of decision (probability) trees. In the learning process, some randomness is introduced into

[13] Several computational models are used together, often one after another, to improve on the overall performance of the individual techniques.
[14] Equal weights are given to each model, in this case each tree developed, to create an overall compounded modelling system of individual trees.

the splitting rule for each individual tree by growing each tree on a random subsample of the data (Robnik-Šikonja, 2004). The collation of the trees aims for minimal repetition of the data classes and information at each terminal node. To achieve this, weighted voting is used to combine the outputs of the individual trees, where new predictions are determined by the highest number of class votes from all trees in the forest (Sirikulviriya and Sinthupinyo, 2011).

5.3.2 Performance Measures

The definition of the performance of a model is multi-dimensional, such that the assessment methods of some performance measures are nonchalant and obscured. Logically, accuracy or misclassification error measures would be employed to evaluate the performance of modelling techniques. However, the accuracy of a model is not acceptable as a performance measure alone as the bias results do not allow for random successes (Ben-David, 2008; Dreiseitl and Ohno-Machado, 2002; Parker, 2011). To ensure this comparative study sufficiently assessed the performance of the classification techniques, other performance measures were explored, categorised as follows (Parker, 2011):

- **Point Measures:** These measures estimate the performance of the model at a specific threshold and ignore all other threshold points. These types of estimates assume that false positives and false negatives are equally undesirable (Parker, 2011). Examples of such measures are the F-score and phi coefficient, and
- **Integrated Measures:** These measures evaluate model performance as the threshold changes, such that the consequences with all incorrect predictions are considered in the evaluation. Performance curves, such as the receiver operating characteristic curve and Cohen's kappa curve, are calculated by plotting such constituents (e.g. sensitivity,

false positive rate, and Cohen's kappa) against each other across various thresholds. Integrating the curve with the associated cost function evaluates the performance of the technique using such measures as area under the receiver operating characteristic curve, the area under the Cohen's kappa curve, and the H-measure (Fawcett, 2006; Parker, 2011; Rogers and Girolami, 2012).

With the changes in output demand of classification techniques, integrated measures are now favoured for assessing the performance of modelling techniques as opposed to point measures (Parker, 2011). Ideally integrated measures would be used for this research; however, because of limitations in the data, the costs (consequences) associated with false positives and false negatives were assumed to be equal, such that it is equally undesirable (e.g. an equal cost function) to have an incorrect prediction of failure, whether that be a false positive or false negative prediction. However, Parker (2011) reported that the phi coefficient was a satisfactory alternative to integrated measures in addition to the comparative performance of the F-score. Therefore, the F-score and phi coefficient, in conjunction with accuracy and misclassification error, were used to assess the classification techniques using a failure threshold of 0.5 ($P(X \geq 0.5) = 1$). These techniques are described further below.

5.3.2.1 Performance (Confusion) Matrix

Figure 5-5 defines the elements of the confusion matrix, each used in the subsequent performance measures.

		DATA FAILURES 0	**DATA FAILURES** 1	
PREDICTIONS	**0**	T_P True Positives	F_P False Positives	N_1 Number of predicted non-failures
PREDICTIONS	**1**	F_N False Negatives	T_N True Negatives	N_2 Number of predicted failures
		N_3 Number of non-failed sites	N_4 Number of failed sites	N_{Total} Total number of sites (or predictions)

Figure 5-5: Confusion matrix (Fawcett, 2006; Sokolova and Lapalme, 2009)

For this research, sound pavements were assigned a zero (0) class, therefore correctly predicted sound pavements were represented by true positives and sites predicted to fail but were in fact sound were defined as false negatives. Failed pavements were assigned a one (1) class and, therefore, true negatives were correctly predicted failures and sites predicted to be sound but in fact had failed were false positives.

Direct relationships from the confusion matrix include:

Precision (positive predictive value): The proportion of predicted non-failures correctly predicted (*what proportion of the predicted sound pavements were in fact sound?*). A higher precision value indicates a higher level of confidence surrounding the predicted sound pavements (Fawcett, 2006; Powers, 2011; Sokolova and Lapalme, 2009).

Equation 5-11

$$Precision = \frac{T_P}{T_P + F_P}$$

$$where: T_P \ and \ F_P \ are \ defined \ in \ the \ confusion \ matrix \ (refer \ Figure \ 5-5)$$

Recall (sensitivity, true positive rate): The proportion of sound road pavement sections correctly predicted (*how many sound pavements were predicted as sound?*).

The result indicates the ability of the technique to correctly identify sound pavements (Fawcett, 2006; Powers, 2011; Rogers and Girolami, 2012; Sokolova and Lapalme, 2009).

Equation 5-12

$$Recall = \frac{T_P}{T_P + F_N}$$

$$where: T_P \; and \; F_N \; are \; defined \; in \; the \; confusion \; matrix \; (refer \; Figure \; 5-5)$$

Negative predictive value: Opposite to the positive predictive value, the proportion of the predicted failed sites that are correctly predicted (*what proportion of the predicted failures had in fact failed?*). A high percentage results in a strong confidence regarding the predicted failures (Powers, 2011).

Equation 5-13

$$Negative \; predictive \; value = \frac{T_N}{F_N + T_N}$$

$$where: T_N \; and \; F_N \; are \; defined \; in \; the \; confusion \; matrix \; (refer \; Figure \; 5-5)$$

Specificity (true negative rate): The proportion of correctly predicted failures given the total number of failed sections in the data (*how many failed pavements were predicted as failed?*), where a high specificity indicates a strong ability of the technique to identify failed pavements (Fawcett, 2006; Powers, 2011; Rogers and Girolami, 2012; Sokolova and Lapalme, 2009).

Equation 5-14

$$Specificity = \frac{T_N}{F_P + T_N}$$

$$where: T_N \ and \ F_P \ are \ defined \ in \ the \ confusion \ matrix \ (refer \ Figure \ 5-5)$$

The performance assessment of the techniques should not be solely based on the direct relationships from the confusion matrix as the above tests rarely conclude the practicality and effectiveness of classification methods.

5.3.2.2 Accuracy and Misclassifications

Both the accuracy and misclassification error measures calculate the percentages of incorrectly predicted pavement sections, as shown below in Equations 5-15 and 5-16 (Powers, 2011; Rogers and Girolami, 2012; Sokolova and Lapalme, 2009). Ideally, a higher accuracy, associated with a low misclassification error, is desirable to determine the percentages of correctly and incorrectly predicted sound and failed pavement sites. However, as stated above, the results from these calculations should not be used alone in assessing the performance of classification techniques.

Equation 5-15

$$Accuracy = \frac{\sum(T_P + T_N)}{N_{Total}} \times 100 \ \%$$

Equation 5-16

$$Misclassification \ Error = \frac{\sum|P_{Predicted} - P_{Actual}|}{N_{Total}} \times 100 \ \% = \frac{\sum(F_P + F_N)}{N_{Total}} \times 100 \ \%$$

$$where: T_P, T_N = Number \ of \ true \ positive \ and \ true \ negative \ predictions \ (refer \ Figure \ 5-5)$$

$$F_P, F_N = Number \ of \ false \ positive \ and \ false \ negative \ predictions \ (refer \ Figure \ 5-5)$$

$$P_{Predicted}, P_{Actual} = Probabilities \ (predicted, actual)$$

5.3.2.3 F-Score

The F-score measures the performance of the binary classifier on a scale of zero (0) to one (1), where the closer the value is to 1, the more accurate the method is regarded (Parker, 2011; Sokolova and Lapalme, 2009). It is a weighted average of the precision and recall values (the predictive power of the model to sound pavements); however, it neglects the number of correctly predicted failures (e.g. the true negative rate). It is a preferred measure of performance over accuracy calculations given its inclusion of incorrect predictions, and is calculated by (Sokolova and Lapalme, 2009):

Equation 5-17

$$F\_Score = \frac{2 \times p \times r}{p + r}$$

$$where: p = precision \ and \ r = recall$$

5.3.2.4 Phi Coefficient

The phi coefficient, also known as Matthews Correlation Coefficient, measures the agreement between the inputs and the output and, therefore, how well the selected technique predicted failed and sound pavements. A negative agreement, equal to -1, suggests the majority of the results are incorrectly predicted, and a positive agreement of +1 would demonstrate that the method is accurate in predicting road pavement failure (Powers, 2011). A value of zero (0) indicates no relationship between the predictions and input variables.

The phi coefficient is often favoured above the F-score (Parker, 2011) because it takes into account negative values (or failed pavement predictions), unlike the F-score which fails to use the predictions of true negatives (failed pavements). It is calculated using the following (Parker, 2011; Powers, 2011):

Equation 5-18

$$\emptyset = \frac{T_P \times T_N - F_P \times F_N}{\sqrt{N_1 \times N_2 \times N_3 \times N_4}}$$

where: $\mathrm{T_P}, \mathrm{T_N} = \text{Number of true positive and true negative predictions (refer Figure } 5-5)$

$\mathrm{F_P}, \mathrm{F_N} = \text{Number of false positive and false negative predictions (refer Figure } 5-5)$

$N_1 to\ N_4\ are\ the\ sums\ of\ the\ rows\ and\ columns\ of\ the\ confusion\ matrix\ (refer\ Figure\ 5-5)$

5.3.3 Methodology for Evaluation of Techniques

Using the *R* functions described in Appendix C, each classification model type was constructed and tested using a reserved portion of the State Highway LTPP dataset. Given the size of the research dataset, only a limited number of datapoints were able to be reserved from this dataset for testing purposes; therefore, to ensure variability of the predictions is accounted for in the results, a 10-fold cross-validation sampling method was employed (Rogers and Girolami, 2012). This approach divided the dataset into 10 subsamples of equal size, randomly selected, reserving one subsample for testing at all times. Therefore, over the 10 repetitions, 90 % of the dataset was used in the training phase of the model and the remaining 10 % was reserved for testing. An average of these 10 repetitions, including each calculated performance measure, was used to evaluate the performance of the classifiers. With the randomness of this method, the testing sample does not guarantee to include an equal number of failed and non-failed pavement observations.

Each of the factor combinations, comprising of the factor groups discussed in Chapter Four (refer Section 4.6), shown in Table 5-3, were modelled individually to:

- Determine the correlation between the factor groups and failure mechanism to validate the statistical computational model, and
- Establish the overall performance of the classification technique.

To do so, the accuracy of each model constructed was reported. An investigation of the distributions of the four performance measures used box and whisker plots (Ayyub and McCuen, 2003) to establish the overall performance of the technique. Density distribution plots explored the performance of each factor combination per failure mechanism, against each technique, to further add to the accuracy results. Finally, the Wilcoxon Signed Rank Test (R Core Team, 2011) analysed the statistical significance between the means of the performance measures. For this comparative study, the performance evaluation remained impartial between the failure mechanisms.

Table 5-3: Factor Combinations of the Factor Groups

Factor Combination	FACTOR GROUPS					
	Traffic	Composition	Strength	Environment	Surface Condition	Subgrade Sensitivity
1	x					
2		x				
3			x			
4				x		
5					x	
6						x
7	x	x				
8	x		x			
9	x			x		
10	x				x	
11	x					x
12		x	x			
13		x		x		
14		x			x	
15		x				x
16			x	x		
17			x		x	
18			x			x
19				x	x	
20				x		x
21					x	x
22	x	x	x			
23	x	x		x		
24	x	x			x	
25	x	x				x
26	x		x	x		
27	x		x		x	
28	x		x			x
29	x			x	x	
30	x			x		x
31	x				x	x
32		x	x	x		
33		x	x		x	
34		x	x			x
35		x		x	x	
36		x		x		x
37		x			x	x
38			x	x	x	
39			x	x		x
40			x		x	x
41				x	x	x
42	x	x	x	x		
43	x	x	x		x	
44	x	x	x			x
45	x	x		x	x	
46	x	x		x		x
47	x	x			x	x
48	x		x	x	x	
49	x		x	x		x
50	x		x		x	x
51	x			x	x	x
52		x	x	x	x	
53		x	x	x		x
54		x	x		x	x
55		x		x	x	x
56			x	x	x	x
57	x	x	x	x	x	
58	x	x	x	x		x
59	x	x	x		x	x
60	x	x		x	x	x
61	x		x	x	x	x
62		x	x	x	x	x
63	x	x	x	x	x	x

5.3.4 Performance of the Classifiers

Based on the predicted probability of failure, the assessment of each classification technique used the performance measures described in Section 5.3.2 to assist in determining the suitability of each technique to the aim of this research.

5.3.4.1 Accuracy and Misclassifications

Tables 5-4, 5-5, and 5-6 present the accuracy and misclassification error percentages based on Equations 5-15 and 5-16, for each failure mechanism, and report the successful factor combinations that failure can be correctly predicted from, given a successful accuracy threshold of 100 % (and 90 % in the case of shear failure). In addition, the factor combinations that performed the poorest for each technique are identified. The accuracy and misclassification error percentages for all factor combinations trialled can be found in Appendix D.

It is difficult from these results to determine the most accurate modelling technique; instead, the technique(s) with the smallest misclassification error across the three failure mechanisms include:

- **Rutting:** Support vector machines and random forests (Table 5-4);
- **Fatigue Cracking:** Support vector machines and random forests (Table 5-5), and
- **Shear:** Support vector machines (Table 5-6).

Table 5-6 shows the performance of all modelling techniques are considerably less accurate in predicting shear failures than that of the other two failure mechanisms. In addition to the considerably higher misclassification error percentages, a greater number of factor combinations were considered *'worst trials'* for the shear failure mechanism, compared to the rutting and fatigue cracking analysis. The inherent behaviour of shear failures and the lack of material information in the State Highway LTPP dataset could be reason for this observation.

The successful factor combinations identified by the support vector machines and logistic regression techniques correlated well with the respective failure charts in Chapter Four, such that combinations containing strength, traffic and composition factors, for example, were associated with predicting rutting failure. However, with the large number of successful factor combinations reported for neural networks and random forests, the correlation may be considered coincidental in some cases. Despite this, the surface condition factor group was only evident in successful fatigue cracking combinations, which suggests forecasting failure and basing maintenance decisions on the surface condition alone is erroneous.

Table 5-4: Summary of the Accuracy and Misclassification Error Results for Rutting Failure (Average of 10-Fold Cross-Validation)

Modelling Technique	Factor Combinations Numbers	Accuracy (%)	Misclassification Error (%)
Logistic Regression			
Best Trial(s)	12, 18, 22, 32, 33, 34, 39, 40, 42, 43, 44, 52, 53, 54, 56, 57, 58, 59, 62, 63	100	0
Worst Trial(s) (< 50 %)	4	38	62
Neural Networks			
Best Trial(s)	1, 2, 3, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 36, 37, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63	100	0
Worst Trial(s) (< 65 %)	5, 6, 20	64, 61, 60	36, 39, 40
Support Vector Machines			
Best Trial(s)	3, 7, 8, 12, 16, 18, 22, 25, 26, 28, 32, 34, 39, 42, 44, 46, 49, 53, 58	100	0
Worst Trial(s) (< 85 %)	4, 6, 20	85	15
Probability Trees			
Best Trial(s)	8, 12, 22, 26, 27, 32, 33, 34, 42, 43, 44, 48, 52, 53, 54, 57, 58, 59, 62, 63	100	0
Worst Trial(s) (<60 %)	6, 20	28, 59	72, 41
Random Forests			
Best Trial(s)	1, 2, 3, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 57, 58, 59, 62, 63	100	0
Worst Trial(s) (< 85 %)	5, 6, 19, 20	84, 85, 84, 85	15, 15, 13, 15

Table 5-5: Summary of the Accuracy and Misclassification Error Results for Fatigue Cracking Failure (Average of 10-Fold Cross-Validation)

Modelling Technique	Factor Combinations Numbers	Accuracy (%)	Misclassification Error (%)
Logistic Regression			
Best Trial(s)	7, 8, 22, 23, 24ab, 25, 26, 27ab, 28, 32, 33ab, 42, 43ab, 44, 45ab, 46, 47ab, 48ab, 49, 50a, 52ab, 57ab, 58, 59ab, 60ab, 61ab, 62ab, 63ab	100	0
Worst Trial(s) (< 50 %)	6	47	54
Neural Networks			
Best Trial(s)	1, 2, 3, 7, 8, 9, 10ab, 11, 12, 13, 14ab, 15, 16, 17ab, 18, 22, 23, 24ab, 25, 26, 27ab, 28, 30, 31ab, 32, 33ab, 34, 35ab, 36, 37ab, 38ab, 39, 40ab, 42, 43ab, 44, 45ab, 46, 47ab, 48ab, 49, 50ab, 51ab, 52ab, 53, 54ab, 55ab, 56ab, 57ab, 58, 59ab, 60ab, 61ab, 62ab, 63ab	100	0
Worst Trial(s) (< 65 %)	6	48	52
Support Vector Machines			
Best Trial(s)	3, 7, 8, 12, 18, 22, 23, 24b, 25, 26, 28, 32, 33ab, 34, 42, 43a, 44, 46, 47b, 49, 50ab,52ab, 53, 54ab, 57ab, 58, 59ab, 62ab, 63ab	100	0
Worst Trial(s) (< 85 %)	5ab, 6, 21ab	84, 84, 84, 85, 84	16, 16, 16, 15, 16
Probability Trees			
Best Trial(s)	1, 3, 7, 8, 9, 10ab, 11, 16, 17ab, 18, 22, 23, 24ab, 25, 26, 27ab, 28, 29ab, 30, 31ab, 38ab, 39, 40ab, 42, 43ab, 44, 45ab, 46, 47ab, 48ab, 49, 50ab, 51ab, 56ab, 57ab, 58, 59ab, 60ab, 61ab, 63ab	100	0
Worst Trial(s) (< 60 %)	6	43	57
Random Forests			
Best Trial(s)	1, 2, 3, 7, 8, 9, 10ab, 11, 12, 13, 15, 16, 17ab, 18, 22, 23, 24ab, 25, 26, 27ab, 28, 29b, 30, 31ab, 32, 33ab, 34, 36, 38ab, 39, 40ab, 42, 43ab, 44, 45ab, 46, 47ab, 48ab, 49, 50ab, 51ab, 52ab, 53, 54ab, 56ab, 57ab, 58, 59ab, 60ab, 61ab, 62ab, 63ab	100	0
Worst Trial(s) (< 85 %)	5ab, 6, 21ab	84, 84, 84, 85, 85	15, 15, 16, 15, 15

Table 5-6: Summary of the Accuracy and Misclassification Error Results for Shear Failure (Average of 10-Fold Cross-Validation)

Modelling Technique	Factor Combinations Numbers	Accuracy (%)	Misclassification Error (%)
Logistic Regression			
Best Trial(s)	44, 58, 59, 63	100	0
Worst Trial(s) (< 50 %)	4, 6	39, 49	61, 51
Neural Networks			
Best Trial(s)	2, 3, 7, 8, 12, 15, 16, 18, 22, 23, 25, 26, 28, 32, 33, 34, 39, 42, 43, 44, 46, 49, 50, 52, 53, 57, 58, 59, 60, 61, 62, 63	100	0
Worst Trial(s) (< 65 %)	4, 5, 6, 19, 20, 21, 41	40, 54, 49, 58, 52, 61, 65	60, 46, 51, 42, 48, 39, 35
Support Vector Machines			
Best Trial(s)	7, 12, 18, 22, 25, 28, 32, 34, 42, 44, 49, 53, 58	90	10
Worst Trial(s) (< 85 %)	1, 2, 4, 5, 6, 9, 10, 11, 13, 14, 15, 19, 20, 21, 29, 30, 31, 35, 36, 37, 38, 41, 51, 55	84, 82, 61, 61, 59, 79, 75, 84, 80, 79, 85, 65, 62, 64, 75, 84, 82, 79, 84, 82, 85, 65, 82, 82	16, 18, 39, 39, 41, 21, 25, 16, 20, 21, 15, 35, 38, 36, 25, 16, 18, 21, 16, 18, 15, 35, 18, 18
Probability Trees			
Best Trial(s)	8, 26, 27, 28, 48, 49, 50, 61	100	0
Worst Trial(s) (< 60 %)	4, 5, 6, 19, 20, 21, 41	39, 50, 49, 50, 51, 52, 58	61, 50, 51, 50, 49, 48, 42
Random Forests			
Best Trial(s)	2, 3, 7, 8, 12, 13, 14, 15, 16, 18, 22, 23, 24, 25, 26, 27, 28, 32, 33, 34, 35, 36, 37, 39, 40, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63	100	0
Worst Trial(s) (< 85 %)	4, 5, 6, 19, 20, 21, 41	67, 59, 61, 64, 65, 65, 69	33, 32, 39, 30, 35, 30, 29

Strong conclusions on the performance of each technique cannot be inferred from the results; however, from the above tables, the successful factor combinations can be further used in the development of the prototype system (see Chapter Six).

5.3.4.2 Distribution Plots

Figure 5-6 compares each technique based on the distributions of the four performance measures. The low phi coefficient indicated a weak relationship between the input variables and predicted outputs for shear failure. As this was evident across all five techniques, it was suggested that the available dataset does not contain enough information, such as pavement material properties, to predict shear failure with certainty.

For the other two failure types, Figure 5-6 concluded that neural networks and random forests perform better than the other three classifiers across all four measures. Probability trees performed marginally better than support vector machines given the accuracy, misclassification error and phi coefficient assessments, but ignoring the shear failure mechanisms the two techniques exhibited equal performance in the F-score. The logistic regression performance was least successful across all four performance measures, with the range present for each performance measure and failure mechanism exceeding that of the other four techniques. The raw data used to construct Figure 5-6 can be found in Appendix D.

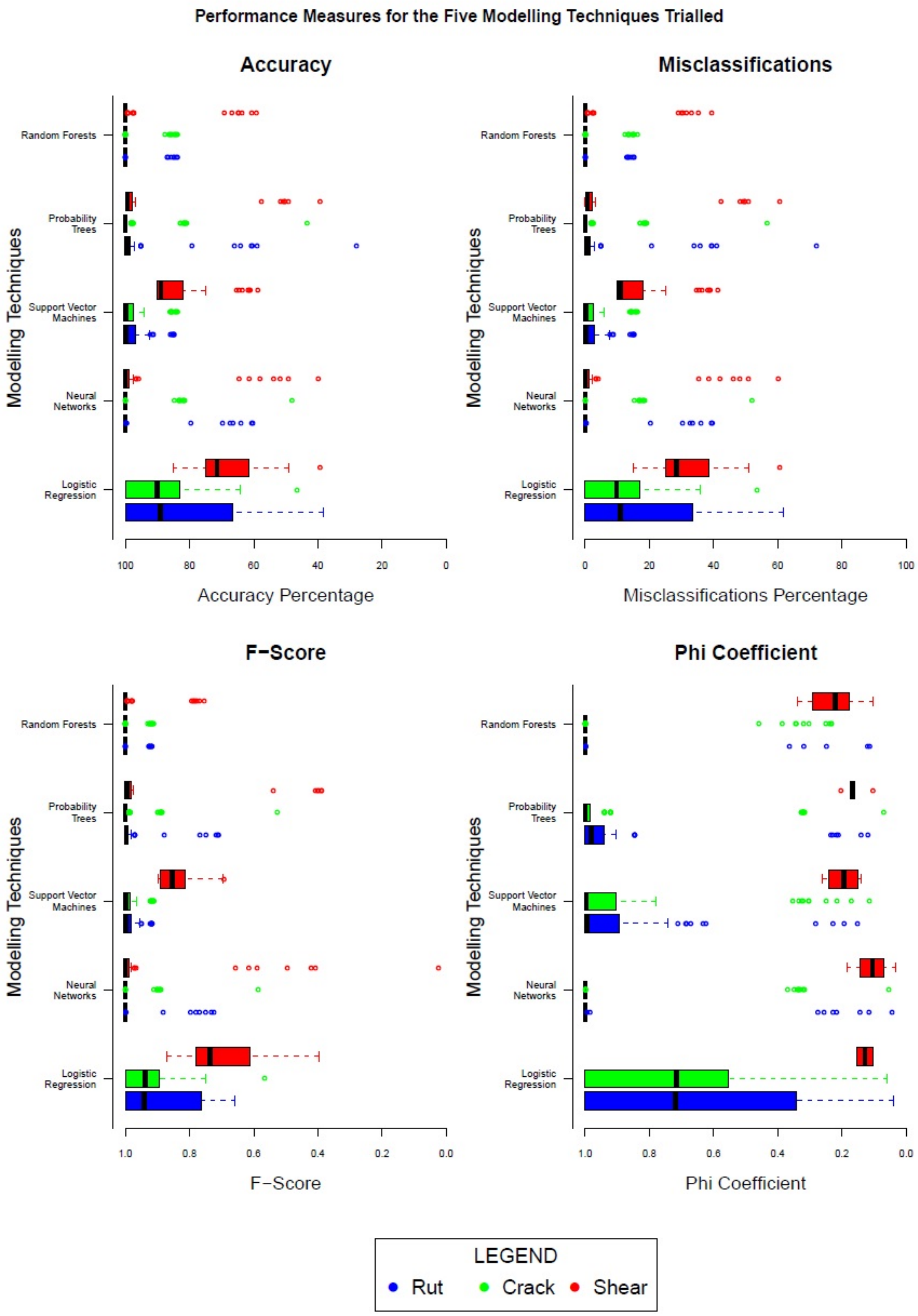


Figure 5-6: Box and whisker plots of the performance measures

From this, the classification techniques are ranked in the following order, given the performance evaluation of each in the above analysis and Figure 5-6:

1. Neural Networks and Random Forests
2. Probability Trees
3. Support Vector Machines
4. Logistic Regression

5.3.4.3 Density Plots

Figures 5-7 and 5-8 present the density distribution plot per factor combination for rutting and fatigue cracking failure mechanisms to compare the performance of each classification technique, given the input variables. Largely, the results reporting on the comparative performance of each technique were inconclusive. The full collection of density plots are presented in Appendix D and are summarised below.

Rutting:

In general, the density distribution plots for the rutting failure mechanism show little difference in the distributions between each of the classification techniques over the four performance measures. However, the plots focussed on the accuracy and misclassification error percentages showed that overall the performance distribution of the neural network technique differs from the other four classifiers, across the majority of the factor combinations. The plots detailing the F-score and phi coefficient have few factor combinations where the distributions of each modelling technique differ. Generally, the accuracy, and the associated misclassification error, density distribution plots for neural networks differ to the other four modelling techniques (see Figure 5-7), suggesting the performance of neural networks is significantly different to the others for predicting rutting failure.

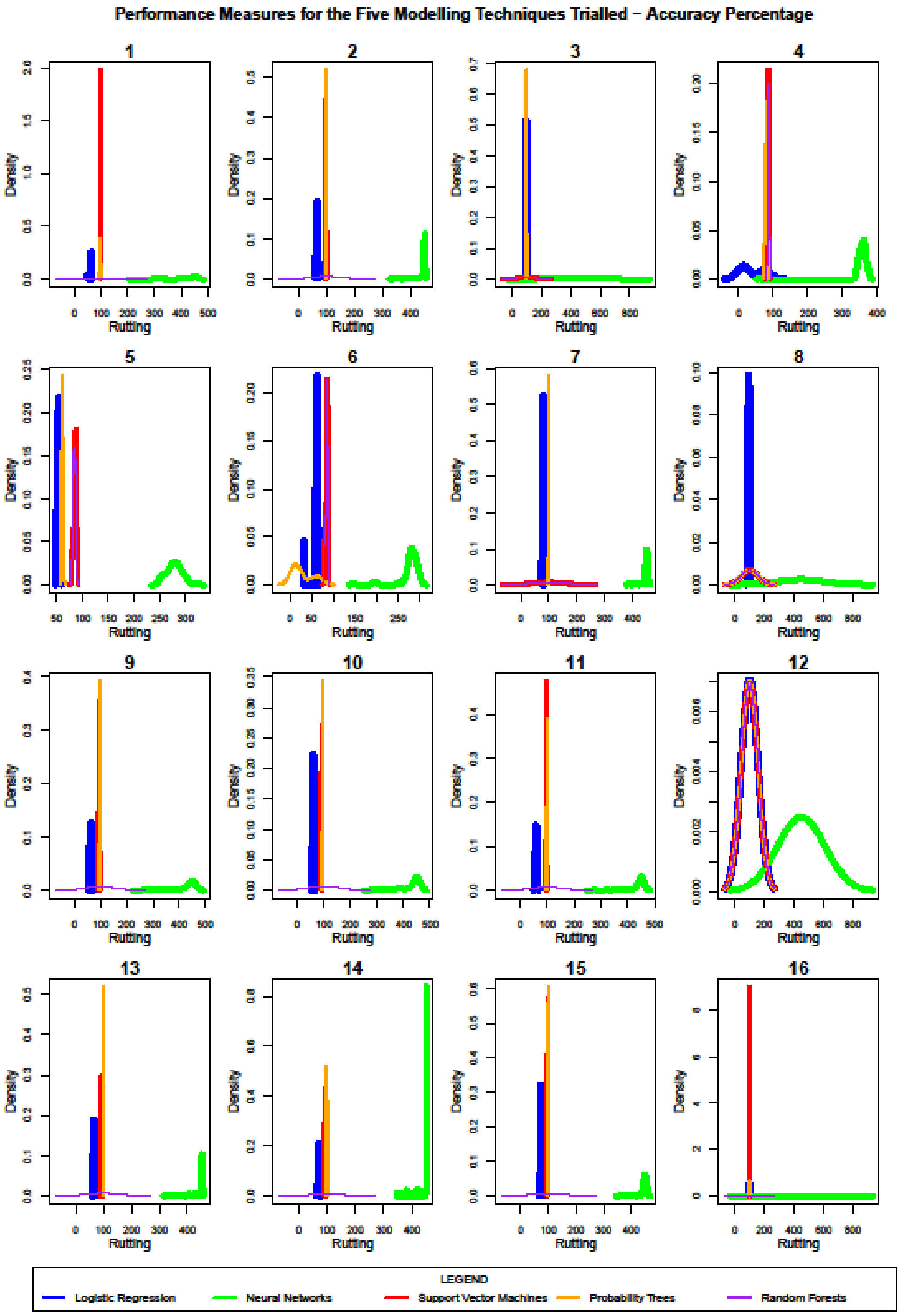


Figure 5-7: Density plots for the accuracy performance measure for the rutting failure mechanism

Fatigue Cracking:

The density distribution plots exploring the misclassification error for fatigue cracking shows a distinct difference between logistic regression and the other four techniques. The centre of the distribution curve for logistic regression tends to be greater than that of the other four techniques (see Figure 5-8), suggesting a higher misclassification error associated with logistic regression. Only approximately 10 factor combinations (out of 95 including all *a* and *b* trials) showed some variance between the modelling techniques for each performance measure.

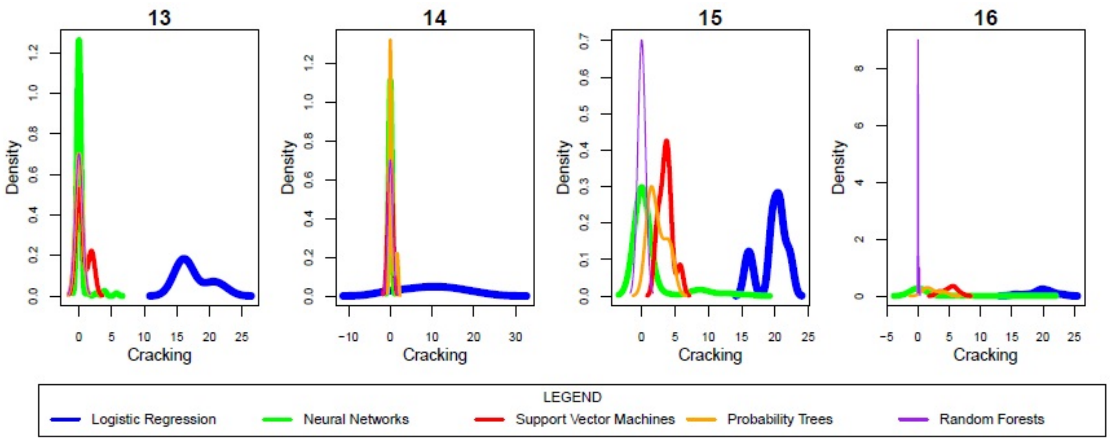


Figure 5-8: Example of density plots for the misclassification error performance measure for fatigue cracking failure

Shear:

The results of the density plots further supported the conclusions from Section 5.3.4.2, in regards to shear failure, where the weak relationships presented in the box and whisker plots were accounted for. Although many of the plots showed a difference in the distributions, Section 5.3.4.2 concluded poor correlations between the input variables and predicted output; therefore, the variation in the distribution plots may be a result of this poor relationship and considered inconclusive for this study.

5.3.4.4 Wilcoxon Signed Rank Test

To determine the statistical difference, if any, in the reported performance of the five modelling techniques, hypothesis testing was employed to establish, statistically, the difference between the distributions of each performance measure. Specifically, the Wilcoxon Signed Rank test (R Core Team, 2011) is a non-parametric statistical hypothesis test used on both small and large samples, which assumes:

1. The data is paired,
2. Each pair is chosen randomly and independent, and
3. The data is measured on an interval scale but does not need to be normal.

Therefore, a paired two-sided test compared the population means of each of the four performance measures with a desired confidence interval for this hypothesis test of 95 %. Given the nature of hypothesis testing, the following is defined:

Null hypothesis = H_0 = difference between the techniques is equal to 0

Alternative hypothesis = H_1 = difference between the techniques is ***NOT*** *equal to 0*

The null hypothesis is accepted if the p-value from the Wilcoxon Signed Rank test is greater than 0.05 (i.e. = 1 – 0.95), otherwise it is rejected and the alternative hypothesis is accepted (p-value < 0.05).

Tables 5-7 and 5-8 show there is no statistical difference between the performance of support vector machines and probability trees as H_0 is accepted across all four of the performance measures for the rutting and fatigue cracking failure mechanisms. There is also no statistical difference in the phi coefficient results between the random forests and neural networks, across the three failure mechanisms.

For the shear failure mechanism, there is a difference between each classification technique for accuracy, misclassification error and F-score measures; however, there is no statistical difference between the performances of the techniques based on the phi coefficient.

Table 5-7: Results from the Wilcoxon Signed Rank Test for the Accuracy and Misclassification Error Performance Measures, showing the Acceptance of the Null Hypothesis (H_0)

Modelling Techniques

	Logistic Regression			Neural Networks			Support Vector Machines			Probability Trees			Random Forests		
	R	*C*	*S*	*R*	*C*	*S*	*R*	*C*	*S*	*R*	*C*	*S*	*R*	*C*	*S*
Logistic Regression															
Neural Networks															
Support Vector Machines										H_0	H_0				
Probability Trees							H_0	H_0							
Random Forests															

Accuracy

Misclassification Error

R – Rutting; *C* – Fatigue Cracking; *S* – Shear

Table 5-8: Results from the Wilcoxon Signed Rank Test for the F-Score and Phi Coefficient Performance Measures, showing the Acceptance of the Null Hypothesis (H_0)

Modelling Techniques

	Logistic Regression			Neural Networks			Support Vector Machines			Probability Trees			Random Forests		
	R	*C*	*S*	*R*	*C*	*S*	*R*	*C*	*S*	*R*	*C*	*S*	*R*	*C*	*S*
Logistic Regression															
Neural Networks			H_0												
Support Vector Machines			H_0			H_0				H_0	H_0				
Probability Trees			H_0			H_0	H_0	H_0	H_0						
Random Forests			H_0	H_0	H_0	H_0			H_0			H_0			

F-Score

Phi Coefficient

R – Rutting; *C* – Fatigue Cracking; *S* – Shear

The Wilcoxon Signed Rank test results conclude that, for both the rutting and fatigue cracking failure mechanism, there is no difference between support vector machines and probability trees. In addition, there is no difference between the distributions of the phi coefficients for neural networks and random forests. Based on this, the techniques can be ranked as followed, given the previous rankings in Section 5.3.4.2:

1. Neural Networks and Random Forests
2. Probability Trees and Support Vector Machines
3. Logistic Regression

5.4 Comparison of Classification Techniques

5.4.1 Selecting the Most Appropriate Technique

The previous section assessed the performance of each of the five classification techniques. This section further builds on those results and bases the choice of the most appropriate technique for further development on three criteria, as follows:

1. The performance of the technique based on the assessment of the four performance measures (accuracy, misclassification error, F-score, and phi coefficient) (Section 5.3.2);
2. The technique must be quick to run, simple to use, and the overall process should be easily understood by the asset manager, enabling the processes involved in predicting failure to be known, and
3. The model generalisation should be adequate so to avoid over-fitting of the model, ensure the performance of the model is not compromised when transferring the prototype system to other road datasets, and ensure the model does not require large amounts of data for new predictions (e.g. avoid data hungry models).

Based on the criteria above, the most appropriate classification technique for this research is not necessarily the one that outperforms the others; instead, the needs of the road management industry have also been accounted for in the criteria for the successful implementation of the system.

5.4.2 Discussion

Table 5-9 lists the five techniques accordingly to the criteria above (refer Section 5.4.1), based on the results from the literature search (refer Table 2-3) and the performance results from

Section 5.3.4. Logistic regression satisfied four of the five categories, which was expected given its linearity and simplicity, suggesting its appropriateness to the problem domain. However, the linearity of this technique resulted in a comparatively poor performance against the other techniques. To avoid selecting a technique that satisfied the interpretability criteria over performance, weightings emphasised the importance of each criterion to the overall system aim. Criterion #1 was assigned the largest weighting of 5, to ensure a poor performing technique was not chosen. The remaining two criterions were assigned smaller weightings to reflect their importance of the usability and interpretability of the system. Criterion #2 was sub-divided so the individual elements of the criteria (run time, usability and interpretability) were independently assessed.

Table 5-9: Ranking of the Modelling Techniques

RANKING	**Performance**	**Running Speed**	**Ease of Use**	**Interpretability**	**Avoid Over-fitting**
	Criterion #1	Criterion #2	Criterion #2	Criterion #2	Criterion #3
1	Neural Networks and Random Forests	Logistic Regression	Logistic Regression	Logistic Regression	Logistic Regression
2	Probability Trees and Support Vector Machines	Probability Trees and Support Vector Machines	Support Vector Machines	Probability Trees	Support Vector Machines
3	Logistic Regression	Random Forests	Probability Trees	Random Forests and Support Vector Machines	Probability Trees
4		Neural Networks	Random Forests and Neural Networks	Neural Networks	Random Forests
5					Neural Networks

From Table 5-10 below, it may be concluded that the most appropriate modelling techniques were both support vector machines and probability trees. This finding is analogous to the accuracy results presented in Section 5.3.4.1. The techniques disregarded include logistic regression due to its poor performance attributed to the linearity of the model function, neural networks because of the associated long running times and lack of user interpretability despite its performance, and random forests due to their somewhat complicated model structure and long running time.

Table 5-10: Selection of the Most Appropriate Modelling Technique

	Criteria						
	#1	**#2**			**#3**		
	Performance	Speed	Ease of Use	Interpretability	Over-fit Properties		**Overall**
Weightings	**5**	**3**	**1**	**1**	**2**	**Total**	**Rank**
Logistic Regression	5	1	1	1	1	25+3+1 +1+2= **32**	3
Neural Networks	1	5	4	5	5	5+15+5 +4+10= **39**	5
Support Vector Machines	3	2	3	3	2	15+6+3+3+ 4= **31**	1
Probability Trees	3	2	2	2	3	15+6+2+2+ 6= **31**	1
Random Forests	1	4	4	3	4	5+12+3+4+ 8= **32**	3

Although probability trees could be easily constructed to follow the developed failure charts (refer Chapter Four), this research selected support vector machines to develop further.

5.5 Summary of the Comparative Study

This chapter presented a methodology to establish the most appropriate classification technique to further develop in the prototype system. Within this methodology, criteria were developed to select five classification techniques from the various approaches identified in the literature review (Chapter Two). A comparative evaluation of the performance of each technique, used in conjunction with objective criteria, selected support vector machines as the most appropriate technique for this research.

A framework consisting of seven criterions selected five techniques to trial to ensure the suitability of each technique to the research. The elements of the conceptual design (inferring engineering knowledge and predicting the probability of failure) were evident in the criteria and model construction. The selected modelling techniques to trial were logistic regression, neural networks, support vector machines, probability trees and random forests.

Four performance measures assessed the suitability of each of these techniques further, including the percentage of correctly classified sites (accuracy), percentage of misclassified sites (misclassification error), F-score and phi coefficient. As accuracy is a poor measure of a model's performance, other measures were explored to ensure a thorough and appropriate evaluation of the techniques, through the use of box and whisker plots, density distributions, and hypothesis testing. From this analysis, the performance of neural networks and random forests exceeded that of the other techniques and the accuracy results identified the successful factor combinations for each failure mechanism, which correlated well with the developed failure charts.

The second set of criteria in addition to the performance of the modelling technique further included an assessment of the running speed, interpretability and ease of use of the model, and generalisation properties of the technique to avoid over-fitting and promote transferability of the trained model to other road network datasets. From this analysis, support vector machines and probability trees outperformed the other classifiers.

Despite the simplicity and interpretability of probability trees, this chapter selected the support vector machines approach for the prototype system. In particular, the comparative study concluded this classification technique to be the most suitable and successful for modelling the road pavement data included in the State Highway LTPP dataset. The development of the system using the chosen modelling techniques is the focus of the subsequent chapter.

Chapter Six

DEVELOPMENT OF THE PROTOTYPE SYSTEM

Predicting the Probability of Failure

6.1 Introduction

The aim of this research is to produce a diagnostic tool capable of quantifying the probability (likelihood) of road pavement failure. To this end, Chapter Three presented the conceptual design of the system, and Chapters Four and Five described the development of the components of the system, as summarised in Figure 6-1. This chapter therefore presents the development of a fully working prototype system to address Objective Three, based on the conceptual design, incorporating the outputs and models developed in Chapters Four and Five.

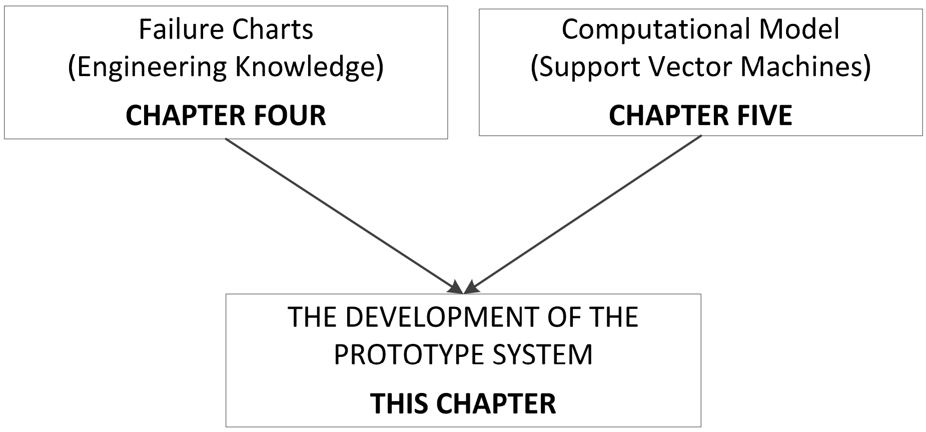


Figure 6-1: The development of the prototype system

6.1.1 Conceptual Framework

The framework of the prototype system includes:

1. Failure charts to diagnose the cause of failure (refer Chapter Four), and
2. A computational model to calculate the probability of road pavement failure (refer Chapter Five).

Failure Charts:

Failure charts (developed in Chapter Four) represent the diagnostic element of the model design. These charts identify the possible failure paths of the failure mechanisms of interest to this research and are based on the inclusion of the five failure factor groups described in Section 4.3. They are utilised in the prototype system to help determine the most probable (critical) failure path and the causes of failure. The predominant failure mechanisms are included in the prototype system herein, namely rutting, fatigue cracking, and shear.

Computational Model:

The support vector machines modelling technique was selected from those investigated in Chapter Five as the most suitable for predicting the likelihood of road pavement failure from the available datasets.

6.2 Development of the Prototype System

Four processes are used to determine the likelihood of pavement failure from road datasets. These are:

1. Determine the failure factors contributing to failure (refer Chapter Four);

2. Model each combination of the failure factors, as per Table 5-4, to determine the successful factor combinations. Each successful combination (refer Chapter Five) represents a different failure path on the developed failure charts;
3. Calculate the probability of each failure mechanism using a computational model to identify from the available research dataset the combinations of factors which could cause failure, based on an accuracy percentage of 100 % for rutting and fatigue cracking failures and 90 % for shear failure. The probability ($P_{Failure}$) for rutting, fatigue cracking, and shear failure is calculated using Equation 6-1, assuming the most probable failure path, and

Equation 6-1

$$P_{Failure} = max[P(A), P(B), \ldots, P(N)]$$

$$where\ A = Failure\ path\ A, B = Failure\ path\ B, and\ N = Failure\ path\ N$$

4. Assess the interactions between the failure mechanisms to calculate the overall failure probability. For example, these calculations may take into account multiple failure mechanisms occurring simultaneously on the pavement section or independence between the failure mechanisms (see Section 6.5.2).

6.2.1 Summarising the Components of the System

Chapter Four identified the five failure factor groups contributing to failure as traffic, composition, strength, environment, and subgrade sensitivity, which were considered in the computational models. Surface condition was also included in the models.

Chapter Five identified the successful factor combinations using the support vector machines technique, as shown in Table 6-1. Each combination was superimposed on the developed failure charts, as shown in Figures 6-2, 6-3, and 6-4, to represent each possible failure path as identified by Table 5-3. From herein, the term *'failure path'* refers to these combinations. The combinations marked with an asterisk in Table 6-2 do not feature on the charts presented here as they may be either a combination of other failure paths or are not significant causal factors. A detailed description of failure causes and path can be found in Appendix E.

Table 6-1: Successful Factor Combinations for the Support Vector Machines Technique

	Factor Combinations	Accuracy (%)	Misclassification Error (%)	Total Number of Combinations
Rutting	3, 7, 8, 12, 16, 18, 22, 25, 26, 28, 32, 34, 39, 42, 44, 46, 49, 53, 58	100	0	**19**
Fatigue Cracking	3, 7, 8, 12, 18, 22, 23, 24b, 25, 26, 28, 32, 33ab, 34, 42, 43ab, 44, 46, 47b, 49, 50ab, 52ab, 53, 54ab, 57ab, 58, 59ab, 62ab, 63ab	100	0	**38**
Shear	7, 12, 18, 22, 25, 28, 32, 34, 42, 44, 49, 53, 58	90	10	**13**

Table 6-2: Factor Combinations of the State Highway LTPP Dataset per Failure Mechanism

Trial	Combinations of Factors	Rutting	Fatigue Cracking	Shear
3	Strength	*Yes	*Yes	-
7	Traffic + Composition	Yes	Yes	Yes
8	Traffic + Strength	Yes	Yes	-
12	Composition + Strength	Yes	Yes	Yes
16	Strength + Environment	*Yes	-	-
18	Strength + Subgrade Sensitivity	Yes	Yes	Yes
22	Traffic + Composition + Strength	Yes	Yes	Yes
23	Traffic + Composition + Environment	-	Yes	-
24	Traffic + Composition + Surface Condition	-	Yes	-
25	Traffic + Composition + Subgrade Sensitivity	Yes	Yes	Yes
26	Traffic + Strength + Environment	Yes	Yes	-
28	Traffic + Strength + Subgrade Sensitivity	Yes	Yes	Yes
32	Composition + Strength + Environment	Yes	Yes	Yes
33	Composition + Strength + Surface Condition	-	Yes	-
34	Composition + Strength + Subgrade Sensitivity	Yes	Yes	Yes
39	Strength + Environment + Subgrade Sensitivity	Yes	-	-
42	Traffic + Composition + Strength + Environment	Yes	Yes	Yes
43	Traffic + Composition + Strength + Surface Condition	-	Yes	-
44	Traffic + Composition + Strength + Subgrade Sensitivity	Yes	Yes	*Yes
46	Traffic + Composition + Environment + Subgrade Sensitivity	Yes	Yes	-
47	Traffic + Composition + Surface Condition + Subgrade Sensitivity	-	Yes	-
49	Traffic + Strength + Environment + Subgrade Sensitivity	Yes	Yes	Yes
50	Traffic + Strength + Surface Condition + Subgrade Sensitivity	-	Yes	-
52	Composition + Strength + Environment + Surface Condition	-	Yes	-
53	Composition + Strength + Environment + Subgrade Sensitivity	Yes	Yes	Yes
54	Composition + Strength + Surface Condition + Subgrade Sensitivity	-	Yes	-
57	Traffic + Composition + Strength + Environment + Surface Condition	-	Yes	-
58	Traffic + Composition + Strength + Environment + Subgrade Sensitivity	*Yes	*Yes	*Yes
59	Traffic + Composition + Strength + Surface Condition + Subgrade Sensitivity	-	Yes	-
62	Composition + Strength + Environment + Surface Condition + Subgrade Sensitivity	-	Yes	-
63	Traffic + Composition + Strength + Environment + Surface Condition + Subgrade Sensitivity	-	*Yes	-
	TOTAL	**19**	**38**	**13**

- Not applicable
* Not shown on the failure charts

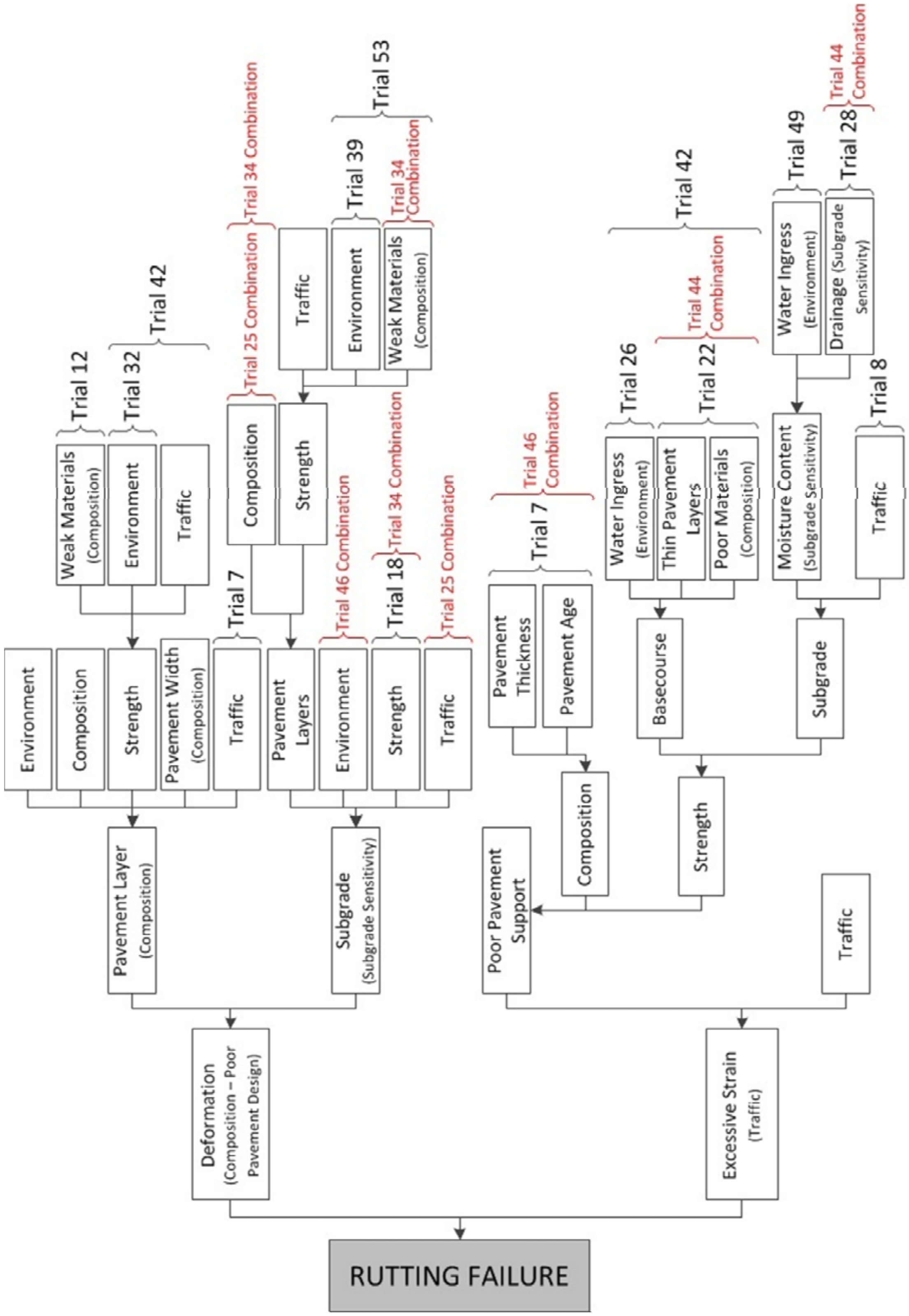


Figure 6-2: Rutting failure paths

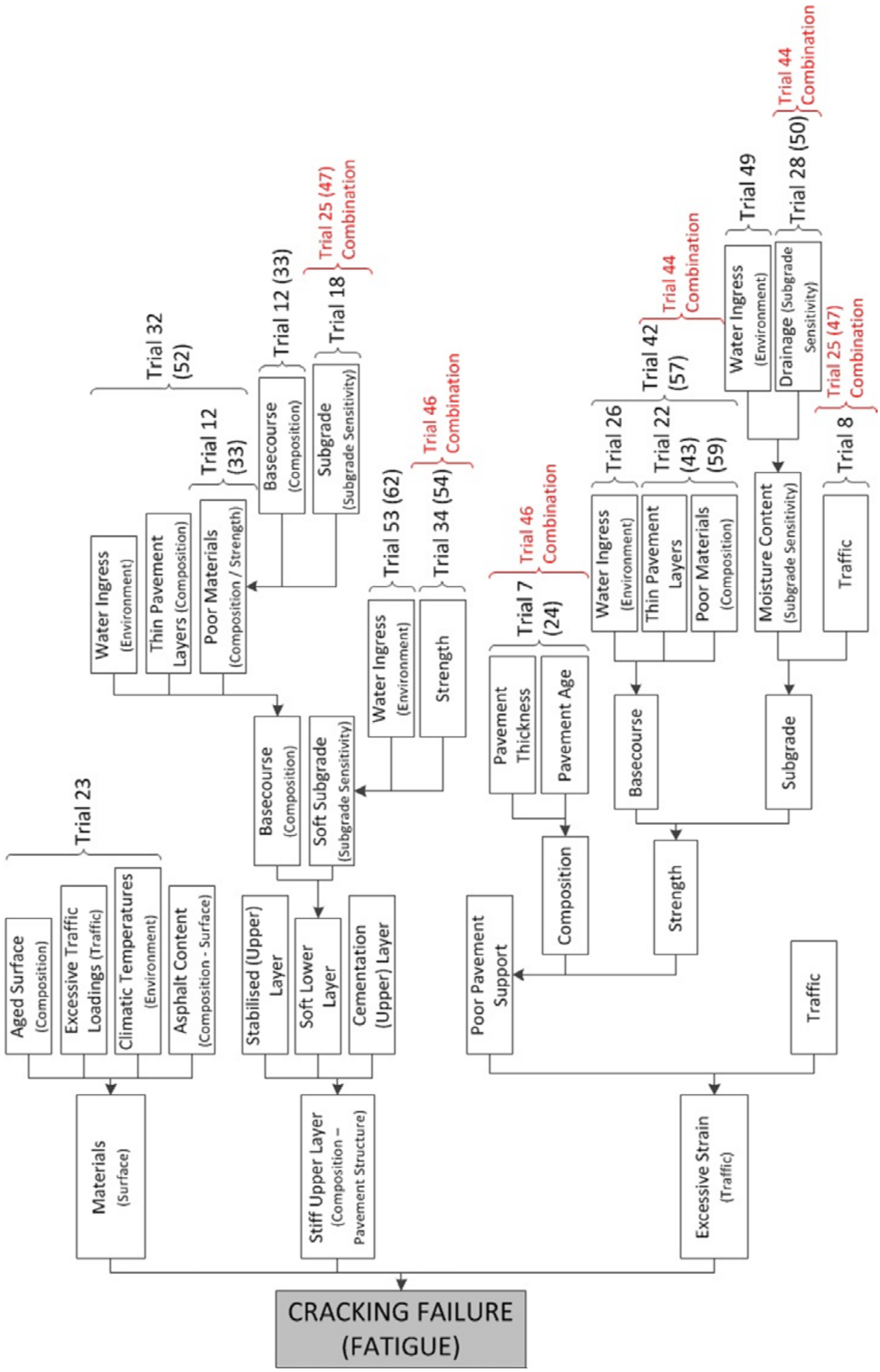


Figure 6-3: Fatigue cracking failure paths

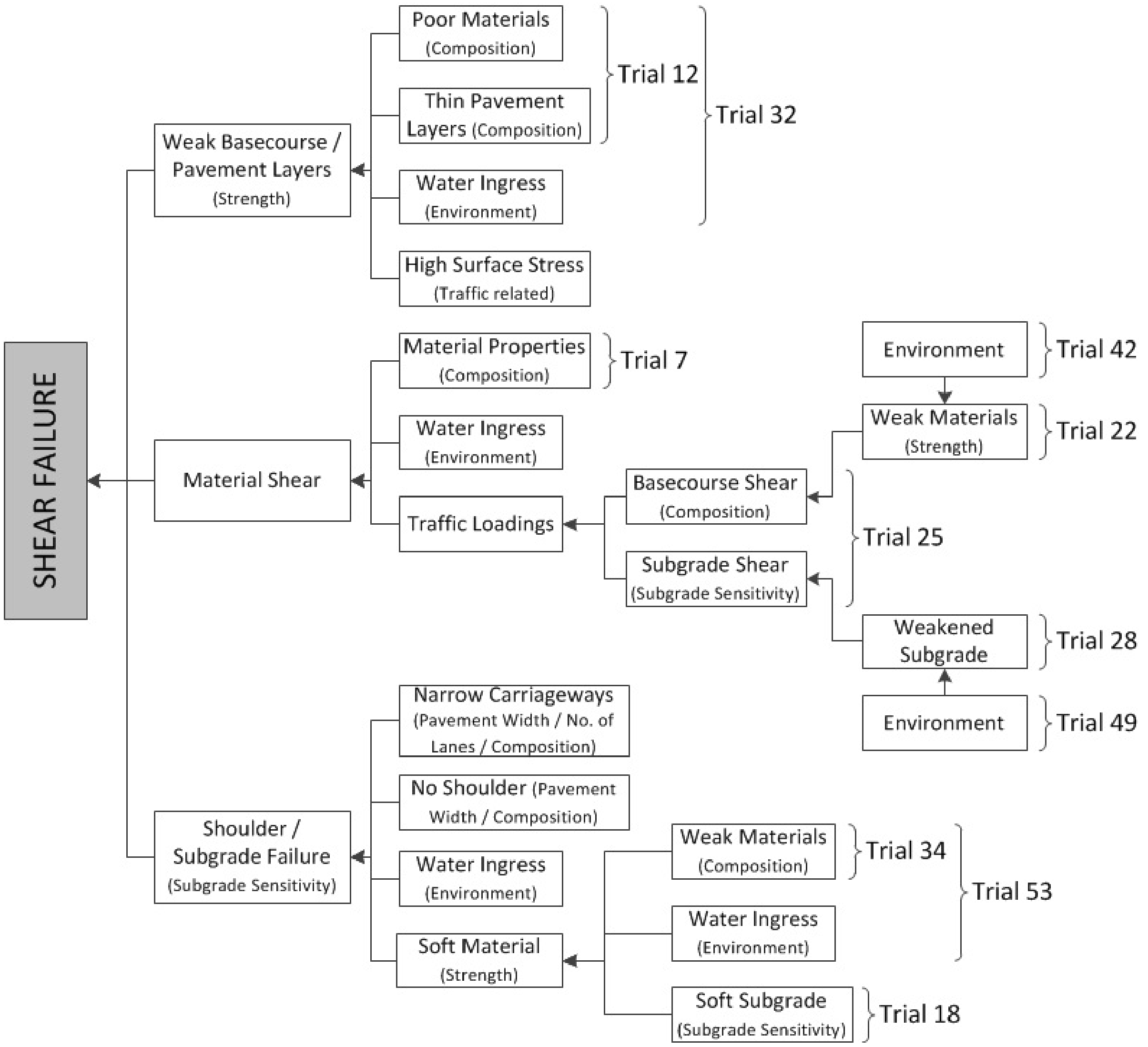


Figure 6-4: Shear failure paths

6.3 Dataset for the Development of the Prototype System

Data from the State Highway LTPP road network (refer Section 3.3.1) was used to further develop and demonstrate the performance of the prototype system. The support vector machines computational model was trained using 90 % of the data, via a 10-fold cross-validation approach (refer Section 5.3.3) (Rogers and Girolami, 2012). Accordingly, the calculations of the performance measures of the prototype system (see Section 6.4) used the predictions (new observations) from the remaining 10 % of the dataset. However, the analysis of the overall failure probability of the State Highway LTPP road network was based on the entire dataset (see Section 6.6).

6.4 Performance of the System

The prototype system was evaluated using the performance measures discussed in Chapter Five (refer Section 5.3.2). To this end, the trained models were evaluated by comparing the failed sites predicted by the support vector machines technique against the number recorded to have failed in the dataset. This was carried out on a failure path basis and the results are shown in Tables 6-3, 6-4, and 6-5. For example, it may be seen from Table 6-3 below that the trained model reported an accuracy of 98 % in predicting rutting failure based on failure path 7. As the system modelled each failure path separately, the performance of the system is based on the results from each factor combination.

6.4.1 Assessment of the System

Tables 6-3, 6-4, and 6-5 show the results the performance analysis of the system in terms of the four measures used. From the results, it may be seen that the system is able to predict the occurrence of rutting and fatigue cracking to a high degree of accuracy (with an average of approximately 98 % in both cases), but less so shear failure (94 %), similar to the poor results presented in Chapter Five for predicting shear failure. This suggests the independent variables do not correlate well with the predictions, such that the dataset does not contain appropriate information for predicting shear failures. The F-Score (99 % for both rutting and fatigue cracking, and 97 % for shear) indicated the superior accuracy and effectiveness of the system to predicting failure. Overall, the results show consistency for each of the factor combinations in terms of the accuracy, misclassification error, and F-score measures, which demonstrates the success of informing the model with the engineering knowledge captured in the failure charts. However, the phi coefficient (average of 0.17, 0.28, and 0.13 for rutting, fatigue cracking, and shear respectively) indicated a weak correlation between the inputs of the system and the predicted failure probabilities, despite the success of the support vector machines technique reported in Figure 5-6. However, in each table, a number of failure paths improve on this average, suggesting that the fatigue cracking and rutting systems in particular sufficiently model the behaviour of pavement failure. The poor result of the shear failure system, similar to that reported earlier in Figure 5-6, further substantiates the reported influence of the aggregate behaviour on shear failures in the literature, as material types were excluded in the research dataset. A phi coefficient equal to zero (0) indicates no correlation between the inputs and predicted output. Although this result can also occur when one of the binary classes is not predicted in any of the 10 cross-validation tests, resulting in the sum of a predicted class is equal to zero (0), the results reported in the tables below list the phi coefficients which were unable to be calculated with a *'-'* to avoid any confusion.

The coefficient of variation was calculated using Equation 6-2 (Madsen, 2011) allowing a direct comparison between populations with different means; therefore, the four performance measures across the three failure mechanisms can be compared. In the tables below, the accuracy and F-score measures are less dispersed than the misclassification error measure.

Equation 6-2

$$Coefficient\ of\ Variation = \frac{Standard\ Deviation}{Mean} \times 100 = \frac{\sigma}{\mu} \times 100$$

Table 6-3: Assessment of the Performance Measures for Rutting Failure (Average of 10-Fold Cross-Validation)

Failure Paths	Accuracy (%)	Misclassification Error (%)	F-Score	Phi Coefficient
3	97.651	2.349	0.988	-
7	97.865	2.135	0.989	0.341
8	97.651	2.349	0.988	-
12	97.794	2.206	0.989	0.301
16	97.669	2.331	0.988	0.114
18	97.651	2.349	0.988	-
22	97.794	2.206	0.989	0.301
25	97.794	2.206	0.989	0.301
26	97.669	2.331	0.988	0.128
28	97.651	2.349	0.988	-
32	97.633	2.367	0.988	0.163
34	97.794	2.206	0.989	0.301
39	97.705	2.295	0.988	0.153
42	97.651	2.349	0.988	0.168
44	97.794	2.206	0.989	0.301
46	97.633	2.367	0.988	0.210
49	97.722	2.278	0.988	0.181
53	97.616	2.384	0.988	0.158
58	97.633	2.367	0.988	0.163
Minimum	**97.616**	**2.135**	**0.988**	**0.114**
Average	**97.704**	**2.296**	**0.988**	**0.219**
Maximum	**97.865**	**2.384**	**0.989**	**0.341**
Coefficient of Variation	**0.5**	**20.1**	**0.3**	**15.7**[1]

[1] – Calculated with the *'-'* omitted

Table 6-4: Assessment of the Performance Measures for Shear Failure (Average of 10-Fold Cross-Validation)

Failure Paths	Accuracy (%)	Misclassification Error (%)	F-Score	Phi Coefficient
7	94.520	5.480	0.972	0.186
12	94.431	5.569	0.971	0.149
18	94.448	5.552	0.971	-
22	94.555	5.445	0.972	0.196
25	94.626	5.374	0.972	0.202
28	94.448	5.552	0.971	-
32	94.520	5.480	0.972	0.138
34	94.591	5.409	0.972	0.188
42	94.502	5.498	0.972	0.133
44	94.662	5.338	0.972	0.214
49	94.413	5.587	0.971	0.053
53	94.520	5.480	0.972	0.133
58	94.520	5.480	0.972	0.138
Minimum	**94.413**	**5.338**	**0.971**	**0.053**
Average	**94.520**	**5.480**	**0.972**	**0.157**
Maximum	**94.662**	**5.587**	**0.972**	**0.214**
Coefficient of Variation	**0.5**	**7.9**	**0.2**	**-**[1]

[1] – Cannot be calculated due to a negative standard deviation (see Appendix F)

Table 6-5: Assessment of the Performance Measures for Fatigue Cracking Failure (Average of 10-Fold Cross-Validation)

Failure Paths	Accuracy (%)	Misclassification Error (%)	F-Score	Phi Coefficient
3	98.078	1.922	0.990	-
7	98.274	1.726	0.991	0.373
8	98.078	1.922	0.990	-
12	98.274	1.726	0.991	0.373
18	98.078	1.922	0.990	-
22	98.220	1.779	0.991	0.358
23	98.292	1.708	0.991	0.363
24b	98.149	1.851	0.991	0.261
25	98.220	1.779	0.991	0.358
26	98.007	1.993	0.990	0.063
28	98.078	1.922	0.990	-
32	98.381	1.619	0.992	0.406
33a	98.185	1.815	0.991	0.275
33b	98.149	1.851	0.991	0.261
34	98.274	1.726	0.991	0.373
42	98.381	1.619	0.992	0.417
43a	98.221	1.779	0.991	0.296
43b	98.185	1.815	0.991	0.282
44	98.221	1.779	0.991	0.383
46	98.292	1.708	0.991	0.374
47b	98.167	1.833	0.991	0.275
49	98.007	1.993	0.990	0.063
50a	98.185	1.815	0.991	0.267
50b	98.149	1.851	0.991	0.253
52a	98.221	1.779	0.991	0.303
52b	98.256	1.744	0.992	0.334
53	98.452	1.548	0.992	0.447
54a	98.149	1.851	0.991	0.253
54b	98.149	1.851	0.991	0.261
57a	98.221	1.779	0.991	0.303
57b	98.274	1.726	0.991	0.346
58	98.381	1.619	0.992	0.417
59a	98.221	1.779	0.991	0.296
59b	98.167	1.833	0.991	0.275
62a	98.203	1.797	0.991	0.296
62b	98.238	1.762	0.991	0.328
63a	98.221	1.779	0.991	0.303
63b	98.256	1.744	0.991	0.334
Minimum	**98.007**	**1.548**	**0.990**	**0.063**
Average	**98.209**	**1.791**	**0.991**	**0.311**
Maximum	**98.452**	**1.993**	**0.992**	**0.447**
Coefficient of Variation	**0.6**	**35.5**	**0.3**	**1.0**[1]

[1] – Calculated with the *'-'* omitted

6.5 Overall Failure Probability

The development of the computational model described above and in Chapter Five has assumed that each of the three failure mechanisms act independently. However, in practice, pavement failure can occur simultaneously as another failure or as a secondary effect of another occurrence of failure. For example, rutting can act alone on the pavement or it can cause a secondary effect of cracking in the wheelpaths. Alternatively, cracked pavements permit water into the lower layers of the pavement resulting in rutting. To address this phenomenon in the computational model, this research explored probability theory to account for the interactions between the failure mechanisms.

6.5.1 Terminology

A number of terms are used to describe inter and independence of events in probability theory (Mendenhall and Beaver, 1991). Definitions of interest to the problem described above are:

> **Mutually Exclusive:** A mutually exclusive event is one in which two events cannot happen at the same time (Ayyub and McCuen, 2003; Mendenhall and Beaver, 1991), as shown in Figure 6-5. The fact that any of the three defects considered herein can be present on a road section with any of the other defects means that their occurrence is **not** mutually exclusive.

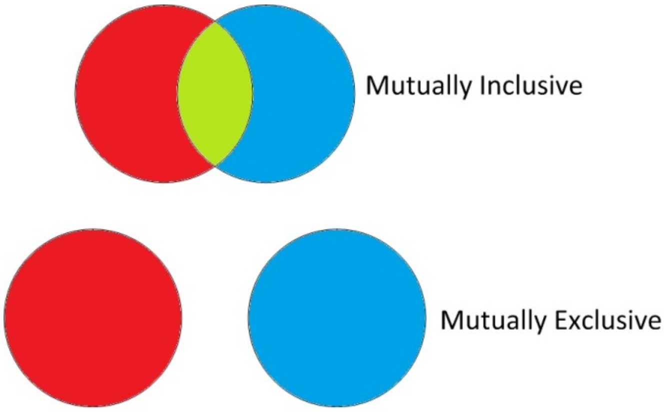


Figure 6-5: Exclusivity terminology

Independence: The occurrence of one event has no influence on another, and provides no further information on the occurrence of another event (Ayyub and McCuen, 2003; Mendenhall and Beaver, 1991). For example, cracking occurring as a secondary effect of rutting would indicate these events are dependent (**not** independent). An example of dependence is shown in Figure 6-6.

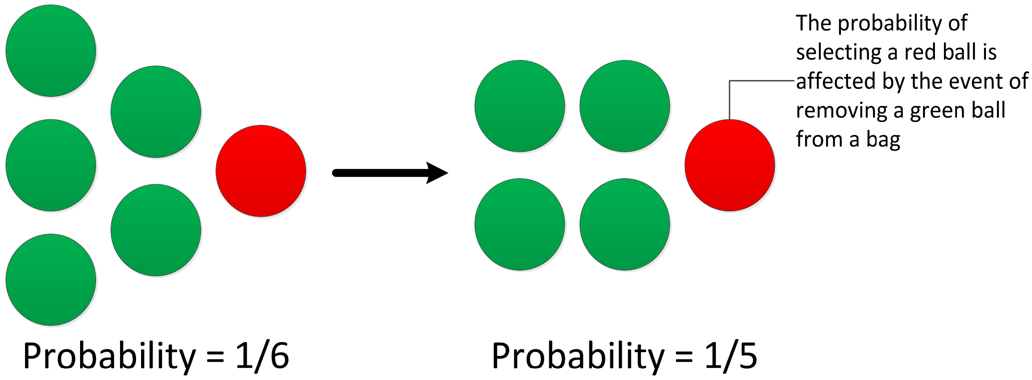


Figure 6-6: The effect of dependence

Probability Unions: The *'Additive Law of Probability'* (Ayyub and McCuen, 2003; Mendenhall and Beaver, 1991) defines the probability of the occurrence of two events as:

Equation 6-3

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If the definition of events A and B is mutually exclusive, $P(A \cap B) = 0$, therefore Equation 6-3 therefore becomes (Mendenhall and Beaver, 1991):

Equation 6-4

$$P(A \cup B) = P(A) + P(B)$$

If events are defined as independent, Equation 6-3 becomes (Ayyub and McCuen, 2003; Mendenhall and Beaver, 1991):

Equation 6-5

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - [P(A) \times P(B)] \end{aligned}$$

6.5.2 Overall Failure Probability Approaches

Given the complexity of the problem of calculating the overall failure probability, three approaches were investigated as follows:

1. **Conservative Approach:** Where the overall failure probability is calculated from the maximum probability of each of the three failure mechanisms (refer Equation 3-1);
2. **Probabilistic Equation:** The overall failure probability is calculated using Equation 6-3, assuming the failure mechanisms are independent of each other (refer Equation 3-2), and
3. **Modelling Joint Failure:** A computational model is trained using the available data, including combined failures, to calculate the overall failure probability.

6.5.2.1 Maximum Failure Probability

With this conservative approach, the maximum of the probabilities from three individual failure mechanisms is taken as the overall failure probability ($P_{Failure}$) of the road section under consideration, as per Equation 6-6:

Equation 6-6

$$P_{Failure} = max\left[P_{Rutting}, P_{Cracking}, P_{Shear}\right]$$

Advantages: This approach follows that of conventional pavement design in which the design of the pavement structure focuses on the critical failure mechanism(s), whether that is for example to prevent rutting or cracking (Austroads, 2012). This approach treats each failure criterion as independent.

Limitations: This approach assumes that each failure mechanism is independent of the others and, therefore, evidence of the likelihood of two or more failure mechanisms on a road section does not increase the probability of failure of the section.

6.5.2.2 Probabilistic Theory

The laws of addition of probabilities can be used to calculate the probability of failure by any one, two or three of the failure mechanisms acting concurrently. (refer Section 3.2.4). The law can be written using *'set'* notation as follows (Ayyub and McCuen, 2003; Mendenhall and Beaver, 1991):

Equation 6-7

$$\begin{aligned} P_{Failure} &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= P_{Rutting} + P_{Cracking} + P_{Shear} - P(Rutting \cap Cracking) \\ &\quad - P(Rutting \cap Shear) - P(Cracking \cap Shear) \\ &\quad + P(Rutting \cap Cracking \cap Shear) \end{aligned}$$

Note if one failure mode is considered not to influence the occurrence of the other (e.g. the failure mechanisms are independent), then $P(A \cap B) = P(A) \times P(B)$ and Equation 6-7 can be written as:

Equation 6-8

$$P_{Failure} = P_{Rutting} + P_{Cracking} + P_{Shear} - [P_{Rutting} \times P_{Cracking}] - [P_{Rutting} \times P_{Shear}] - [P_{Cracking} \times P_{Shear}] + [P_{Rutting} \times P_{Cracking} \times P_{Shear}]$$

Advantages: This approach takes into account the possibility of multiple failures occurring.

Limitations: Although this approach accounts for multiple failures in the calculations, there is not enough information in the dataset to determine the timings of failure. More specifically, it is not possible to determine from the dataset whether the failure modes that occurred simultaneously were in fact secondary effects of each other (e.g. one mode influenced the occurrence of another failure mode).

6.5.2.3 Computational Model

The overall failure probability could be determined using a modelling approach to predict the occurrence of combined failures. To achieve this, individual models would be developed for each combined failure occurrence, much like the individual failure paths in the failure charts (refer the methodology presented in Section 6.4). The model outputs would then predict the individual probabilities for each failure type and combined failures, as shown in Equation 6-9.

Equation 6-9

$$P_{Failure} = \begin{bmatrix} P_{Rutting}, P_{Cracking}, P_{Shear}, P_{Rutting+Cracking}, P_{Rutting+Shear}, \\ P_{Cracking+Shear}, P_{Rutting+Cracking+Shear} \end{bmatrix}$$

Advantages: This approach presents a superior method of calculating the overall failure probability.

Limitations: In order to replicate pavement failure accurately, sufficient data is required to train the computational model. As the State Highway LTPP dataset contains only a small number of multiple failure occurrences (see Table 6-6) and limited information on the occurrences of multiple failure modes, it was not possible to develop such a process in this research.

Table 6-6: Distribution of Failures in the State Highway LTPP Dataset

	Occurrences in the Dataset	
Failure Mechanism	**Number**	**Percentage**
Rutting	132	2.35
Fatigue Cracking	108	1.92
Shear	312	5.54
Rutting and Fatigue Cracking	36	0.64
Rutting and Shear	108	1.92
Fatigue Cracking and Shear	84	1.49
Rutting and Fatigue Cracking and Shear	24	0.43

6.5.3 Selection of the Most Suitable Approach

The above section explored three approaches to determine an overall failure probability for a road section. The third approach was found to be difficult to achieve using the available dataset and therefore will not be considered further. For the purposes of the prototype system, it was assumed the failure mechanisms were independent, such that one mode of failure has no influence on the occurrence of another. For a more accurate assessment, the dataset would need to include more information regarding the timing of failure types and the independence between failure mechanisms. Consequently, approaches one and two described above were further explored using the State Highway LTPP dataset, as described below.

6.6 Analysis of the State Highway Road Network

In order to compare the results obtained from the two approaches described above, distribution plots of the overall failure probabilities across the State Highway LTPP road network were generated and compared. The performance measures, defined in Section 5.3.2, were also used to assist this process.

6.6.1 Conservative Approach (Maximum Probability)

Figure 6-7 presents the failure probability distribution of the network using the conservative (maximum probability) approach to calculate the overall failure. By inspection, the actual failure data, where the failure variable was converted to an estimate of the probability density function, of this network indicated a similar failure distribution, as shown by the black density curve in Figure 6-7. A large number of sites in the State Highway LTPP dataset have not failed, hence the majority of the distribution has a failure probability of less than 20 %.

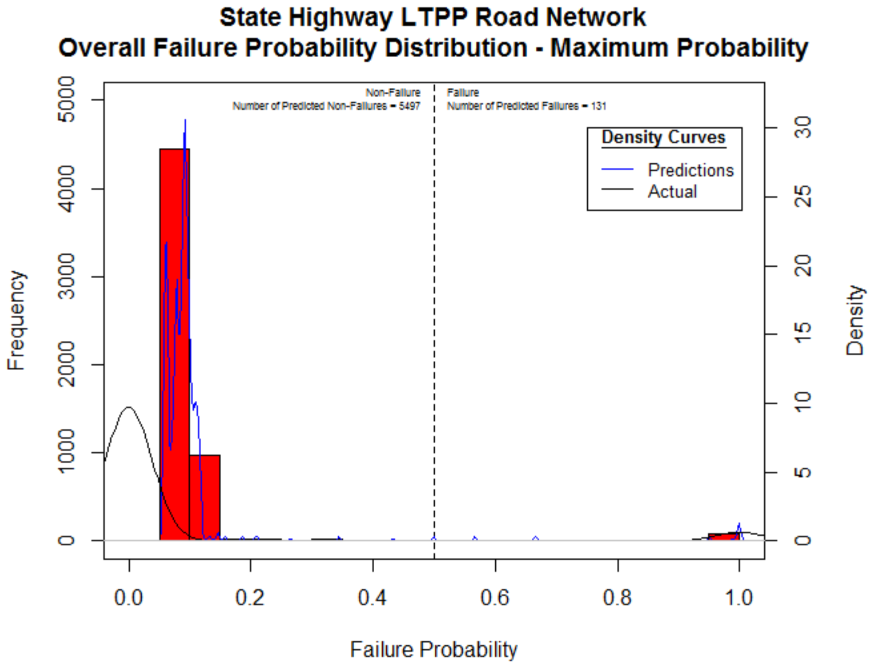


Figure 6-7: Distribution of the overall failure probabilities, calculated using the maximum probabilities, of the State Highway LTPP road network

6.6.2 Probability Theory Approach (Probability Equation)

Figure 6-8 presents the distribution of failure probabilities determined using the second (probability theory) approach. In comparison to Figure 6-7, although the peak of the distribution is smaller than obtained by the first approach, the shape of the distribution remains the same with a slight shift to the right suggesting that more sites have a higher likelihood of failure. This result could be expected given the addition of the individual failure probabilities in the equation (refer Equations 6-7 and 6-8).

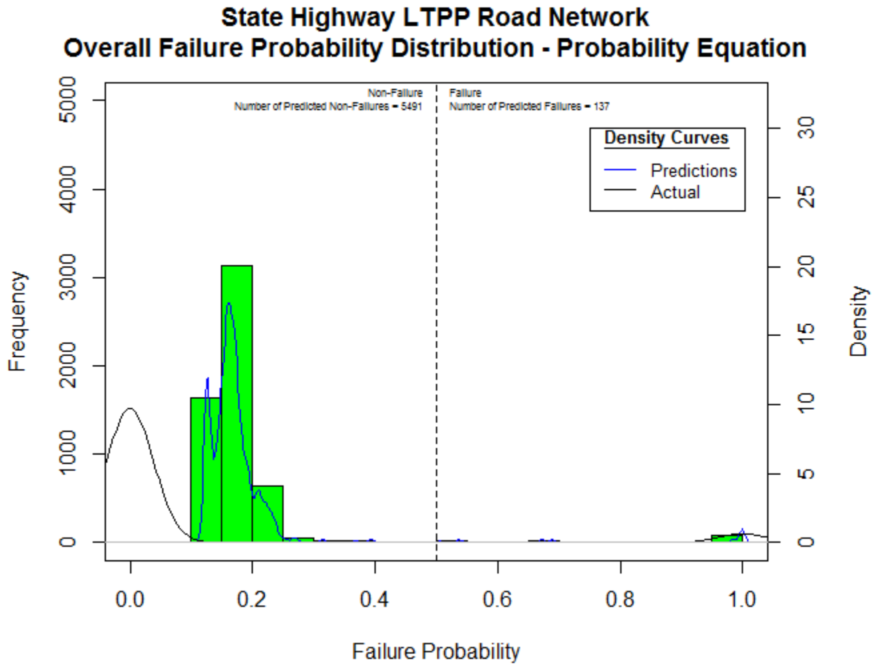


Figure 6-8: Distribution of the overall failure probabilities, calculated using the probability equation, of the State Highway LTPP road network

6.6.3 Comparison of the Two Approaches

Figure 6-9 directly compares the distributions of the two approaches, where the differences in the distributions, such that the overall failure probability calculated using the probability theory (probability equation) approach would always exceed that of the conservative (maximum probability) approach, is apparent.

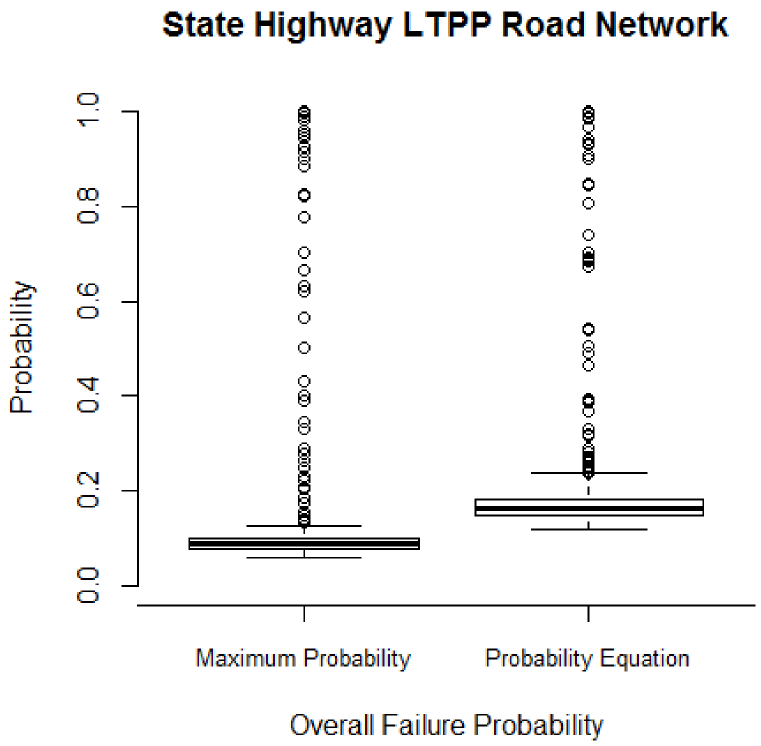


Figure 6-9: A comparison of the overall failure probability approaches

To further compare the results from the two approaches, confusion matrices (refer Chapter Five) were generated, as shown in Figure 6-10, based on the outputs using a failure threshold of $P(X \geq 0.5) = 1$. The results for the two approaches are very similar, with the only difference being the very small increase (0.107 %) in the number of misclassified non-failures from the second (probability equation) approach (Figure 6-10b).

DATA FAILURES

PREDICTIONS	**0**	**1**	
0	5226	271	5497
1	54	77	131
	5280	348	**5628**

(a) from the Conservative Approach (Maximum Probability)

DATA FAILURES

PREDICTIONS	**0**	**1**	
0	5220	271	5491
1	60	77	137
	5280	348	**5628**

(b) from the Probability Theory Approach (Probability Equation)

Figure 6-10: Confusion matrix of the results

The data in Figure 6-10 was further analysed; the results from the performance measures for each approach is summarised in Table 6-7. The results of these analyses show that in terms of the performance measures considered there is an insignificant difference between the two approaches and that either approach may be suitable for the dataset considered.

Table 6-7: Performance Measures for the Conservative and Probability Theory Approaches

	Accuracy	Misclassification Error	F-Score	Phi Coefficient
Conservative Approach (Maximum Probability)	94.23 %	5.77 %	0.97	0.34
Probability Theory Approach (Probability Equation)	94.12 %	5.88 %	0.97	0.33

6.7 Summary of the Prototype System Development

This chapter has presented a prototype system which includes failure charts (developed in Chapter Four) and a computational model (selected in Chapter Five). The system comprises of the following four sequential steps:

- Determining the failure factor groups;
- Modelling each possible combination of the factor groups to select the successful combinations;
- Using these combinations to calculate the probability of each failure mechanism; and
- Assessing the approaches to determine the overall failure probability using the predicted probability outputs from the prototype system for each failure mechanism.

Chapter Five identified the successful factor combinations and this chapter presented them on the developed failure charts (from Chapter Four). Each combination represents a failure path for each failure mechanism. Not all the same combinations are present on each failure chart, as the mechanisms associated with each failure are different. However, the successful factor combinations demonstrate the key factor groups and failure paths appropriately for each failure mechanism.

To calculate the failure probabilities of each failure mechanism, the probability associated with the most probable failure path (e.g. the maximum probability for the failure mechanism) was selected. With this approach, the factor combination associated with the greatest probability becomes the critical failure path. Identifying this path and combination of factors on the failure paths indicates the most likely causes of the failure.

This chapter assessed the performance of the prototype system based on the four performance measures introduced in Chapter Five, from which it was found that the performance of the support vector machine models was satisfactory.

To calculate the overall failure probability, three approaches were investigated and, of these, two were selected for further analysis. The dataset did not contain information on the timings of failure, in order to determine if failure of two or more mechanisms influenced the occurrence of another, which inhibited the implementation of the computational modelling approach.

The two approaches selected were used to produce the distributions of the overall failure probabilities for the State Highway LTPP dataset. The first (conservative) approach selected the maximum probability from the three failure mechanisms as the overall failure probability.

The second approach used the principles of probability unions to develop an equation to combine the individual probabilities. A comparison of the two approaches failed to demonstrate a significant difference between them in terms of the failure probability distributions of the road sections analysed.

The following chapter further assesses the performance of the developed system, given both approaches, as well as discussing the practical applications of the research outcomes.

Chapter Seven

NEW ZEALAND CASE STUDY
Practical Applications of the System

7.1 Introduction

To demonstrate the effectiveness of the prototype system and the practical applications of this research outcome, as stated in Objective Four, this chapter presents the results of the analysis of an independent dataset from the New Zealand LTPP programme. This case study aims to use the developed prototype system to:

- Calculate the probability of failure for road sections in the dataset, and determine the most probable failure mechanism(s);
- Identify the critical failure path and causes of failure;
- Identify the symptomatic problems across the pavement sites included the dataset and road network, and
- Critically analyse the shifts in the probability distributions to quantify the effect the changes in the environment have on the road network, such as an increase in traffic loadings acting alone or coupled with abnormal weather.

The above includes both project and network level applications of the prototype system. With such an analysis, asset managers can make informed decisions on the future maintenance demands of their network, and identify potential failures. This chapter discusses the results from the analysis illustrating the performance of the sites included in the Local Authority LTPP road dataset.

7.1.1 Methodology for Case Study

To demonstrate the use of the system, the case study followed the methodology below.

1. Normalise the testing dataset using the assumption of a straight-line transformation (refer Chapter Five);
2. Process the data using the support vector machines models developed in Chapter Six (note: the models were not retrained using the testing dataset; instead the existing models from Chapter Six are used);
3. Calculate the individual failure probabilities per failure mechanism using Equation 6-1 (refer Chapter Six);
4. Assess the accuracy of the predictions from the system using the predicted failure probabilities per failure mechanism, and
5. Compare the overall failure probability using the two approaches discussed in Chapter Six (Equations 6-5 and 6-6, respectively).

Statistical plots, such as histograms, barplots, and pie charts, were used to gain an understanding of the network performance from the predicted probabilities.

7.1.2 Local Authority LTPP Dataset

The Local Authority LTPP dataset (refer Section 3.3.2) was processed using the prototype system developed in Chapter Six to further test the performance of the system on an independent dataset, one which was not used to develop and refine the prototype system, and evaluate the transferability of the system to other road networks. The similar data collection methods employed under the LTPP programme (refer Section 3.3) consequently led to similar variables included in the Local Authority LTPP dataset as those included in the State Highway LTPP dataset, and subsequently the successful implementation of the prototype system to another dataset. Alternative data collection methods are implemented by RCAs on typical local authority road networks, such as the network analysed in Schlotjes and Henning (2012), resulting in distinct differences in the variables included in the datasets and databases in comparison to the research dataset. Thus, evaluating the performance of the prototype system on such a network would vary the predictions for the pavements and ultimately mislead and falsify the performance results, given the expected high percentage of missing data from RCA road networks.

Testing of the system to real-life data, such as New Zealand RCA road networks, requires manipulation of the network data to maximise the amount of data available for the prototype system. Deng and Henning (2012) and Schlotjes and Henning (2012) tested the system with network data from a number of local New Zealand road networks.

Table 7-1 summarises the failure distribution of the Local Authority LTPP dataset by geographical region. Similar percentages of failure were seen on the State Highway LTPP road network (refer Table 6-6), although the discernible difference is a reduction of 1 % from the State Highway LTPP dataset in rutting failures, and an increase of 1.1 % for the Local Authority LTPP dataset in fatigue cracking failures. The majority of the road sites in the State Highway LTPP dataset, which was employed in the development of the prototype system, were located in rural environments. Therefore, to ensure consistent and indicative results from the system testing, this case study disregarded any urban sites included in the Local Authority LTPP dataset as the probabilities of these sites were not able to be appropriately predicted by the prototype system. Although the case study only concerned the rural LTPP sites and, as a result, reduced the number of datapoints included in the independent testing dataset, the rural sites provided 4136 datapoints for the testing dataset. The size of this testing dataset was similar to the size of the development dataset (State Highway LTPP dataset), which contained 4512 datapoints, and therefore can be considered sufficient for the purpose of this research.

The rural sites included in the LTPP programme are situated in several geographical regions of New Zealand. Generally, the traffic loadings on these sites is much less than the demand on the State Highway road network, resulting in a thinner pavement design and often lacking a sub-base layer. Despite this, a number of less popular State Highway routes are similar in design, construction, traffic and environmental conditions, and regional locations as the Local Authority road pavements, resulting in the verification of the prototype system using this dataset feasible.

Table 7-1: Distribution of Failures in the Local Authority LTPP Dataset

Failure Mechanism	Occurrences in the Dataset					
	Urban Sites (2340)		Rural Sites (4136)		ALL SITES (6476)	
	Number	Percentage	Number	Percentage	Number	Percentage
Rutting	36	1.54	48	1.16	84	1.30
Fatigue Cracking	96	4.10	100	2.42	196	3.03
Shear	180	7.69	204	4.93	384	5.93

7.1.3 Performance of the System

Table 7-2 presents a comparison between the recorded failures given in the rural Local Authority LTPP dataset and those predicted by the prototype system. Overall, the system was less accurate at predicting failure for the rural Local Authority LTPP dataset than that used for the development of the prototype system in Chapter Six. This was expected as using the system to analyse another road dataset, although of the same detail and including similar pavement types, was expected to bring about a reduction in performance as it is common for computational models to lose their predictive power and robustness through validation of the model with an independent dataset (Austroads, 2009b). To address this in practice, it is suggested further refinement of the system is required, which may include adequately redefining the developed failure charts when the system is used on other datasets and pavement types.

Unlike the results reported in Chapter Six, Table 7-2 shows a similarity between the accuracy of the rutting and shear failure systems. This result was unexpected and suggests that the failure mechanisms described in Figure 4-7, and subsequently the prototype system, may not include all methods of rutting failure included in this particular testing dataset. This also demonstrates the need to adequately redefine the developed failure charts when the system is used on other datasets. Furthermore, it suggests the performance of the rutting system is compromised, more so than the other two, when transferring the developed support vector machines models to other road datasets. Despite this, the performance of the prototype system, and subsequently the failure probability predictions, were considered adequate for the testing dataset, such that the transferability of the prototype system to another road dataset was successful.

Table 7-2: Accuracy of Predictions, based on the Confusion Matrices

Failure Modes	Accuracy (%)	Misclassification Error (%)
Rutting	79.25	20.75
Fatigue Cracking	85.76	14.24
Shear	78.32	21.68

7.2 Project Level Applications

The project level applications of the proposed system include the ability to:

- Quantify the probability of road pavement failure per road section;
- Identify the most probable failure mode, and
- Recognise the causes of failure.

Using these outputs in conjunction with the failure charts developed in Chapter Four facilitates an appropriate diagnosis of the cause(s) of failure and the selection of suitable maintenance treatment(s) for a road section. These aspects are discussed further in subsequent sections of this chapter.

7.2.1 Analysis of the Predictions

In order to demonstrate the project level applications of the prototype system, 12 sites were selected at random[15] from the entire rural Local Authority LTPP dataset, as presented in Table 7-3. The predicted failure state of each road section was predicted using the developed prototype system and Equation 6-6. Sections which have been predicted to have failed are indicated with a 1 in the '*Failure*' column. Then, for the purpose of this case study, the overall probability of failure was calculated using the conservative (maximum probability) approach (refer Section 6.5.2.1). The most probable failure mode and, subsequently, the associated most probable failure path are also presented in Table 7-3. In the instances where failure was not predicted to occur, the identified failure mode indicates the mechanism by which the road site is likely to fail in the future.

[15] Site numbers were randomly generated by the random number generator *sample* function in *R*.
random_sites <- sample (1 : 4136 , 12 , replace = F)

Table 7-3: Inferred Project Level Statistics from the Rural Local Authority LTPP Dataset

Site No.[16]	Actual Failure	Primary Failure Predictions				
		Failure	Probability of Failure	Failure Mode	Failure Path	Factors
512	Rutting	1	0.875	Rutting	22	Traffic + Composition + Strength
1855	Shear	1	0.742	Shear	22	Traffic + Composition + Strength
3286	None	0	0.081	Shear	18	Strength + Subgrade Sensitivity
4493	None	0	0.105	Shear	58	Traffic + Composition + Strength + Environment + Subgrade Sensitivity
4494	None	0	0.111	Rutting	44	Traffic + Composition + Strength + Subgrade Sensitivity
4962	Shear	1	0.506	Shear	42	Traffic + Composition + Strength + Environment
5204	Fatigue Cracking + Shear	1	0.808	Fatigue Cracking	7	Traffic + Composition
5396	Fatigue Cracking + Shear	1	0.821	Fatigue Cracking	7	Traffic + Composition
6473	None	0	0.083	Shear	18	Strength + Subgrade Sensitivity
1608	None	1	0.668	Rutting	7	Traffic + Composition
5434	None	1	0.571	Fatigue Cracking	7	Traffic + Composition
5571	None	1	0.554	Shear	49	Traffic + Strength + Environment + Subgrade Sensitivity

To further demonstrate the use of the system, Table 7-4 presents some additional information on secondary failure predictions for the 12 sites, which has been determined using the second greatest predicted probability of the three failure mechanisms.

[16] Despite the site numbers exceeding 4136, the sites reported in Table 7-3 are rural sites only. The urban sites are removed from the dataset yet the original site numbers remain from the entire Local Authority LTPP dataset.

Table 7-4: Secondary Failure Predictions of the Rural Local Authority LTPP Dataset

Site No.	Actual Failure	Primary Failure Predictions			Secondary Failure Predictions		
		Primary Failure	**Probability of Failure**	**Failure Mode**	**Probability of Failure**	**Failure Mode**	**Failure Path**
512	Rutting	1	0.875	Rutting	0.733	Shear	25
1855	Shear	1	0.742	Shear	0.086	Rutting	7
3286	None	0	0.081	Shear	0.037	Shear	7
4493	None	0	0.105	Shear	0.104	Rutting	44
4494	None	0	0.111	Rutting	0.109	Shear	58
4962	Shear	1	0.506	Shear	0.347	Rutting	49
5204	Fatigue Cracking + Shear	1	0.808	Fatigue Cracking	0.265	Rutting	22
5396	Fatigue Cracking + Shear	1	0.821	Fatigue Cracking	0.254	Rutting	26
6473	None	0	0.083	Shear	0.054	Rutting	25
1608	None	1	0.668	Rutting	0.171	Shear	22
5434	None	1	0.571	Fatigue Cracking	0.255	Rutting	26
5571	None	1	0.554	Shear	0.224	Rutting	49

Two sites are further explained to demonstrate the project level applications. Site # 512, as an example, is likely to fail in rutting with a probability of 0.875, which correlates well with the raw data. The predicted critical failure path for this site is # 22 (refer Table 7-3), indicating that a combination of the traffic, poor composition, and weak pavement strength are the contributing factors to the rutting failure. Figure 7-1 shows the use of the failure charts for this task, and with reference to the developed rutting failure chart presented in Chapter Four (Figure 4-7) and again in Chapter Six (Figure 6-2), failure path # 22 is defined as:

"Excessive Strain → Poor Pavement Support → Strength → Basecourse

→Thin Pavement Layers OR Poor Materials"

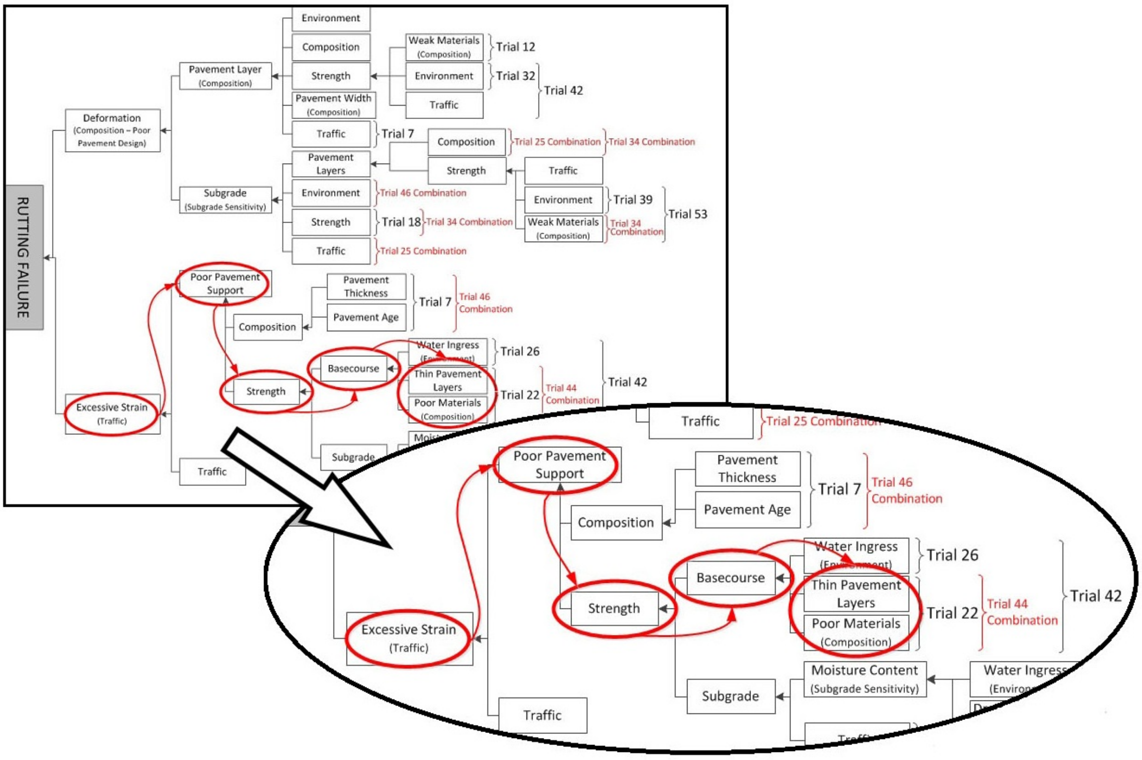


Figure 7-1: Critical failure path for site # 512

Possible causes of failure in this case could be an under-designed pavement structure, resulting in pavement layers that are too thin, or the use of poor aggregates, which are unable to spread the traffic load sufficiently to the underlying layers. A suitable maintenance treatment could be considered to be a granular structural overlay of suitable thickness to reduce the strains adequately in the underlying layers for the traffic loadings, thus preventing rutting from occurring in the future. The required thickness could be calculated using an appropriate overlay design procedure.

Site # 4962 failed in shear, and the prototype system predicted shear failure as the most probable failure mode with an associated probability of 0.506 and a predicted secondary failure of rutting with a probability of 0.347 (refer Table 7-4). The system identified the critical (shear) failure path as # 42 (refer Figures 4-11 and 6-4) and Figure 7-2 identifies the failure path as:

“Material Shear → Traffic Loadings → Basecourse Shear → Weak Materials → Environment”

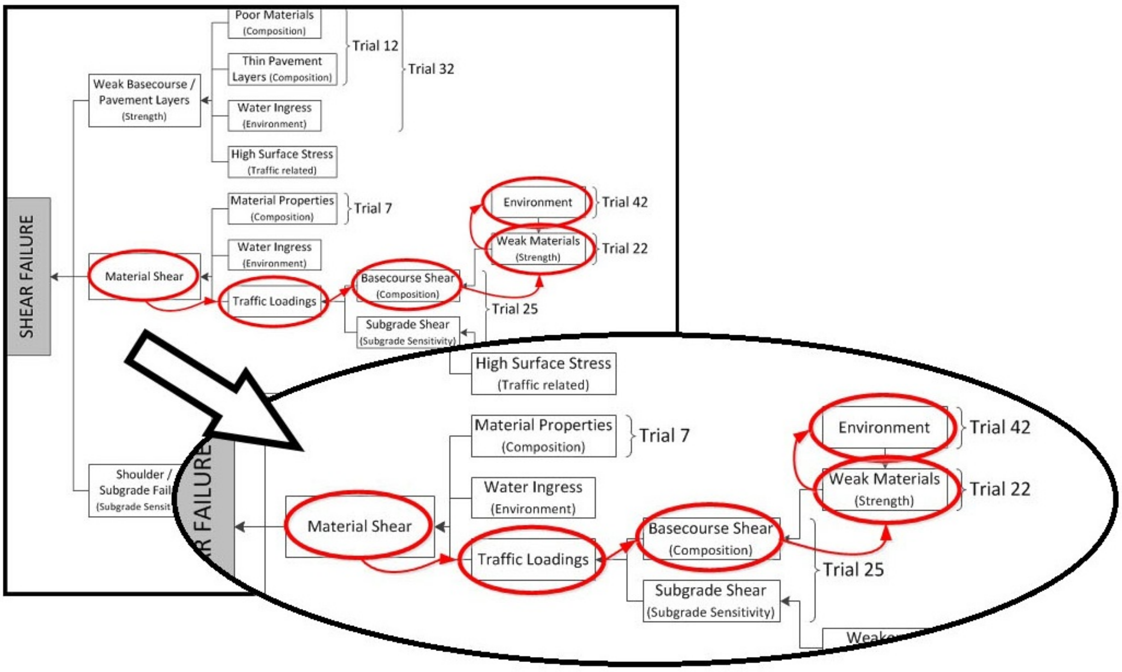


Figure 7-2: Critical failure path for site # 4962

Failure path # 42 includes the environmental factor group indicating this shear failure can be attributed to water ingress in the structural layers of the pavement, which in turn weakens the materials. Under repetitive traffic loads, the pavement results in material shear of the basecourse layer. Cracks in the surface layer or a sub-terrain drainage problem are the likely causes of the water ingress in this pavement. Therefore, the appropriate maintenance treatment would be to seal the cracks with a resealing exercise and / or remedy the drainage problem, the latter being a more extensive and expensive task.

Furthermore, it is suggested to investigate the possible causes of rutting (relating to the failure path # 49) to prevent the occurrence of the predicted secondary failure. Based on Figure 6-2, secondary rutting is likely to occur due to a weakened subgrade as a result of water ingress, which coincides with the above problem resulting in shear failure. Therefore, addressing the ingress of water into this pavement structure will likely prevent both primary and secondary failures.

7.3 How does the Rural Local Authority LTPP Sections Perform?

At a network level, the developed prototype system can assess the performance of the road network based on the following outputs:

- The distribution of the failure probabilities, indicating the likelihood of failure profile of the network;
- The proportion of the network susceptible to failure and the distribution of the failure modes of the predicted failed sections, and
- The distribution of the most probable causes of failure (critical failure paths) of the predicted failed road sections, identifying any symptomatic network-wide problems.

7.3.1 Profile of Likelihood of Failure

Figure 7-3 presents the distribution of the overall failure probabilities, based on the conservative (maximum probability) approach (refer Section 6.5.2.1). Overall, the distribution of the predicted probabilities is similar to that of the actual failure distribution (see Figure 7-3) where the peaks of the distribution curve occur in similar probability ranges of the failure spectrum, despite the peak at the lower end of the probability scale having shifted slightly to the right. Approximately one-third of the rural sites included in the rural Local Authority LTPP dataset are predicted to fail, given sites with a predicted probability of overall failure greater than 0.5. However, shifting the failure threshold to a higher value (0.8) resulted in approximately 20 % of the sites being predicted to fail. Although the predicted failure curve has shifted to the right of the actual failure density curve, the predicted failure distribution presents a large proportion of road sections with low probability of failure (probability less than 0.2) and a secondary peak with a higher failure probability (greater than 0.95).

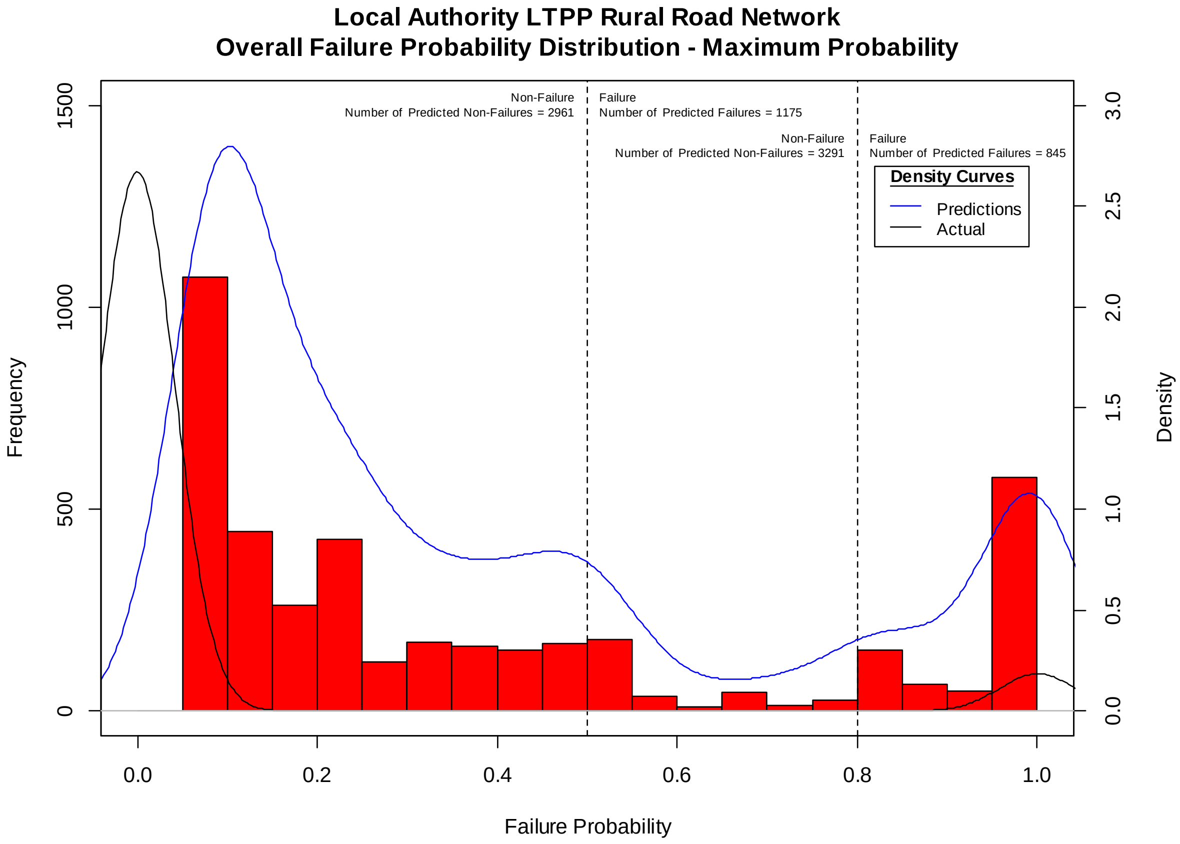


Figure 7-3: Failure profile of the rural sites of the Local Authority LTPP dataset, using the conservative (maximum probability) approach

The second approach to calculating the overall failure probability, using additive probability theory (refer Section 6.5.2.2), also assessed the performance of the rural Local Authority LTPP dataset. Again, Figure 7-4 reported similarities between the predicted and actual failure distributions; however, the peak centred around 0.15 is noticeably greater than that previously reported in Figure 7-3. With this approach, a higher percentage (36 %) of the road sections are predicted to fail (see Figure 7-4), reducing to 22 % when the failure threshold is increased to 0.8.

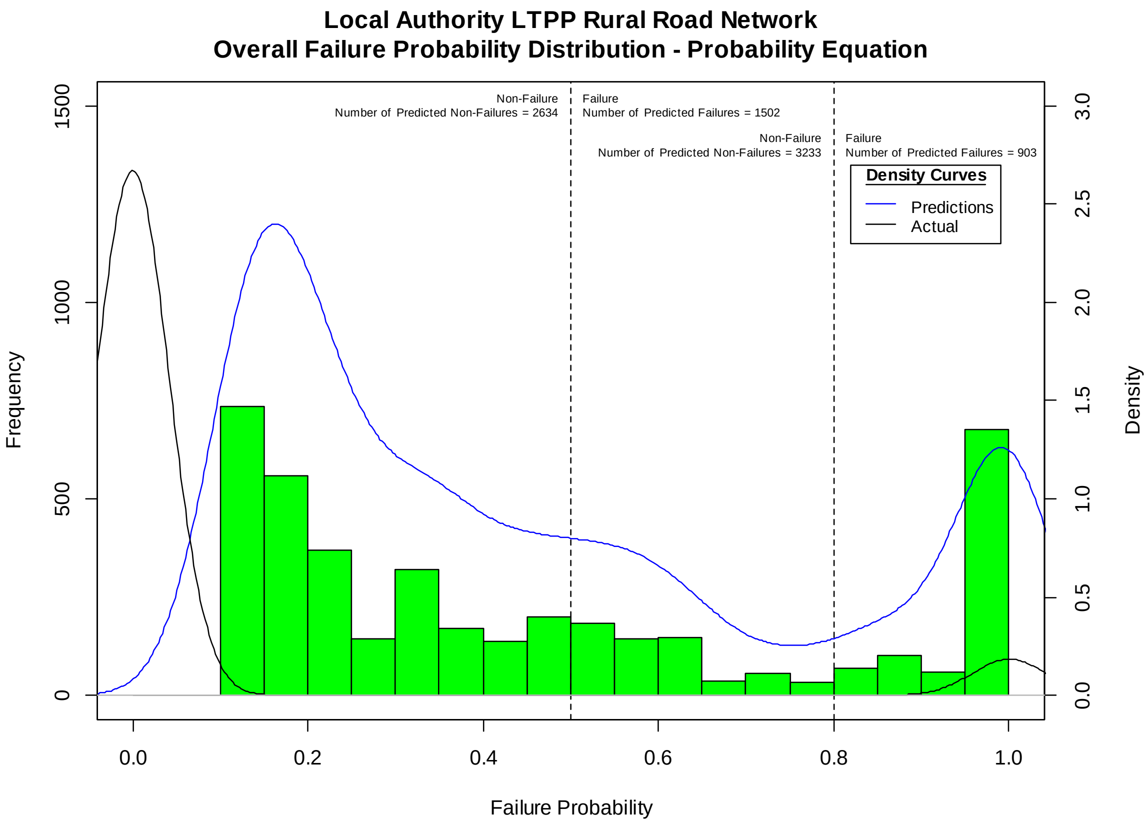


Figure 7-4: Failure profile of the rural sites of the Local Authority LTPP dataset, using the probability theory (probability equation) approach

Although the differences and shifts between Figures 7-3 and 7-4 are a result of two methods of calculating the overall failure probability, the figures provide an example of identifying the changes in the failure profiles. For example, given a change in the road network's environment, such as an increase in traffic loadings coupled with environmental changes and / or increased rainfall, the data for the modified (increased traffic) scenario can be used in the prototype system. The predicted probabilities for both the original and modified scenario can then be analysed visually to attain the shifts in the distribution profiles, similar to the above Figures 7-3 and 7-4, providing valuable information to the asset manager. An application of this approach would be useful in cases for example where a re-routing exercise increases the traffic loadings on a less strategic road, which typically had not been designed to carry the increased traffic loading levels. Schlotjes and Henning (2012) reported on this application of the developed prototype system using a typical road network in New Zealand, which recently experienced an increase in traffic loading.

7.3.2 Pavement Susceptibility to Failure Modes

The most probable failure modes predicted by the prototype system were used to determine the susceptibility of the rural Local Authority LTPP road pavements to specific failure types. The two approaches, to calculate the overall failure probability described in Chapter Six, were used to determine the susceptibility of the network to failure types comparatively, as shown in Figure 7-5. From the distribution of failure modes associated with the predicted failed road sections, the majority of the network is predicted to fail in rutting, as expected. However, fewer sites are predicted to fail in rutting using the probability theory (probability equation) approach (refer Section 6.5.2.2) and a different failure mode distribution is presented.

When the failure threshold is increased from 0.5 to 0.8 for both approaches, the proportion of rutting failures also increased, suggesting a susceptibility of this network to rutting. The fatigue cracking proportions remain unchanged, yet the shear failures at the higher end of failure probability distribution decreased, in some cases by half. The apparent difference between the two approaches is not evident in Figure 7-5.

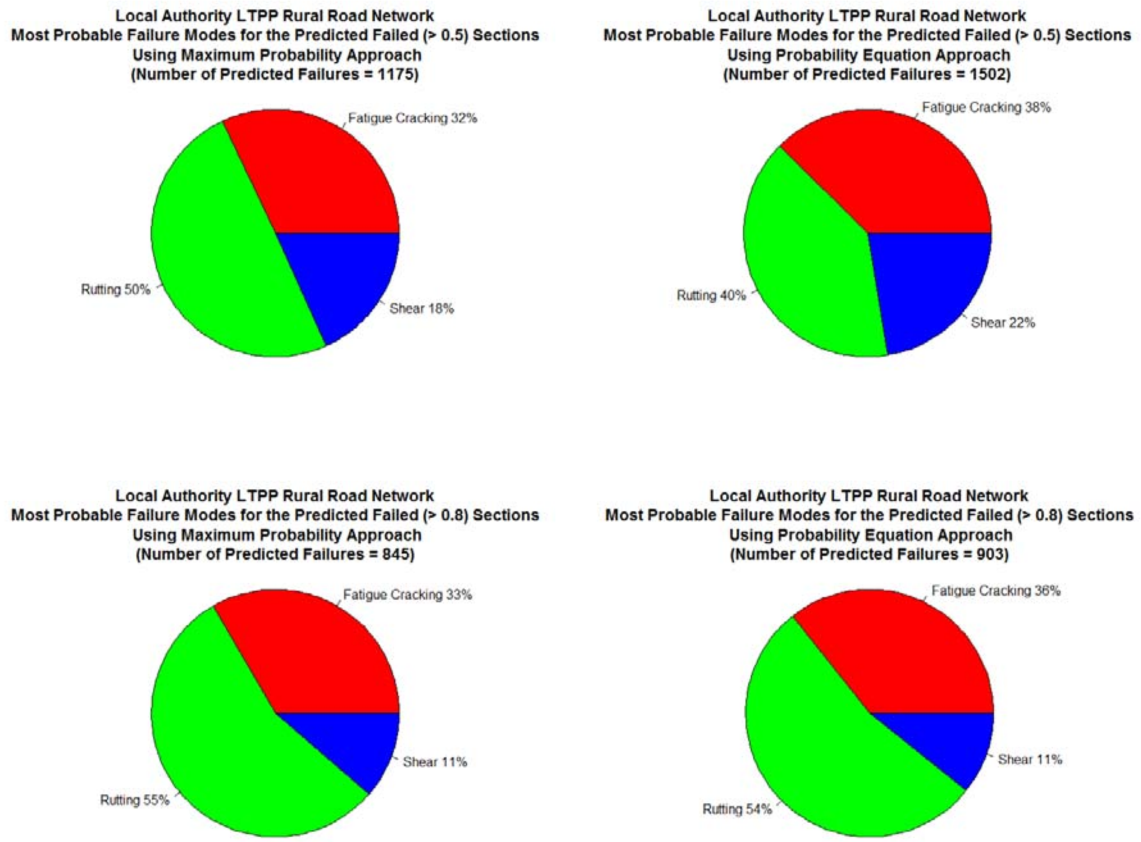


Figure 7-5: Predicted failure modes of the rural sites of the Local Authority LTPP dataset

7.3.3 Symptomatic Problems of the Network

Analysing the predicted critical failure paths identifies the symptomatic problems across the network, usually as a result of site-specific environments or maintenance practices. The rural Local Authority LTPP dataset includes a range of environments. To demonstrate this application, Figures 7-6, 7-7, and 7-8 present the frequencies of occurrence of the critical failure paths for the pavements included in the dataset.

Figure 7-6 shows little difference between the two approaches used to determine the probability of failure, for the predicted failure paths of rutting failures. The predominant causes of network-wide rutting were traffic loadings, pavement composition and pavement strength (failure path # 22). From Figure 7-8, weaknesses in the pavement composition coupled with traffic loadings (failure path # 7) caused the majority (44 % and 34 %) of network-wide fatigue cracking, given the respective approaches of calculating the overall failure. However, two distinct failure paths were identified in Figure 7-7; therefore, the predominant causes of shear include traffic loadings, pavement composition, and pavement strength (failure path # 22), and the above (failure path # 22) in conjunction with subgrade sensitivity (failure path # 44).

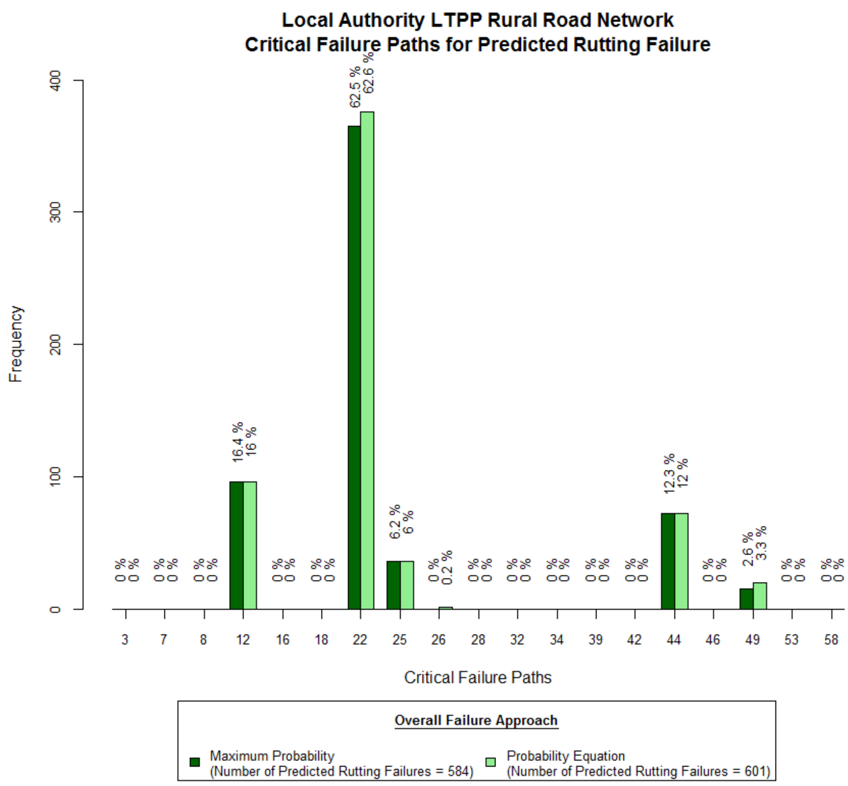


Figure 7-6: Critical failure paths for predicted rutting failures

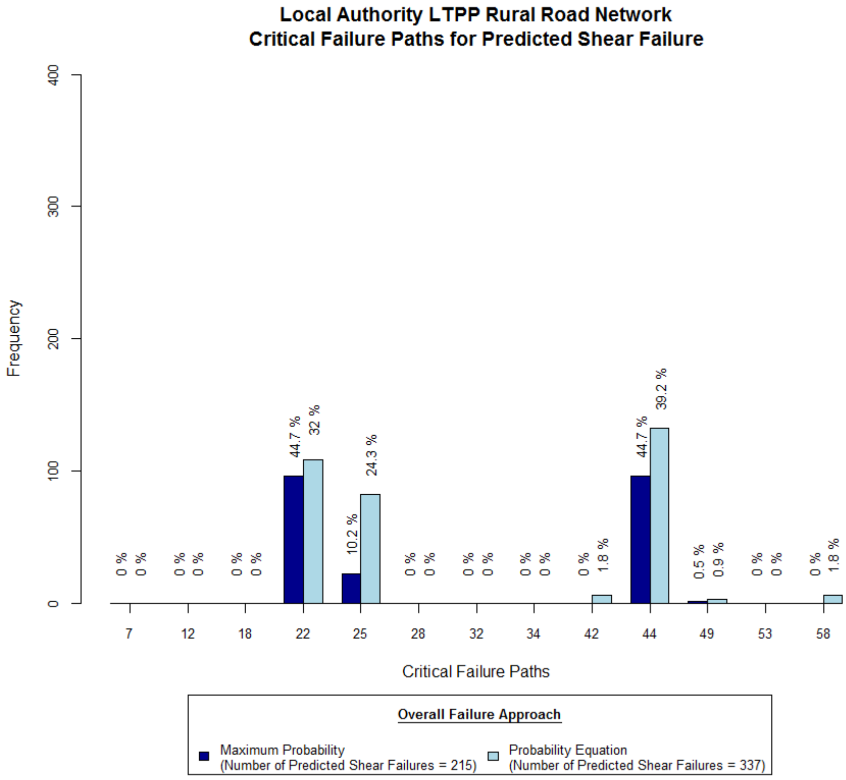


Figure 7-7: Critical failure paths for predicted shear failures

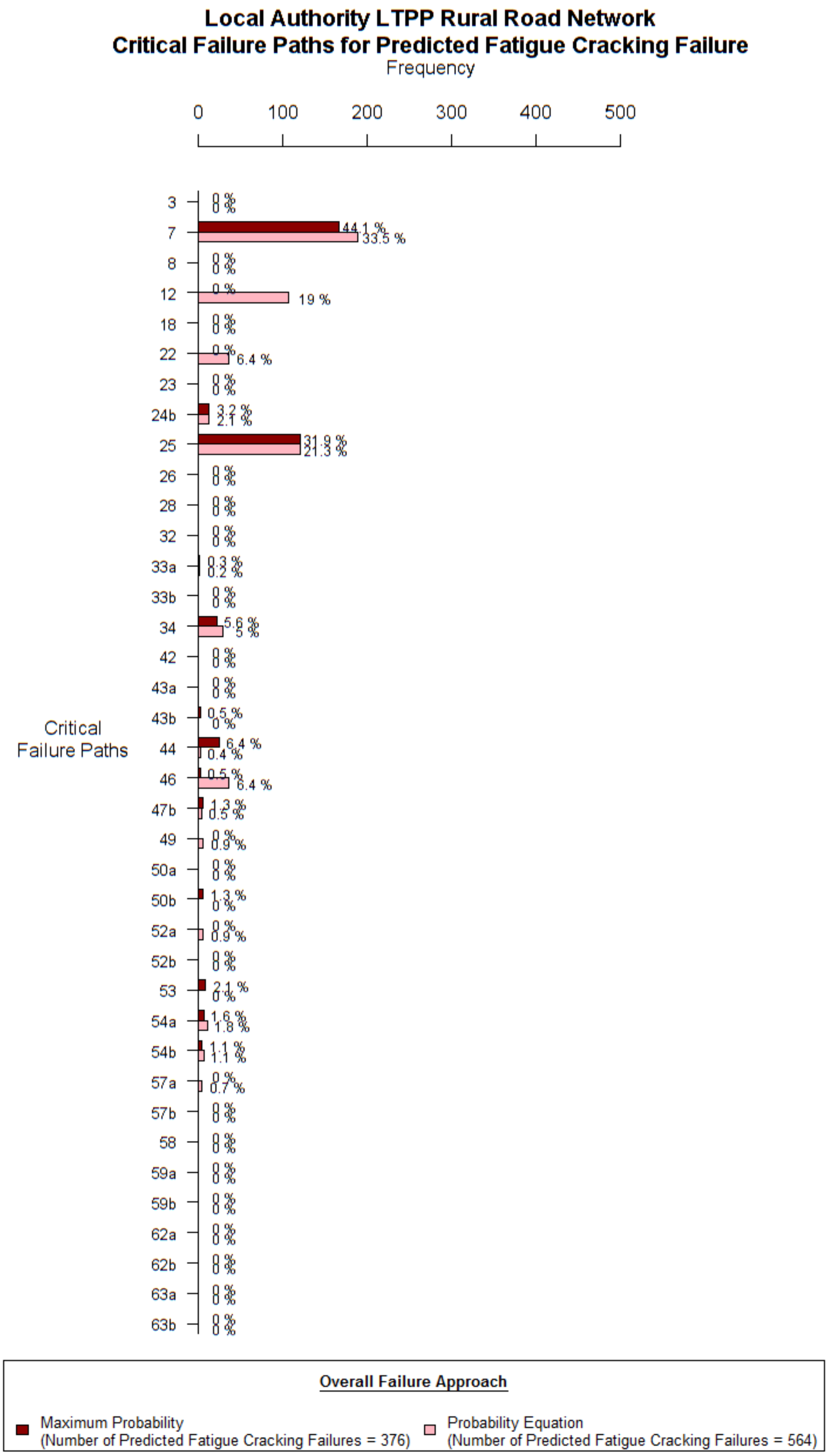


Figure 7-8: Critical failure paths for predicted fatigue cracking failure

7.4 Summary of the Case Study

This chapter analysed the rural Local Authority LTPP dataset to demonstrate the practical applications, both at project level and network level, of the developed prototype system and evaluate the performance of the system on an independent dataset. The system performed adequately on the independent testing dataset; however, the performance of the rutting system in comparison to Chapter Six was inferior. This may be attributable to the system not recognising or including all methods of rutting failure, which were present in the Local Authority LTPP dataset (testing dataset), and suggests the need to refine the rutting system further.

From the network dataset, 12 sites were selected at random to demonstrate the use of the system. For each, the probability of failure was reported and the most probable causes of failure were identified. This information can assist the asset manager in maintenance decisions.

At a network level, Figures 7-3 and 7-4 showed the distributions of the predicted failure probabilities followed that of the actual failures presented in the rural Local Authority LTPP dataset and a large portion of the network were not predicted to fail ($P(X < 0.5)$). The primary failure mode of the rural Local Authority LTPP road pavements was rutting. Such information can be used as a measure of the overall performance of the network.

Overall, the developed prototype system can be employed in the management of road networks to:

- Determine the overall failure probability of a road pavement section;
- Identify the most probable causes of failure to streamline maintenance;

- Identify the symptomatic problems across the network, and
- Evaluate the vulnerability of networks to the effect of changes in the environment, such as an increase in traffic or the predicted effects of climate change.

The results of the prototype system may be regarded as promising for the potential of an effective non-destructive diagnostic tool. However, further testing of the developed failure system is both desired and necessary before it can be considered to be fully developed. This should include further testing of the performance of the system on network data. The following chapter summarises the reported research and reviews the methodology and approaches taken to the research.

Chapter Eight

SUMMARY AND DISCUSSION
A Review of the Research

8.1 Introduction

The development of a prototype system that predicts the probability and diagnoses road pavement failure is a novel approach to supplement the current PMS and road reporting processes. Current practices fail to present a robust methodology for predicting the failure probability of road pavements for further use in network assessments of failure risk. To address this, the framework developed in this research provides a means to predict the probability of failure of road sections and to identify the most probable causes of failure.

The purpose of this chapter is to provide a critical review of the methodology followed in this research. The research objectives outlined in Chapter One were accomplished by:

- Developing a comprehensive knowledge of flexible road pavement failure and identifying the factors contributing to failure;
- Assessing the performance of a number of classification methods for road failure prediction using the State Highway LTPP road network;
- Developing a prototype system based on incorporating the engineering (failure) knowledge with the computational model, and

- Assessing the effectiveness of the developed prototype system and methodology using the Local Authority LTPP road network.

8.2 Defining the Research Task

The research aim, defined in Chapter One, was to predict the probability of road pavement failure. In the LTPP datasets, failure was represented with a binary class label, such that a road section was defined either as a failed site or a sound site; thus, mathematically defining the research task as a binary classification problem, focussing on the end of life state of the pavement as opposed to the gradual deterioration of pavements over time. Although other modelling approaches have been successful in predicting pavement performance, namely deterioration models, the definition of the research task limited the modelling approaches considered in this research to classification techniques. Chapter Two discussed the performance of several classification techniques, which were specifically evaluated in Chapter Five using the research dataset.

8.3 Summary of the Research

This research developed a generic framework to predict the probability of road pavement failure by inferring engineering knowledge into computational models.

Based on the findings in the literature review (Chapter Two), a methodology was proposed to develop a classification model, selected through the comparative study, which quantified the probability of road pavement failure. Through the analysis of the system, it was concluded that the methodology followed in this research successfully inferred engineering knowledge into the design of a pavement performance model.

To achieve this, the objectives of this research were to:

1. Develop charts for rutting, cracking, and shear failures that supplements the current knowledge base and identifies the factors contributing to failure and interactions between these causes;
2. Comparatively assess the performance of suitable classification techniques, using the State Highway LTPP dataset, based on a developed selection criteria process and a review of suitable classification techniques. This study aimed to evaluate the performance and effectiveness of the classification techniques, such that one technique could be further developed in the remaining research;
3. Develop a prototype system to predict the probability of road pavement failure while inferring engineering knowledge into the computational system. The predicted outputs from the system included a probability of road pavement failure for each failure mechanism, a predicted overall failure probability, and a suggested prediction of the most probable causes of failure, and
4. Evaluate the developed prototype system with an independent road dataset to demonstrate the effectiveness of the system and methodology.

Failure charts illustrating the causes of rutting, cracking, and shear failures were developed.

Chapter Four reviewed the fundamental knowledge of each failure mechanism, canvassed expert opinion on road pavement failure, and explored potential relationships in two New Zealand road datasets. The developed failure charts captured this engineering knowledge of pavement failure by presenting the possible causes and illustrated the failure paths associated with each failure mechanism. Five failure factor groups were inferred from the developed failure charts for further use in modelling processes, as described in Chapter Five. The charts were revisited in Chapter Seven and used to identify the most probable cause(s) of failure.

In addition to the application demonstrated in this thesis, the failure charts can provide asset managers, civil engineers, and RCAs a comprehensive background to the causes of rutting, cracking, and shear failure in the field.

A comparative study evaluated the performance and effectiveness of five classification techniques.

The reviewed literature revealed several classification techniques that performed well in the respective study domains. Five of these were selected to assess their suitability for the use with road pavement data using a number of criteria, including the advantages and limitations of each approach. Four point performance measures evaluated the performance of each modelling technique. Hypothesis testing was employed to determine if the difference between the results of the performance measures, for each modelling technique, was statistically significant or not, and to conclude whether one modelling technique outperformed the others. Based on a further set of criteria including model performance, user interpretability, and computer efficiency, the performance and effectiveness of the five classification techniques were evaluated. This demonstrated that the support vector machines technique was most suitable for the task at hand and accordingly it was selected to develop further into a prototype system.

Through this research, a prototype system was developed successfully integrating engineering knowledge into the support vector machine classification models.

Chapter Six demonstrated a methodology to infer engineering knowledge captured in the failure charts into computational models. Each failure mechanism was modelled individually and the resulting failure probability for each mode of failure was assumed to be the maximum probability across all failure paths modelled. Three approaches were considered to calculate the overall failure probability; however, the analysis failed to conclude on one superior approach. Only two approaches were pertinent to this research, namely the conservative (maximum probability) approach (refer Section 6.5.2.1) and an approach based on probability theory (refer Section 6.5.2.2). The proposed computational modelling approach (refer Section 6.5.2.3) was thought to be the best calculation of the overall failure probability; however, given the few occurrences of multiple failures in the research dataset, the implementation of this approach was excluded in this research.

The effectiveness of the prototype system was demonstrated using the Local Authority LTPP dataset.

Chapter Seven presented the results from the system testing on an independent New Zealand road network. The accuracy of the system was found to be sufficient and demonstrated the transferability of the system to a new dataset. The chapter further discussed the practical project and network level applications of the system including:

- Predicting the probability of road pavement failure for both overall and the individual failure modes;
- Identifying the critical failure path and causes of failure;

- Identifying any symptomatic problems on the network, and
- Assessing the impact of changes in the road environment and evaluating the susceptibility of the network to failure types.

8.4 Critical Review of the Research

The methodology adopted for the development of a system, which quantifies the probability of pavement failure, focussed on the end of life probability state of the pavement as opposed to the life-cycle performance of the pavement. In hindsight, alternative approaches to the methodologies adopted in this research could have been followed, which are explored and discussed below under the following headings:

1. Failure knowledge;
2. Comparative study of classification techniques;
3. Development of the prototype system;
4. Practical application of the system, and
5. Data issues.

8.4.1 Failure Knowledge

The developed failure charts made use of expert opinion, the literature, and knowledge obtained from scrutinising road datasets in presenting the causes of rutting, cracking, and shear failures of flexible road pavements. Adopting the robust methodology resulted in a comprehensive understanding of road pavement failure obtained from the three sources of knowledge. However, the causes included in these charts were site-specific to New Zealand road environments; thus, for other road environments, amendments to the failure charts may be required to include additional or alternative causes of failure.

Inferring such engineering knowledge into the computational model identified the limitation of assuming independence between the failure mechanisms in the development of the failure charts, such that the support vector machines computational models developed for each failure mechanism did not recognise the interactions between rutting, fatigue cracking, and shear failures in the probability calculations. Rather, this aspect was addressed by using probability theory (refer Chapter Six). To address this aspect, it is recommended that any future work, which further refines the system developed herein, should focus on further developing the failure charts so that they are able to depict possible interactions between failure mechanisms.

8.4.2 Comparative Study of Classification Techniques

Chapter Five presented a comparative study of five classification techniques using road pavement data, as the performance of a classifier is highly dependent on the particulars of the research dataset, thus the direct transfer of a single algorithm to another research domain is uncertain (Dreiseitl and Ohno-Machado, 2002; Perlich et al., 2003). Other comparative studies adopted techniques based on the popularity of the technique in the respective study fields. This research developed objective-based criteria to select the techniques most suitable for the research dataset, based on the performance of the techniques, user interpretability, and computational inefficiencies faced in the practical implementation of such approaches. However, the exclusion of other techniques at this stage of the research was reliant on the published results from other studies and data.

A number of assumptions were adopted for the comparative study (Chapter Five) as described below:

1. Replacing missing data values with a nominal value;

2. Straight-line transformation to normalise the research dataset;
3. Applying a weighting factor to account for unequal portions of failures versus non-failures;
4. Employing a 10-fold cross-validation sampling method, and
5. Assuming a uniform cost impact (consequence) of incorrect predictions, such that this research assumed it was equally undesirable for both incorrect predicted failures and non-failures.

8.4.2.1 Missing Data Values

It is inevitable for road network data to include missing values, as a result of financial, equipment, and time restrictions imposed on the RCAs, such that inventory data in the network databases is limited. For the State Highway LTPP dataset, missing values were assigned a nominal value of zero (0) (refer Section 5.1.2.1), opposed to omitting such road sections which would ultimately reduce the available data significantly. In many of the input fields, this assumption was considered acceptable where it resulted in the worst possible scenario of pavement composition and strength. For example, much of the State Highway LTPP dataset failed to provide evidence of a sub-base layer and, in these cases, the thickness and age of this layer was assigned a zero (0), indicating a thinner pavement. With this approach, the predictions will be based off a pavement more susceptible to failure; this could in turn distort the predictions of the new observations. For example, the model may assume a thin pavement is sound, based on the data, but in fact it is a thicker pavement with an unknown thickness of one of the layers that is sound. To address this, the data needs to be verified to ensure completeness and accuracy of the composition information. Furthermore, assuming a zero (0) traffic loading represents the best possible scenario suggesting the pavement's deterioration is wholly attributed to environmental loading. Perlich et al. (2003) suggested adjusting the missing values using the mean of the input variables. For the State Highway LTPP dataset, this approach would distort the representative sample of the network, indicating the pavements were thicker and stronger than in reality, although for the environment, traffic loadings, and condition data, the averages would represent the pavement conditions realistically.

By adopting this assumption, the majority of the network assumed the worst possible scenario, as the State Highway LTPP datasets included complete records of traffic loadings and condition data. In the case of rainfall, incomplete data was cross-referenced to external sources for completion. However, for road pavements, adopting this approach is recognised to be misrepresentative of the data, and it is suggested to either adopt a mean nominal value per input variable or a known default value appropriate for the respective network. Further testing of the prototype system employed the latter to network datasets (Deng and Henning, 2012), who reported sufficient results from the road networks applicable to the scope of this research.

8.4.2.2 Straight-Line Transformation

Normalisation of the research dataset eliminates any preference from the model towards one input variable over another, such that the amount of data considered by the model is maximised. The performance of classifiers is improved when the input values range between zero (0) and one (1) (Roiger and Geatz, 2003), typically achieved by the following normalisation processes:

- **Decimal scaling:** Where scaling of the variables by a nominal power of 10;
- **Min-max normalisation:** Variable scaling using the maximum and minimum values as follows in Equation 8-1;

Equation 8-1

$$new\ value = \frac{original\ value - minimum\ value}{maximum\ value - minimum\ value}$$

- **Normalisation using straight-line transformation:** Where statistical parameters of the variable, such as the mean and standard deviation, are used to compute the new value, following Equation 8-2, and

Equation 8-2

$$new\ value = \frac{original\ value - mean}{standard\ deviation}$$

- **Logarithmic normalisation:** Replacing the variable values with their logarithms.

Although road pavement data seldom follows a normal distribution, the straight-line normalisation methodology adopted consistently across all independent variables was assumed sufficient for this study, given the properties of the research dataset. As discussed in Section 5.1.2.1, the uncertainty of the pavement behaviour and variability in road network data distorts the distributions of the variables (Reigle, 2000), such that the distributions of the variables are heavily dependent on the contents of the research dataset(s). For future adoption of the framework, this research suggests investigating the distributions of the possible model inputs to understand the nature of the variables and, subsequently, establish the appropriate normalisation method applied consistently across all inputs.

8.4.2.3 Removing Failure Biasness from the Dataset

Classifiers perform best with a balanced dataset that has equal representation of classes, in order for the model to assume the equal possibility of predicting a class for the new observations. With unbalanced data classes, the trained model will always attempt to minimise the error, thus the majority of the new predictions will belong to the larger of the two classes (Prinzie and Van den Poel, 2008). Failure on road pavements is avoided with the implementation of proactive maintenance and, therefore, the reality of road pavement data results in an unbalanced dataset. To offset this imbalance, and subsequently remove any biasness from the predictions of the trained model, a weighting factor was applied to the data. Following Equation 8-3, the weighting factor of a datapoint is based on the ratio of non-failures to failures across the dataset:

Equation 8-3

$$w_i = \left[\frac{n - \sum_{i=1}^{n}(x_i)}{\sum_{i=1}^{n}(x_i)} - 1\right] \times x_i + 1$$

$$where{:}\, w_i = weighting\ factor\ for\ datapoint$$

$$x_i = binary\ failure\ state\ (0\ or\ 1)$$

$$n = number\ of\ datapoints$$

8.4.2.4 Cross-Validation Approach

It is good practice to reserve a portion of the entire dataset for testing (Gil and Johnnson, 2011), if possible, although researchers have assorted viewpoints on the size of the evaluation datasets required for binary classifiers (see Table 8-1).

Table 8-1: Size of Evaluation Datasets from the Literature

Study	Percentage of dataset reserved for testing	Sampling Method (if applicable)
Austin et al. (2010)	-	1000 bootstrapping
Bhattacharya and Solomatine (2006)	35 - 39	-
Caruana and Niculescu-Mizil (2006)	47 - 88	-
Chandra et al. (2009)	10	10-fold cross-validation
Eftekhar et al. (2005)	33	-
Gil and Johnnson (2011)	20	5-fold cross-validation
Jagielska et al. (1999)	33	-
Kaseko and Ritchie (1993)	48	-
Pal (2006)	31 - 32	-
Perlich et al. (2003)	25 - 33	-
Prinzie and Van den Poel (2008)	50	-
Saghafi et al. (2009)	20	-
Tso et al. (1998)	33	-
Verikas et al. (2011)	10	10-fold cross-validation
Wu et al. (2002)	17	-

- Not applicable

The datasets for the majority of the studies listed in Table 8-1 were sufficient in size to reserve an adequately sized evaluation dataset to assess the performance of the models. In some instances, the evaluation dataset was in excess of 30,000 samples (Caruana and Niculescu-Mizil, 2006; Prinzie and Van den Poel, 2008). The studies with smaller testing sets employed sampling methods, such as bootstrapping or cross-validation, due to the insufficient amount of data available for the study. Alternative sampling techniques are available (Raudys, 2001; Roiger and Geatz, 2003) if the ideal volume of data is not available:

- **N-fold cross-validation:** The dataset is divided into *n* random subsamples. One subsample is reserved for testing, and the remaining dataset (*n - 1* subsamples) is used in training the model. The training of the model and testing is repeated *n* times, once on each test subsample, and the results are averaged over the number of repetitions;

- **Bootstrapping:** The training set is sampled with replacement from the complete dataset. Mathematically, bootstrapping a dataset containing *n* instances *n* times will result in a training dataset of two-third *n* instances and a testing dataset of one-third *n* instances, and
- **Leave-one-out cross-validation:** In extreme cases, only one datapoint is used for testing and the remainder of the dataset is used for training. This method will be repeated *n* times, where *n* = number of datapoints in the complete dataset.

The State Highway LTPP research dataset did not contain sufficient datapoints for separate training and evaluation datasets, unlike the studies in alternative research domains identified in Chapter Two and Table 8-1 above. The need for training computational models on small datasets is a requirement for road pavement data. However, this research dataset was of adequate size to disregard the leave-one-out cross-validation method. Although Dreiseitl and Ohno-Machado (2002) reported the superiority of bootstrapping over cross-validation, this method is often overly optimistic despite its simplicity. Therefore, a 10-fold cross-validation sampling method was adopted in this research to ensure validity of the results of the performance assessment of the techniques. To achieve this, the entire dataset was randomly divided into 10 subsamples, with each 10 % subsample reserved once for testing the model that had been trained on the remaining 90 % of the data. However, extensive run time was required to repeat the training and testing of the model, resulting in *n*-times the single computation time to evaluate the method. The randomness of the sampling method did not guarantee the same testing dataset was used for each technique; however, the outcome of the comparative study was not negatively influenced by this given the method of hypothesis testing employed.

8.4.2.5 Assuming Equal Cost of Incorrect Predictions

Point and integrated performance measures were identified in the literature review to evaluate the performance of the classification techniques, in conjunction with the model accuracy. Since accuracy is greatly influenced by the class distribution of the dataset, sole reliance on this performance measure results in a bias assessment of the technique's predictive power (Ben-David, 2008; Dreiseitl and Ohno-Machado, 2002; Parker, 2011). Thus, alternative performance measures were included in the comparative study, specifically point measures with an assumed failure threshold of 0.5 (i.e. $P(X \geq 0.5) = 1$).

Integrated measures have provided superior assessments of the performance of modelling techniques (Parker, 2011), as these calculate the predictive power of the model over all possible thresholds. To report on such measures, such as the popular area under the receiver operating characteristic curve (e.g. integration functions), the user requires a cost function (Dreiseitl and Ohno-Machado, 2002). This research assumed the cost to be uniform between the predictions that incorrectly predicted failed and sound pavements, such that it is equally undesirable for false negatives and false positives; therefore, calculating such measures measure would add no additional value to the results. The consequence of failure varies depending on the required maintenance treatment. For example, the costs, both financial and social, of a full rehabilitation are much more extensive than a resealing exercise. Without this information, the assumption adopted in this research was considered to be adequate, given the available data. For a more accurate representation of the performance of the classifier, collation of this information would be required to make well-educated judgments on incorrect predictions and included in an appropriate cost analysis of the model performance.

8.4.3 Prototype System Development

Chapter Six presented the development of the prototype system, which, by following the conceptual design of the research, successfully inferred engineering knowledge through the use of failure charts (developed in Chapter Four) into the computational model (selected in Chapter Five). The discussion of this methodology focuses on the following:

1. Experimental design for inferring human knowledge, and
2. Methods for calculating the overall failure probability.

8.4.3.1 Experimental Design and Human Knowledge

The conceptual design of this research focussed on inferring engineering knowledge into the design of the computational system, such that the development of the support vector machines models were informed by the developed failure charts. Modelling each successful factor combination individually, selected by 100 % accuracy (and 90 % in the case of shear failure), represented a possible failure path on the developed failure charts. Although these factor combinations correlated well with the failure charts, the use of computational outputs identifying the failure paths has the potential of excluding critical failure paths, especially if the factor combinations are not well represented in the data. This hinders the transferability of the developed prototype system to other road networks; however, the resulting failure paths included in the system were considered appropriate for the State Highway LTPP dataset.

As each failure path (factor combination) is modelled separately, the computational model excluded possible interactions between the failure paths, assuming the critical failure path was solely responsible for failure. The failure charts account for the interactions between the causes of failure, as shown per failure path; however, there was no provision included in the system for interactions between the failure paths of each failure mechanism (refer Section 8.4.1).

8.4.3.2 Overall Failure Probability

Chapter Six demonstrated the difficulties of representing real-life pavement failures with mathematical equations. The occurrence of two or more failure mechanisms was dealt with using two approaches, as follows:

1. Assumed the probability of overall failure did not increase with the occurrence of multiple failures (conservative approach), and
2. Assumed independence between the failure mechanisms, such that the overall failure probability accounted for multiple failures but secondary effects were not considered (probability theory approach).

The limitations of each approach were discussed in Section 6.5.2. Although these two approaches were considered sufficient given the available data, a third approach was proposed, which was proposed to determine the probabilities of multiple failures using computational models. To do so adequately, the research dataset would require a greater number of observations where multiple failure modes were apparent and detailed information regarding the timing of these failure(s), in order to differentiate between secondary effects and combined (or multiple) failures. Such data and information on these types of failure is seldom included in road network data, not to say such failures do not occur on the networks. In most cases, a secondary defect is often masked by the occurrence of primary failure, or so determined by the site inspection, and subsequently the maintenance fails to address the secondary failure causes. This research however used alternative approaches to compute the probability of multiple failures.

Notwithstanding, the expected differences between the two approaches adopted in this research were apparent in the results in Section 6.6; naturally greater probabilities were yielded by the probability theory (probability equation) approach, due to the addition of the individual probabilities, compared to the conservative (maximum probability) approach. The comparison of the two approaches failed to conclude a superior approach, based on the performance measures, although Figures 6-5, 6-6 and 6-7 showed a discernible difference in the probability distributions. Integrated performance measures would assess this difference sufficiently if the consequences of incorrect predictions were known and, therefore, provide a conclusive evaluation of the system.

8.4.4 Practical Applications of the System

Chapter Seven demonstrated the use of the developed prototype system and evaluated the effectiveness of the system given an independent testing dataset. The Local Authority LTPP dataset included identical parameters to the research dataset enabling it to be analysed by the prototype system. However, typical road network datasets lack the sophistication and detail included in the LTPP datasets and, therefore, it should not be assumed that the results in Chapter Seven are equivalent to a network test. Such differences in the datasets were evident in the preliminary data analysis of the Southland District Council road network and the LTPP networks (refer Section 4.4.3 and see Appendix A).

The applications demonstrated in Chapter Seven rely heavily on the predictions from the system. However, if data needed for the implementation of the prototype system is absent, as a result of discernible differences between data collection methods and collation of the research dataset and RCAs, it is expected that the performance of the prototype system would further decrease. Consequently, the information obtained from the analysis and manipulation of the raw probabilities will be ambiguous to the user. To address this, the system could be retrained on the available data, yet the performance of the new computational models would need to be evaluated.

8.4.5 Data Issues

The previous sections of this chapter have discussed the issues with the available research data, such as:

- Missing data values (refer Section 8.4.2.1);
- Unbalanced representation of failure data in the datasets (refer Section 8.4.2.3);
- Restriction of testing methods given the size of the research dataset (refer Section 8.4.2.4), and
- Restricted use of performance measures given unavailable information (refer Section 8.4.2.5).

Furthermore, this research chose to exclude various data items from the study given the categorical nature[17] of the information and / or the high level of incompleteness. The failure charts (Figures 4-7 to 4-11) developed in Chapter Four included causes of failure that, due to the lack of corresponding data, were unable to be included in the classification models. This included aggregate properties, surfacing material types, construction quality, horizontal gradient, and inadequate compaction. The LTPP datasets listed material types as categorical

[17] Data items listed as categories or classes, such as material types, as opposed to numerical data.

data with many possibilities, if the information was available. Given the subjective nature of construction quality, including compaction characteristics and asphalt contents, collating information on such data items poses difficult with current data collection methods and therefore was not available for this research. As a result, the computational models ignored any irrelevant, unavailable, or categorical data. This research suggests the development of a robust methodology for the collection of subjective causes of failure for further consideration in the pavement performance models.

As stated previously (refer Section 8.4.4), the differences between the research dataset and network data are discernible. Chapter Four identified the failure factors involved in road pavement failure. Further to this, Chapter Six concluded on the successful factor combinations from the comparative study presented in Chapter Five. As Table 8-2 shows, this research concluded strength, composition, and traffic factors were the influential in predicting the probability of each pavement failure mechanism. These factors should be collected robustly from road networks, for the success of the prototype system in road network management, and to be considered in future pavement performance studies.

Table 8-2: Summary of the Factor Combinations included in the Prototype System

Failure Factors	Rutting	Fatigue Cracking	Shear
Traffic	11	19	8
Composition	11	22	10
Strength	16	23	11
Environment	9	12	5
Subgrade Sensitivity	10	14	8
Surface Condition	0	11	0

8.5 Overall Performance of the System

The performance of the developed prototype system for analysing the failure probabilities of the State Highway LTPP road network was discussed in Chapter Six. As reported, the performance of the shear support vector machines system was surpassed by the other two failure mechanisms. This was a direct result from excluding the information about material compositions from the research dataset. Further to the above, such information is required for the performance of the prototype system to improve.

Despite the reported success of the support vector machines technique in the comparative study, the weak results from the phi coefficient suggested little correlation between the input variables and predicted outputs. The inferred knowledge from the failure paths, although successful for rutting and fatigue cracking, was not as successful in the shear failure system. However, given the other reported performance measures, the developed system could be considered to have performed adequately overall, with respect to predicting rutting and fatigue cracking.

8.6 Value of the Research

The value of the research was demonstrated by:

- Developing comprehensive failure charts;
- A comparative study of the performance of classification techniques given road pavement data;
- Inferring engineering knowledge into computational models, and
- Predicting the probability of road pavement failure with support vector machines.

Failure Charts:

Representing knowledge of road pavement failure in the form of failure charts was not evident in the literature review. By presenting the causes of failure on failure charts, the causes, including the combinations of causes, are easily identified for future reference and added a diagnostic element to the computational model. This research based the development of the failure charts for rutting, cracking, and shear road failure mechanisms on approaches previously employed for other infrastructure assets. This work has been submitted to the Australian Road Research Board *Road and Transport Research* journal for publication (Schlotjes et al., 2012b).

Comparative Study:

The literature identified one study (Chandra et al., 2009) that had compared the classification techniques considered for the prototype system in Chapter Five; however, the focus of this study was outside of the transportation sector. Comparative studies of classification methods were not evident in the literature within the transportation sector and, more specifically, for predicting the end of life probability of an infrastructure asset. Methodologies from the reviewed studies using classification techniques acknowledged a number of considerations for the future use of classifiers. The performance of the suitable classification techniques was reported in Chapter Five and this research concluded the appropriateness of classifiers when using pavement data.

Inference of Engineering Knowledge:

The preceding two elements of the design were used as the foundation of the prototype system. Although literature on mechanistic-empirical models detailed a similar conceptual design, the inference of engineering knowledge in this research, such that the development of the computational model is informed by the failure charts, is a new approach to modelling pavement performance and the probability of pavement failure. In doing so, this research recognised data factors that played an important role in the prediction of pavement failure, which further identified the importance of strength, composition, and traffic data in predicting pavement performance and the future management of road networks.

Support Vector Machines:

Using support vector machines to predict the probability of road pavement failure is a novel approach in the transportation sector. The majority of the literature on pavement performance modelling focussed on pavement deterioration modelling opposed to the probability of the end of asset life failure. Schlotjes et al. (2013a) presented the appropriateness of using support vector machines in the transportation industry and the work was further submitted to the Institution of Civil Engineers *Transport* journal (Schlotjes et al., 2013b).

8.7 Summary of the Discussion

This chapter has critically reviewed the research methodology adopted in the development of a prototype system for predicting the probability of road pavement failure. In particular, the effectiveness of the system and assumptions made throughout the research were discussed and, where appropriate, suggestions have been offered to facilitate future developmental work and improvements to the system.

The inference of engineering knowledge into the computational model was achieved through the use of failure charts. Although the overall system performed well, the assumptions made in this research affected the ability of shear failures to be predicted with as great an accuracy to the other two modes of failure considered. The developed failure charts informed the development of the computational models, such that there was a parallel association between the failure charts and model. However, the exclusion of categorical and unattainable data from the modelling process resulted in a number of the possible failure paths having to be disregarded.

Furthermore, suggestions for further development of the system included evaluating the interactions between the failure mechanisms with the use of computational techniques on sufficient road pavement data, evaluating the modelling technique with integrated performance measures, and improvements to the current data collection processes for such tasks.

Further testing on typical road network datasets is recommended. The study concluded that strength, composition, and traffic failure factors are essential in obtaining sufficient predictions with the prototype system. Therefore, for the implementation of the developed prototype system to other road network datasets, it is imperative that RCAs employ robust data collection and collation processes in order to attain the input variables required for the system.

The value of this research, focussing on predicting the end of life probability for road pavements, was discussed with reference to the literature. As studies in this topic area were not evident, the use of support vector machines in predicting the probability of pavements is a new approach to pavement performance modelling.

Conclusions from the research together with recommendations for future work are presented in the following chapter.

Chapter Nine

CONCLUSIONS AND RECOMMENDATIONS

9.1 Accomplished Work and Main Findings

As discussed in the previous chapter, this research has demonstrated the objectives outlined in Chapter One by:

1. Developing failure charts for rutting, cracking, and shear road pavement failure based on a thorough understanding of the mechanisms associated with each failure type. These failure charts enable this knowledge to be utilised in the computational models developed to predict the likelihood of road pavement failure;
2. Evaluating the performance of classification techniques using the research dataset to establish the most appropriate technique for this research task;
3. Developing a prototype system based on the outcomes from the above two objectives, which quantifies the probability of road pavement failure, and
4. Assessing the effectiveness of the system on an independent dataset.

The research described above concluded:

- A number of computational methods are available to model pavement performance. However, **Chapter Two** reported that investigating the applicability of each technique is imperative to the successful performance of such methods in practice. To do so, the performance of each technique based on the research dataset should be compared. The methodology presented in **Chapter Three** offers a framework that could be used in future comparative studies. It was found to be particularly useful in evaluating the methods for both the performance and the implementation of the techniques for the given research dataset, despite the reported success of each technique in the literature;

- A variety of causes contribute to road pavement failure, as discussed in **Chapter Four**. Furthermore, failure is not always attributed to a sole cause; instead, it is common on road pavements for failure mechanisms to act jointly with or as secondary effects of others. This research not only yielded results that confirmed the practical knowledge but it also explored approaches to quantify this;
- Classification methods are appropriate for modelling the performance of road pavements, as demonstrated in **Chapters Six** and **Seven**. **Chapter Five** concluded that support vector machines, from those methods considered, was the most appropriate classification technique for this research, although probability trees also appeared to have a number of merits. However, support vector machines proved to be more suitable because of its superior model format and decision boundary versus the limited expressiveness and simplicity of the model format for, and subsequently the predictions from, probability trees;
- The developed prototype system was effective in predicting probabilities of failure for the independent testing dataset, as shown in **Chapter Seven**. However, using the prototype system without proper calibration on other road datasets can compromise the performance of the system. For the dataset tested in this research, this was found to be particularly the case for rutting;

- Furthermore, the developed prototype system can be used as a network management tool to determine the failure distribution across the network, the susceptibility to specific failure modes, and any symptomatic problems on the network, as shown in **Chapter Seven**. The failure charts developed in **Chapter Four** add a diagnostic element to the predictions enabling the most probable causes of failure to be identified;
- **Chapter Eight** discussed the importance of strength, composition, and traffic failure factors in predicting the probability of road pavement failure. Therefore, the collection and use of such information is recommended in order to improve the accuracy of predicting pavement performance;
- The inference of engineering knowledge into the development of the computational models improves the results in predicting pavement performance as presented in Schlotjes et al. (2011). Purely data driven processes are successful in situations where large amounts of unbiased data is available; however, the results will undoubtedly be resultant on the trends and relationships within the data. On the other hand, mechanistic techniques rely on the fundamental knowledge available, often overlooking site-specific causes of failure. To address this, the methodology incorporated engineering knowledge into the computational models trained on the research data, and
- The success of any computational model depends greatly on the quality of the data available to train the model with. The inputs included in the prototype system reflect the data collected in the LTPP programme and, given the detail included in the research dataset, may not always be available in practice.

9.2 Recommendations for Implementing the Research

While the results presented in Chapter Seven have demonstrated the promise of the system, it is recommended that in order for the developed support vector models to be implemented in New Zealand, the model needs to be calibrated and tested on a wide variety of different road networks following the procedure described in Chapter Seven. At the time of submission of this research, such testing was being facilitated on a variety of New Zealand road networks, with funding from the NZTA.

To further develop and improve the accessibility of the prototype system to practitioners, it is recommended that a computer language, appropriate for the software packages typically employed in the industry, be developed to call in the support vector machines model files (currently in a *.txt* file format in *R* language) for the direct implementation of the developed system. For the implementation of the system in a domain similar to that reported in this thesis, such as a network containing pavements of similar composition to the development of the prototype system, the network dataset can be passed through the uploaded model files, providing the inputs from the network dataset correspond to those of the system. However, for alternative road networks comprising of pavement compositions and traffic volumes other than those considered in this research, it is recommended that new support vector machines model files are developed, specific for the new dataset, following the methodology adopted throughout this thesis.

The methodology adopted in this research can be transferred to other road pavement types for the development of similar pavement performance systems; however, the assumptions of this research may not be suitable to other pavement domains. Internationally, pavement types and construction practices differ, such that alternative causes are responsible for failure of other pavement types than the knowledge inferred in this research, which captured the causes of flexible road pavement failure. Therefore, it is suggested to reconsider the causes included in the failure charts for true replication of alternative pavement failure. Although this research concluded support vector machines were the most appropriate classification technique for this case study, the performance of this technique may not be reproduced when using alternative road pavement datasets. Therefore, this research further recommends adopting the methodology of Chapter Five to evaluate the performance of alternative classification techniques, given the respective pavement data.

9.3 Further Work

A summary of the recommendations for the improvement of the prototype system are listed in the following sections, namely addressing the limitations of the research and for further work.

9.3.1 Recommendations to Address the Limitations of the Research

Chapter Eight discussed the limitations of the research; therefore, the following recommendations are suggested to address the issues identified.

Developed failure charts:

Despite the comprehensive methodology followed in Chapter Four, additional causes of failure may not have been represented in the failure charts. The methodology included site-specific failure causes; however, an analysis of a greater number of New Zealand road networks may further identify regional causes of failure. Furthermore, it is recommended to take into account the interactions between the failure mechanisms in the failure charts.

Comparative study:

To improve the comparative study, this research recommends evaluating the classification techniques using integrated performance measures and alternative sampling methods, in addition to the reported results. To do so, it is important that the research dataset includes information regarding the cost functions of incorrect predictions for the evaluation of the classification techniques using integrated performance measures, and larger amounts of failure data for the implementation of alternative sampling methods.

It is also recommended a cost effective analysis is undertaken, such as reported in (Levin and McEwan, 2001), to determine the overall effectiveness of the system.

Design of the prototype system:

In the design of the prototype system, basing the selection of successful factor combinations on 100 % accuracy threshold (or 90 % for the case of the shear models) may have excluded possible failure paths from the resultant system. For example, the factor combination(s) presenting an accuracy of 99 % were omitted from the remainder of the study, despite the minute difference between the accuracy and the target threshold. Reviewing this approach, it is suggested to cross-reference the excluded failure paths with the failure charts to ensure the omitted failure paths are not in fact possible for the respective failure mechanism. One method to address this would be to reduce the accuracy threshold, for example.

Multiple failure events should also be modelled with the support vector machines technique providing the data is forthcoming.

Prototype system testing:

This research tested the developed prototype system on an independent dataset in Chapter Seven; however, this testing dataset is atypical of network data. Although the results in Chapter Seven suggested that the developed rutting system is compromised when used with an independent dataset, further network testing is advised to determine the effectiveness of the system on typical road networks. It is suggested that the network datasets involved in the further testing of the system include as many input variables possible. Those associated with the strength of the pavement and its composition, as well as reliable traffic data, were shown to be particularly important in maximising the accuracy of the results.

Improvements to the data:

Possible improvements to the research dataset include information on the costs of failures for the implementation of integrated performance measures, information relating to the timings of multiple failures, processes to collate and quantify subjective and categorical data, and greater amounts of data representing multiple failures to use in modelling such failure types.

9.3.2 Recommendations for Further Work

The following recommendations are suggested to improve the developed prototype system:

Other classification techniques:

Chapter Two reviewed the advantages and disadvantages of a list of discriminative and generative classifiers, yet the comparative study in Chapter Five investigated the performance of only five methods. The selection of these methods was based on objective criteria, assessing the applicability and suitability of each of these methods to the study. However, the exclusion of the other methods (refer Section 2.6) resulted in the unknown performance of these to the research dataset. Therefore, this research suggests evaluating the excluded classification methods based on the research dataset.

Alternative failure mechanisms:

The scope of this research limited the development of the prototype system to three predominant structural failure mechanisms on New Zealand road networks. Other failure modes, generally associated with the surfacing layers, impede the function and integrity of road pavements. For a complete failure system, the framework of this research should be replicated focussing on predicting the probability of other failure types.

Forecasting of the predictions:

Employing pavement deterioration models to forecast the future condition state of the pavement has positively impacted on forecasting maintenance. As this research complements the current PMS practices and failure knowledge base, this research should be further developed so to include forecasting capabilities in the design of the prototype system; so over time, a measurable shift in network failure probabilities can be used in future maintenance decisions.

Uncertainty:

Comparisons of the research dataset and RCA network datasets identified a number of inconsistencies. The limited resources and funding available to RCAs restricts the data collected in road inspections, the collection methods employed, and the timings of road inspections. Further research is proposed to evaluate the confidence around the predicted probabilities by incorporating an uncertainty element into the system, taking into account:

- Missing data records resulting from uncollected information;

- Unreliable and inconsistent collection methods;
- Human error, and
- Subjective model inputs.

9.4 Lessons Learnt from this Research

This research has proposed a novel approach to predicting the probability of road pavement failure, contributing to the existing knowledge base of pavement performance models. The most significant lessons learnt, based on the findings of this research, include:

- Purely empirical modelling approaches lack the **inference of human knowledge** and purely mechanistic modelling approaches lack the **sophisticated technology offered by computational models**. The successful factor combinations highlighted in Chapter Five, and revisited in Chapter Six, correlated well with the developed failure charts from Chapter Four, since the model foundations were based on the failure paths. The combination of the two elements resulted in a well performing and accurate system;
- A number of modelling techniques are available, given current computer efficiency and power. However, **no single algorithm can be assumed superior**, without a robust comparison of the techniques using the same data;
- **Classification methods** are appropriate for implementation in modelling road pavement performance, and
- **Improvements in the current pavement data** are required for effective and accurate modelling of multiple failure incidents. Current methods can only

estimate the occurrence of such failures; however, it is expected computational models will be able to model these phenomena effectively.

REFERENCE LIST

Abraham, D.M., Wirahadikusumah, R., Short, T.J. and Shahbahrami, S. 1998. Optimization modeling for sewer network management. *Journal of Construction Engineering and Management*, Vol. 124, No.5, pp. 402-410.

Aitkin, M. and Foxall, R. 2003. Statistical modelling of artificial neural networks using the multi-layer perceptron. *Statistics and Computing*, Vol. 13, Issue 3, pp. 227-239.

Amador-Jiménez, L.U. and Mrawira, D. 2011. Reliability-based initial pavement performance deterioration modelling. *International Journal of Pavement Engineering*, Vol. 12, No. 2, pp. 177-186.

Anghel, C.I. 2009. Risk assessment for pipelines with active defects based on artificial intelligence methods. *International Journal of Pressure Vessels and Piping*, Vol. 86, Issue 7, pp. 403-411.

Arambula, E., George, R., Xiong, W. and Hall, G. 2011. Development and validation of pavement performance models for the state of Maryland. *Transportation Research Record*, Vol. 2225, pp. 25-31.

Arampamoorthy, H. and Patrick, J.E. 2010. Failure probability of New Zealand pavements. NZ Transport Agency Research Report 421. Wellington, New Zealand: NZ Transport Agency.

Ariaratnam, S.T., El-Assaly, A. and Yang, Y. 2001. Assessment of infrastructure inspection needs using logistic models. *Journal of Infrastructure Systems*, Vol. 7, No. 4, pp. 160-165.

Austin, P.C., Tu, J.V. and Lee, D.S. 2010. Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure. *Journal of Clinical Epidemiology*, Vol. 63, pp. 1145-1155.

Austroads. 2006. *Guide to Asset Management: Part 5E: Cracking*. Sydney, Australia: Austroads Inc.

Austroads. 2007a. *Guide to Asset Management: Part 5B: Roughness*. Sydney, Australia: Austroads Inc.

Austroads. 2007b. *Guide to Asset Management: Part 5C: Rutting*. Sydney, Australia: Austroads Inc.

Austroads. 2009a. *Guide to Asset Management: Part 5: Pavement Performance.* Sydney, Australia: Austroads Inc.

Austroads. 2009b. *Guide to Asset Management: Part 5H: Performance Modelling.* Sydney, Australia: Austroads Inc.

Austroads. 2012. *Guide to Pavement Technology Part 2: Pavement Structural Design*. Sydney, Australia: Austroads Ltd.

Ayyub, B.M. and McCuen, R.H. 2003. *Probability, Statistics, and Reliability for Engineers and Scientists.* Second Edition. Florida, United States of America: Chapman and Hall/CRC Press LLC.

Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Peterson, M., Van Kenhove, P. and Vanthienen, J. 2004. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. *European Journal of Operational Research*, Vol. 156, Issue 2, pp. 508-523.

Baik, H-S., Jeong, H.S. and Abraham, D.M. 2006. Estimating transition probability in Markov chain-based deterioration models for management of wastewater systems. *Journal of Water Resources Planning and Management*, Vol. 132, Issue 1, pp. 15-24.

Bailey, R., Patrick, J.E. and Jackett, R. 2006. Relationship between design and predicted performance of New Zealand pavements. Land Transport New Zealand Research Report 259. Wellington, New Zealand: Land Transport New Zealand.

Bapat, R.B. 2000. *Linear Algebra and Linear Models*. Second Edition. New York, United States of America: Springer-Verlag Inc.

Basheer, P.A.M., Chidiac, S.E. and Long, A.E. 1996. Predictive models for deterioration of concrete structures. *Construction and Building Materials*, Vol. 16, No. 1, pp. 27-37.

Bell, D.D., Cox, L., Jackson, S. and Schaefer, P. 1992. Using casual reasoning for automated failure modes and effects analysis. Presented at the Annual Reliability and Maintainability Symposium. Las Vegas, United States of America, pp. 343-353.

Ben-Akiva, M. and Gopinath, D. 1995. Modeling infrastructure performance and user costs. *Journal of Infrastructure Systems*, Vol. 1, No.1, pp. 33-43.

Ben-David, A. 2008. About the relationship between ROC curves and Cohen's kappa. *Engineering Applications of Artificial Intelligence,* Vol. 21, Issue 6, pp. 874-882.

Bernard, J., Gruening, J. and Hoffmeister, K. 1998. Evaluation of vehicle/driver performance using genetic algorithms. Paper No. 980227. Pennsylvania, United States of America: Society of Automotive Engineers, Inc.

Bhattacharya, B. and Solomatine, D.P. 2006. Machine learning in soil classification. *Neural Networks*, Vol. 19, Issue 2, pp. 186-195.

Bicego, M., Pękalska, E., Tax, D.M.J. and Duin, R.P.W. 2009. Component-based discriminative classification for hidden Markov models. *Pattern Recognition*, Vol. 42, Issue 11, pp. 2637-2648.

Cai, K-Y. 1996. System failure engineering and fuzzy methodology: An introductory overview. *Fuzzy Sets and Systems*, Vol. 83, Issue 2, pp. 113-133.

Caruana, R. and Niculescu-Mizil, A. 2006. An empirical comparison of supervised learning algorithms. Presented at the 23rd International Conference on Machine Learning. Pennsylvania, United States of America.

Chandra, D.K., Ravi, V. and Bose, I. 2009. Failure prediction of dotcom companies using hybrid intelligent techniques. *Expert Systems with Applications*, Vol. 36, Issue 3, Part 1, pp. 4830-4837.

Chen, M., Zheng, A.X., Lloyd, J., Jordan, M.I. and Brewer, E. 2004. Failure diagnosis using decision trees. Presented at the 1st IEEE International Conference on Autonomic Computing (ICAC). New York, United States of America.

Committee of State Road Authorities (CSRA). 1992. *TMH 9: 1992: Pavement Management System: Standard Visual Assessment Manual for Flexible Pavements*. Pretoria, South Africa.

Contini, S. and Matuzas, V. 2011. Analysis of large fault trees based on functional decomposition. *Reliability Engineering and System Safety,* Vol. 96, Issue 3, pp. 383-390.

Cortes, C. and Vapnik, V. 1995. Support-vector networks. *Machine Learning*, Vol. 20, Issue 3, pp. 273-297.

Creagh, M. 2005. Risk assessment for unbound granular materials and pavement performance. Presented at the Road System and Engineering Technology Forum. Brisbane, Australia.

Davies, S. and Ihaka, R. 2010. *Fitting generalized linear models*. R Documentation. [Online]. Available: http://127.0.0.1:25106/library/stats/html/glm.html. [October 5, 2010].

Deng, Z. and Henning, T.F.P. 2012. Testing of new NZTA condition performance measurers. NZ Transport Agency Research Report. Unpublished.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D. and Weingessel, A. 2011. *Package 'e1071'*. R Documentation. [Online]. Available: http://cran.r-project.org/web/packages/e1071/e1071.pdf. [October 5, 2010].

Dreiseitl, S. and Ohno-Machado, L. 2002. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Information*, Vol. 35, Issue 5-6, pp. 352-359.

Eastaugh, J.L., Smye, S.W., Snowden, S., Walker, J.J., Dear, P.R.F. and Farrin, A. 1997. Comparison of neural networks and statistical models to predict gestational age at birth. *Neural Computing and Applications*, Vol. 6, Issue 3, pp. 158-164.

Eftekhar, B., Mohammad, K., Ardebili, H.E., Ghodsi, M. and Ketabchi, E. 2005. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Medical Informatics and Decision Making*, Vol. 5, No. 3.

Emery, S. J. 1992. *The Prediction of Moisture Content in Untreated Pavement Layers and an Application to Design in Southern Africa. DRTT Bulletin 20*. Pretoria: Division of Roads and Transport Technology (DRTT), Council for Scientific and Industrial Research (CSIR).

Everitt, B.S. and Hothorn, T. 2006. *A Handbook of Statistical Analyses Using R.* Florida, United States of America: Taylor and Francis Group.

Faraway, J.J. 2006. *Extending the Linear Model with R*. Florida, United States of America: Taylor and Francis Group.

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters*, Vol. 27, Issue 8, pp. 861-874.

Foote, J.T. 1994. Decision-tree probability modeling for HMM speech recognition. Rhode Island, United States of America: Brown University [PhD Thesis].

Gao, L., Aguiar-Moya, J.P. and Zhang, Z. 2012. Bayesian analysis of heterogeneity in modeling of pavement fatigue cracking. *Journal of Computing in Civil Engineering*, Vol. 26, No.1, pp. 37-43.

Gil, D. and Johnsson, M. 2011. Support Vector Machines in Medical Classification Tasks, in Boyle, B.H. (Ed.). *Support Vector Machines: Data Analysis, Machine Learning and Applications*. New York, United States of America: Nova Science Publishers, Inc.

Gribble, M. and Patrick, J. 2008. Adaptation of the AUSTROADS pavement design guide for New Zealand conditions. Land Transport New Zealand Research Report 305. Wellington, New Zealand: Land Transport New Zealand.

Gualtieri, J.A., Chettri, S.R., Cromp, R.F. and Johnson, L.F. 1999. Support vector machine classifiers as applied to AVIRIS data. Presented at the 8th Annual JPL Airborne Earth Science Workshop. California, United States of America.

Gupta, A. and Lam, S.M. 1998. Weight decay backpropagation for noisy data. *Neural Networks*, Vol. 11, Issue 6, pp. 1127-1137.

Gupta, A., Kumar, P. and Rastogi, R. 1997. Pavement deterioration and maintenance model for low volume roads. *International Journal of Pavement Research and Technology*, Vol. 4, Issue 4, pp. 195-202.

Haas, R., Hudson, W.R. and Zaniewski, J.P. 1994. *Modern Pavement Management*. Malabar, Florida, United States of America: Krieger Publishing Co.

Halilčević, S.S., Gubina, F. and Gubina, A.F. 2011. The composite fuzzy reliability index of power systems. *Engineering Applications of Artificial Intelligence*, Vol. 24, Issue 6, pp. 1026-1034.

Hawkins, P.G. and Woollons, D.J. 1998. Failure modes and effects analysis of complex engineering systems using functional models. *Artificial Intelligence in Engineering*, Vol. 12, Issue 4, pp. 375-397.

Hayward, B.J. 2006. Investigation of road base shear strains using in-situ instrumentation. Christchurch, New Zealand: The University of Canterbury [ME (Transportation) Thesis].

He, J., Zhang, Y., Li, X. and Shi, P. 2012. Learning naïve Bayes classifiers from positive and unlabelled examples with uncertainty. *International Journal of Systems Science*, Vol. 43, No. 10, pp. 1805-1825.

Henning, T.F.P. 2008. The development of pavement deterioration models on the state highway network of New Zealand. Auckland, New Zealand: The University of Auckland [PhD Thesis].

Henning, T.F.P. and Roux, D.C. 2008. Pavement deterioration models for asphalt-surfaced pavements in New Zealand. NZ Transport Agency Research Report 367. Wellington, New Zealand: NZ Transport Agency.

Henning, T.F.P., Roux, D.C. and Alabaster, D. 2007. Benchmarking pavement performance between Transit's LTPP and CAPTIF programs. Land Transport New Zealand Research Report 319. Wellington, New Zealand: Land Transport New Zealand.

Henning, T.F.P., Costello, S.B., Dunn, R.C.M., Parkman, C.C. and Hart, G. 2004. The establishment of a long-term pavement performance study on the New Zealand state highway network. *Road and Transport Research*, Vol. 13, Issue 2, pp. 17-32.

Henning, T.F.P., Costello, S.B., Parkman, C.C., Bennett, C.R. and Dunn, R.C.M. 2006. A complete review of the crack initiation models for New Zealand. Presented at the 22nd ARRB Conference – Research into Practice. Canberra, Australia.

Horak, E. 2008. Benchmarking the structural condition of flexible pavements with deflection bowl parameters. *Journal of the South African Institution of Civil Engineering*, Vol. 50, No. 2, pp. 2-9.

Hu, H., Du, X., Xu, C., Zhao, F., Lin, X. and Bo, Z. 2011. Risk assessment of cascading failure in power systems based on uncertainty theory. Presented at the IEEE Power and Energy Society General Meeting. Detroit, Michigan, United States of America.

Hu, Y-C. 2010. Analytic network process for pattern classification problems using genetic algorithms. *Information Sciences*, Vol. 180, Issue 13, pp. 2528-2539.

Hussain, J., Henning, T., Wilson, D.J. and Alabaster, D. 2011. What happens when it rains? Performance of unbound flexible pavements in accelerated pavement testing. *Road and Transport Research,* Vol. 20, Issue, 4, pp. 3-15.

Isa, A.H.M., Ma'soem, D.M. and Hwa, L.T. 2005. Pavement performance model for federal roads. Presented at Eastern Asia Society for Transportation Studies (EASTS). Bangkok, Thailand, Vol. 5, pp. 428-440.

Jagielska, I., Matthews, C. and Whitfort, T. 1999. An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems. *Neurocomputing*, Vol. 24, Issue 1-3, pp. 34-54.

Karatzoglou, A., Meyer, D. and Hornik, K. 2006. Support vector machines in R. *Journal of Statistical Software*, Vol. 15, Issue 9, pp. 1-28.

Kaseko, M.S. and Ritchie, S.G. 1993. A neural network-based methodology for pavement crack detection and classification. *Transport Research Part C: Emerging Technologies*, Vol. 1, No. 4, pp. 275-291.

Kecman, V. 2005. Support Vector Machines – An Introduction, in Wang, L. (Ed.). *Support Vector Machines: Theory and Applications*. Heidelberg, Germany: Springer-Verlag Berlin Heidelberg.

Kleiner, Y., Adams, B.J. and Rogers, J.S. 2001. Water distribution network renewal planning. *Journal of Computing in Civil Engineering*, Vol. 15, No.1, pp. 15-26.

Kuncheva, L.I. 2006. On the optimality of naïve Bayes with dependent binary features. *Pattern Recognition Letters*, Vol. 27, Issue 7, pp. 830-837.

Land Transport New Zealand. 2008. *Land Transport New Zealand road network statistics for the year ending 30 June 2008.* [Online]. Available: http://www.landtransport.govt.nz/performance/2008/docs/network-statistics.pdf. [March 10, 2009].

Lazzaroni, M., Cristaldi, L., Peretto, L., Rinaldi, P. and Catelani, M. 2011. *Reliability Engineering: Basic Concepts and Applications in ICT*. Heidelberg, Germany: Springer-Verlag Berlin Heidelberg.

Levin, H.M. and McEwan, P.J. 2001. *Cost-Effectiveness Analysis: Methods and Applications*. Second Edition. California, United States of America: Sage Publications, Inc.

Liaw, A. and Wiener, T. 2012. *Package 'randomForest'*. R Documentation. [Online]. Available: http://cran.r-project.org/web/packages/randomForest/randomForest.pdf. [December 2, 2010].

Lim, T-S., Loh, W-Y. and Shih, Y-S. 2000. A comparison of prediction accuracy, complexity, and training time for thirty-three old and new classification algorithms. *Machine Learning*, Vol. 40, Issue 3, pp. 203-228.

Lindhe, A., Rosén, L., Norberg, T. and Bergstedt, O. 2009. Fault tree analysis for integrated and probabilistic risk analysis of drinking water systems. *Water Research*, Vol. 43, Issue 6, pp. 1641-1653.

Long, P.M., Servedio, R.A. and Simon, H.U. 2007. Discriminative learning can succeed where generative learning fails. *Information Processing Letters*, Vol. 103, Issue 4, pp. 131-135.

Long, W., Sato, Y. and Horigome, M. 2000. Quantification of sequential failure logic for fault tree analysis. *Reliability Engineering and System Safety*, Vol. 67, Issue 3, pp. 269-274.

Lounis, Z. and Madanat, S.M. 2002. Integrating mechanistic and statistical deterioration models for effective bridge management. Presented at the 7th International Conference on Applications of Advanced Technologies in Transportation. Massachusetts, United States of America.

Lytton, R.L. 1987. Concepts of pavement performance and modelling. Presented at the 2nd North American Conference on Managing Pavements. Ontario, Canada.

Madanat, S.M., Karlaftis, M.G. and McCarthy, P.S. 1997. Probabilistic infrastructure deterioration models with panel data. *Journal of Infrastructure Systems*, Vol. 3, No.1, pp. 4-9.

Madsen, B. 2011. *Statistics for Non-Statisticians*. Heidelberg, Germany: Springer-Verlag Berlin Heidelberg.

Manik, A., Gopalakrishnan, K. and Khaitan, S.K. 2007. Statistical modeling of accelerated pavement failure using response surface methodology. *International Journal of Computer, Information, and Systems Science, and Engineering*, Vol. 1, Issue 1, pp. 59-67.

Martin, T.C. 2008. Predicting sealed granular pavement deterioration at a road network level. Victoria, Australia: Monash University [PhD Thesis].

Menard, S. 2002. *Applied Logistic Regression Analysis*. Second Edition. Sage University Paper No. 106. London, United Kingdom: Sage Publications, Inc.

Menard, S. 2010. *Logistic Regression: From Introductory to Advanced Concepts and Applications*. California, United States of America: Sage Publications, Inc.

Mendenhall, W. and Beaver, R.J. 1991. *Introduction to Probability and Statistics*. Eighth Edition. Massachusetts, United States of America: PWS-KENT Publishing Company.

Morcous, G. and Lounis, Z. 2005. Maintenance optimization of infrastructure networks using genetic algorithms. *Automation in Construction*, Vol. 14, Issue 1, pp. 129-142.

Morcous, G., Rivard, H. and Hanna, A.M. 2002. Case-based reasoning system for modeling infrastructure deterioration. *Journal of Computing in Civil Engineering*, Vol. 16, No. 2, pp. 104-114.

Noble, W.S. 2004. Support Vector Machine Applications in Computational Biology, in Schoekkopf, B., Tsuda, K. And Vert, J.P. (Eds). *Kernel Methods in Computational Biology.* Cambridge, Massachusetts, United States of America: MIT Press, pp. 71-92.

Örkcü, H.H. and Bal, H. 2011. Comparing performances of backpropagation and genetic algorithms in the data classification. *Expert Systems with Applications*, Vol. 38, Issue 4, pp. 3703-3709.

Ortmeier, F. and Schellhorn, G. 2007. Formal fault tree analysis – practical experiences. *Electronic Notes in Theoretical Computer Science*, Vol. 185, Special Issue, pp. 139-151.

Osman, H. and Bainbridge, K. 2011. Comparison of statistical deterioration models for water distribution networks. *Journal of Performance of Constructed Facilities*, Vol. 25, No. 3, pp. 259-266.

Pal, M. 2006. Support vector machines-based modelling of seismic liquefaction potential. *International Journal for Numerical and Analytical Methods in Geomechanics*, Vol. 30, Issue 10, pp. 983-996.

Pan, N-F., Ko, C-H., Yang, M-D. and Hsu, K-C. 2011. Pavement performance prediction through fuzzy regression. *Expert Systems with Applications*, Vol. 38, Issue 8, pp. 10010-10017.

Pan, N-F., Lin, T-C. and Pan, N-H. 2009. Estimating bridge performance based on a matrix-driven fuzzy linear regression model. *Automation in Construction*, Vol. 18, Issue 5, pp. 578-586.

Papagiannakis, A.T. and Masad, E.A. 2008. *Pavement Design and Materials*. New Jersey, United States of America: John Wiley and Sons.

Park, D-Y., Buch, N. and Suh, Y-C. 2001. Development of fatigue cracking prediction model for flexible pavements. *KSCE Journal of Civil Engineering*, Vol. 5, No. 4, pp. 397-402.

Parker, C. 2011. An analysis of performance measures for binary classifiers. Presented at the 11th IEEE International Conference on Data Mining. Vancouver, Canada.

Patev, R.C., Putcha, C. and Foltz, S.D. 2005. *Methodology for Risk Analysis of Dam Gates and Associated Operating Equipment Using Fault Tree Analysis*. United States of America: US Army Corps of Engineers.

Patrick, J. 2009. The waterproofness of first-coat chipseals. NZ Transport Agency Research Report 390. Wellington, New Zealand: NZ Transport Agency.

Peng, C-Y.J., Lee, K.L. and Ingersoll, G.M. 2002. An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, Vol. 96, No. 1, pp. 3-14.

Perlich, C., Provost, F. and Simonoff, J.S. 2003. Tree induction vs logistic regression: A learning-curve analysis. *Journal of Machine Learning Research*, Vol. 4, Issue 2, pp. 211-255.

Pickard, K., Müller, P. and Bertsche, B. 2005. Multiple failure mode and effects analysis – An approach to risk assessment of multiple failures with FMEA. Presented at the Annual Reliability and Maintainability Symposium. Alexandria, Virginia, United States of America, pp. 457-462.

Powers, D.M.W. 2011. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, Vol. 2, Issue 1, pp. 37-63.

Prinzie, A. and Van den Poel, D. 2008. Random forests for multiclass classification: Random multinomial logit. *Expert Systems with Applications*, Vol. 34, Issue 3, pp. 1721-1732.

Prozzi, J.A. and Madanat, S.M. 2002. A nonlinear model for predicting pavement serviceability. Presented at the 7th International Conference on Applications of Advanced Technologies in Transportation. Massachusetts, United States of America.

Pruim, R. 2010. *Foundations and Applications of Statistics: An Introduction Using R.* Rhode Island, United States of America: American Mathematical Society.

R Core Team. 2011. *Wilcoxon rank sum and signed rank tests*. R Documentation. [Online]. Available: http://stat.ethz.ch/R-manual/R-patched/library/stats/html/wilcox.test.html. [February 2, 2011].

Rau, C.A., Becker, D.G., Besuner, P.M., Cipolla, R.C., Egan, G.R., Gupta, P., Johnson, D.P., Omry, U., Tetelman, A.S., Rettig, T.W. and Peters, D.C. 1977. Failure analysis and failure prevention in electric power systems. *Nuclear Engineering and Design*, Vol. 43, Issue 1, pp. 1-97.

Raudys, Š. 2001. *Statistical and Neural Classifiers: An Integrated Approach to Design*. London, United Kingdom: Springer-Verlag London Limited.

Razi, M.A. and Athappilly, K. 2005. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, Vol. 29, Issue 1, pp. 65-74.

Reay, K.A. and Andrews, J.D. 2002. A fault tree analysis strategy using binary decision diagrams. *Reliability Engineering and System Safety*, Vol. 78, Issue 1, pp. 45-56.

Reed, R.D. and Marks II, R.J. 1999. *Neural Smithing*. Massachusetts, United States of America: The MIT Press.

Reigle, J. 2000. Development of an integrated project-level pavement management model using risk analysis. West Virginia, United States of America: West Virginia University [Ph.D. Thesis].

Remenyte-Prescott, R. and Andrews, J.D. 2008. An enhanced component connection method for conversion of fault trees to binary decision diagrams. *Reliability Engineering and Safety System*, Vol. 93, Issue 10, pp. 1534-1550.

Ripley, B. 2012. *Package 'nnet'*. R Documentation. [Online]. Available: http://cran.r-project.org/web/packages/nnet/nnet.pdf. [October 5, 2010].

Robinson, R. 2008. *Restructuring Road Institutions, Finance and Management Volume 1: Concepts and Principles*. Birmingham, United Kingdom: The University of Birmingham.

Robnik-Šikonja, M. 2004. Improving random forests. Presented at the 15th European Conference on Machine Learning (ECML). Pisa, Italy.

Rogers, S. and Girolami, M. 2012. *A First Course in Machine Learning*. Florida, United States of America: Taylor and Francis Group.

Roiger, R.J. and Geatz, M.W. 2003. *Data Mining: A Tutorial-Based Primer*. United States of America: Pearson Education Inc.

Sadek, A.W., Freeman, T.E. and Demetsky, M.J. 1996. Deterioration prediction modeling of Virginia's interstate highway system. *Transportation Research Record*, Vol. 1524, pp. 118-129.

Saghafi, B., Hassani, A., Noori, R. and Bustos, M.G. 2009. Artificial neural networks and regression analysis for predicting faulting in jointed concrete pavements considering base condition. *International Journal of Pavement Research and Technology*, Vol. 2, No. 1, pp. 20-25.

Salt, G., Henning, T.F.P., Stevens, D. and Roux, D.C. 2010. Rationalisation of the structural capacity definition and quantification of roads based on falling weight deflectometer tests. NZ Transport Agency Research Report 401. Wellington, New Zealand: NZ Transport Agency.

Schlotjes, M.R. and Henning, T.F.P. 2012. Knowing your risk envelope – An enhanced understanding of the health of road networks. Presented at the 25th Ingenium Annual Conference. Rotorua, New Zealand.

Schlotjes, M.R., Henning, T.F.P. and Burrow, M.P.N. 2011. Improved network understanding – a diagnostic approach to the risk of pavement failure. Presented at the 8th International Conference on Managing Pavement Assets (ICMPA). Santiago, Chile.

Schlotjes, M.R., Henning, T.F.P. and Burrow, M.P.N. 2012a. The development of a diagnostic approach that predicts failure risk of flexible road pavements. Presented at the 7th International Conference on Maintenance and Rehabilitation of Pavements and Technological Control (MAIREPAV). Auckland, New Zealand.

Schlotjes, M.R., Henning, T.F.P. and St George, J.D. 2009. The risks associated with pavement failure in New Zealand. Presented at NZIHT and NZTA 10th Annual Conference. Rotorua, New Zealand.

Schlotjes, M.R., Burrow, M.P.N., Evdorides, H.T. and Henning, T.F.P. 2013a. A methodology to assess the causes and probability of road pavement structural failure. Presented at the Transportation Research Board (TRB) 92nd Annual Meeting. Washington D.C., United States of America.

Schlotjes, M.R., Burrow, M.P.N., Evdorides, H.T. and Henning, T.F.P. 2013b. Using SVM to predict the probability of pavement failure. Manuscript accepted for publication in *Proceedings of the Institution of Civil Engineers - Transport*.

Schlotjes, M.R., Henning, T.F.P., Burrow, M.P.N. and St. George, J.D. 2012b. Incorporating engineering knowledge to diagnose structural pavement failures. Manuscript submitted for publication.

Seyed-Hosseini, S.M., Safaei, N. and Asgharpour, M.J. 2006. Reprioritization of failures in a system failure mode and effects analysis by decision making trial and evaluation laboratory technique. *Reliability Engineering and System Safety*, Vol. 91, Issue 8, pp. 872-881.

Sharabah, A., Setunge, S. and Zeephongsekul, P. 2006. Use of Markov chain for deterioration modeling and risk management of infrastructure assets. Presented at the IEEE International Conference on Information Acquisition (ICIA). Weihai, China.

Sirikulviriya, N. and Sinthupinyo, S. 2011. Integration of rules from a random forest. Presented at the 2011 International Conference on Information and Electronics Engineering. Singapore.

Sokolova, M. and Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, Vol. 45, Issue 4, pp. 427-437.

Son, C-S., Kim, Y-N., Kim, H-S., Park, H-S. and Kim, M-S. 2012. Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. *Journal of Biomedical Informatics*, Vol. 45, Issue 5, pp. 999-1008.

Southern Africa Transport and Communications Commission (SATCC). 2003. *Guideline: Low-volume Sealed Roads.* Gaborone, Botswana: Southern African Development Community.

Themeau, T.M. and Atkinson, B. 2012. *Package 'rpart'.* R Documentation. [Online]. Available: http://cran.r-project.org/web/packages/rpart/rpart.pdf. [November 17, 2010].

Thom, N. 2008. *Principles of Pavement Engineering*. London, United Kingdom: Thomas Telford Limited.

Tso, S.K., Gu, X.P., Zeng, Q.Y. and Lo, K.L. 1998. An ANN-based multilevel classification approach using decomposed input space for transient stability assessment. *Electric Power Systems Research*, Vol. 46, Issue 3, pp. 259-266.

Tu, J. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, Vol. 49, No. 22, pp. 1225-1231.

Van Looy, S., Verplancke, T., Benoit, D., Hoste, E., Van Maele, G., De Turck, F. and Decruyenaere, J. 2007. A novel approach for prediction of tacrolimus blood concentration in liver transplantation patients in the intensive care unit through support vector regression. *Critical Care*, Vol. 11, Issue 4.

Verikas, A., Gelzinis, A. and Bacauskiene, M. 2011. Mining data with random forests: A survey and new results of new tests. *Pattern Recognition*, Vol. 44, Issue 2, pp. 330-349.

Volkanovski, A., Čepin, M. and Mavko, B. 2009. Application of the fault tree analysis for assessment of power system reliability. *Reliability Engineering and System Safety*, Vol. 94, Issue 6, pp. 1116-1127.

Wang, J., Ruxton, T. and Labrie, C.R. 1995. Design for safety of engineering systems with multiple failure state variables. *Reliability Engineering and System Safety*, Vol. 50, Issue 3, pp. 271-284.

Wang, K.C.P., Zaniewski, J. and Way, G. 1994. Probabilistic behavior of pavements. *Journal of Transportation Engineering*, Vol. 120, No. 3, pp. 358-375.

Wirahadikusumah, R., Abraham, D.M. and Iseley, T. 2001. Challenging issues in modeling deterioration of combined sewers. *Journal of Infrastructure Systems*, Vol. 7, No. 2, pp. 77-84.

Wu, Y., Ianakiecv, K. and Govindaraju, V. 2002. Improved k-nearest neighbor classification. *Pattern Recognition*, Vol. 35, Issue 10, pp. 2311-2318.

Xiao, N., Huang, H.Z., Li, Y., He, L. and Jin, T. 2011. Multiple failure modes analysis and weighted risk priority number evaluation in FMEA. *Engineering Failure Analysis*, Vol. 18, Issue 4, pp. 1162-1170.

Xu, K., Tang, L.C., Xie, M., Ho, S.L. and Zhu, M.L. 2002. Fuzzy assessment of FMEA for engine systems. *Reliability Engineering and System Safety*, Vol. 75, Issue 1, pp. 17-29.

Yang, J., Gunaratne, M., Lu, J.J. and Dietrich, B. 2005. Use of recurrent Markov chains for modeling the crack performance of flexible pavements. *Journal of Transportation Engineering*, Vol. 131, No. 11, pp. 861-872.

Yang, J., Lu, J.J., Gunaratne, M. and Dietrich, B. 2006. Modeling crack deterioration of flexible pavements: Comparison of recurrent Markov chains and artificial neural networks. *Transportation Research Record*, Vol. 1974, pp. 18-25.

Younis, R. and Knight, M.A. 2010. A probability model for investigating the trend of structural deterioration of wastewater pipelines. *Tunnelling and Underground Space Technology*, Vol. 25, Issue 6, pp. 670-680.

Yuhua, D. and Datao, Y. 2005. Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis. *Journal of Loss Prevention in the Process Industries*, Vol. 18, Issue 2, pp. 83-88.

Zhong, D., Sun, Y. and Li, M. 2011. Dam break threshold value and risk probability assessment for an earth dam. *Natural Hazards*, Vol. 59, Issue 1, pp. 129-147.

APPENDICES

Appendix A: Southland District Council Site Visit Notes

A research site visit to the Southland District, New Zealand, was completed between the dates of 20th April 2009 and 23rd April 2009. The purpose of this site visit was to:

- Become familiar with this road network;
- Gain sufficient background information into the appearance and properties of typical New Zealand pavement failures, and
- Gain experience with characteristic RCA road data specifically data from the RAMM database.

The information from this site visit aided the development of the failure charts. Appendix A presents the site visit notes from each road section.

Appendix B: Preliminary Data Analysis

B.1 SDC Data Analysis

Appendix B.1 provides the results from the preliminary data analysis using the Southland District Council road dataset. The data included in this dataset was obtained primarily from the RAMM database, and provides this research with data characteristically typical of RCA databases. This analysis aided in the development of the specific failure charts for rutting, cracking, and shear failures.

B.2 LTPP Data Analysis

Data from the LTPP programme established two datasets for this research. Appendix B.2 presents the results from the preliminary data analysis using both the Local Authority (LA) and State Highway (SH) LTPP datasets. This analysis aided in the development of the failure charts in Chapter Four and furthermore the State Highway LTPP analysis was used as a reference in Chapter Five.

Appendix C: Computational Model *R* Code for Each Classification Technique

The computational code employed in the *R* statistical programme throughout this research is presented below. The source files for each modelling technique are included in the Appendix folder (on the CD-rom):

- source_file_log_reg.r: for logistic regression;
- source_file_neural_networks.r: for neural networks;
- source_file_prob_tree.r: for probability trees;
- source_file_random_forest.r: for random forests, and
- source_file_svm.r: for support vector machines.

Further to the above, the following (Appendices C.1 to C.5) detail the individual code for each classification technique explored in Chapter Five.

C.1 Logistic Regression

The *glm()* function (Davies and Ihaka, 2010) was used to construct a logistic regression model in *R*, and a number of the arguments used in this function remained as the default arguments. The following shows the non-default arguments defined in this research:

Model = glm(formula, family = binomial(link = 'logit'), data, weights = wgts)

where:
formula = variable description of the model to be fit
data = dataset used to train the model, which contains all the variables included in the 'formula' argument
weights = weighting vector to account for the imbalance of the failures versus non-failures of the research dataset

The *family* argument described the error distribution and link function used in the model where, within this function, the *binomial* family was chosen as this permits the function for generating a logistic regression model (Pruim, 2010).

The *predict()* function below was used to predict the probability of failure; the *type* argument of *"response"* was selected to ensure the output is a probability.

predict(model, test_data, type="response")

where:
model= the trained logistic regression model, from above
test_data= the testing dataset to assess the modelling technique, and contains all the variables included in the model training

C.2 Neural Networks

The *nnet* package in *R* (Ripley, 2012) was used to construct a feed-forward neural network with a single hidden layer, as below:

Model = nnet(formula, size=size, decay=0.001, skip=T, Hess=T, maxit=1000, entropy=T, data=data, weights=wgts)

where:
formula = variable description of the model to be fit
data = dataset used to train the model, which contains all the variables included in the 'formula' argument
weights = weighting vector to account for the imbalance of the failures versus non-failures of the research dataset

The *size* argument defined the number of hidden units (nodes) in the hidden layer and, in this research, a range of 1 to 10 nodes inclusive were trialled. An arbitrary value for *decay* was selected to lower the magnitude of the weights, resulting in a more optimal solution. The *skip* argument was set to *TRUE* to allow the network to skip a layer, if necessary, so the optimal solution can connect straight from the input nodes to the output layer. *Hess* was also set to

TRUE so the measure of fit at the best set of weights found was returned in the output summary. The maximum number of iterations for the training of the model was set to 1000, using *maxit*, and *entropy* was set to *TRUE* to fit maximum conditional likelihood and provide probabilistic outputs.

The *predict()* function was used in the same manner as the logistic regression technique previously, with the *type* argument omitted, as the probability output was already defined in the model *nnet()* function.

predict(model, test_data)

where:
model= the trained neural network model, from above
test_data= the testing dataset to assess the modelling technique, and contains all the variables included in the model training

C.3 Support Vector Machines

The *svm()* function within the *e1071* package (Dimitriadou et al., 2011) in *R* was used to construct the support vector machines model. The research chose the radial basis function kernel, defined by the *kernel* argument, because of its ability to handle non-linear attributes and class relationships (Gil and Johnsson, 2011). The *type* argument was set to *"C"* to indicate a classification type of model, and the inclusion of the *probability* argument applies the maximum conditional likelihood concept, alike neural networks, to the decision values. The remaining arguments remained as the default values.

Model = svm(formula, kernel="radial", type="C", data, weights=wgts, probability=TRUE)

where:
formula = variable description of the model to be fit
data = dataset used to train the model, which contains all the variables included in the 'formula' argument
weights = weighting vector to account for the imbalance of the failures versus non-failures of the research dataset

Like the previous methods, the *predict()* function was used to establish the predictions of the testing data and included in the arguments were *probability=TRUE* and *decision.values=TRUE* to ensure the new predictions were presented as probabilities.

predict(model, test_data, probability=TRUE, decision.values=TRUE)

where:
model= the trained support vector machines model, from above
test_data= the testing dataset to assess the modelling technique, and contains all the variables included in the model training

C.4 Probability Trees

In *R*, the *rpart* package and *rpart()* function (Themeau and Atkinson, 2012) was used to construct a probability tree model . The *method* argument of this function defined the classification nature of the probability tree (i.e. the splitting rule).

Model = rpart(formula, data=data, weights=wgts, method="class")

where:
formula = variable description of the model to be fit
data = dataset used to train the model, which contains all the variables included in the 'formula' argument
weights = weighting vector to account for the imbalance of the failures versus non-failures of the research dataset

Pruning unnecessary branches optimised the model output; therefore, in the same package, the *prune()* function was applied to the *model* tree. This research selected the complexity parameter (*cp*) (i.e. the complexity of the tree) that was associated with the smallest cross-validated error, ensuring the optimal tree was delivered.

Pruned_Model=prune(model, cp=model$cptable[which.min(model$cptable[,"xerror"]), "CP"])

The *predict()* function was used to establish the predictions for the testing dataset and the type argument defined the output as a probability.

predict(model, test_data, type="prob")

where:
model= the trained probability tree model, from above
test_data= the testing dataset to assess the modelling technique, and contains all the variables included in the model training

C.5 Random Forests

The *randomForest()* function, from the package of the same name (Liaw and Wiener, 2012), in *R* was used to construct a random forest ensemble.

Model = randomForest(formula, data=data, weights=wgts, xtest=NULL, importance=TRUE, keep.forest=TRUE)

where:
formula = variable description of the model to be fit
data = dataset used to train the model, which contains all the variables included in the 'formula' argument
weights = weighting vector to account for the imbalance of the failures versus non-failures of the research dataset

The *xtest* was not needed in this research, thus this argument was set as *NULL*. Throughout the construction of the forest, the importance of the predictors was assessed, defined by *'importance = TRUE'*. The argument *keep.forest=TRUE* ensured the forest information was retained in the output object for future reference. The remainder of the model arguments remained as their default parameters.

Identical to probability trees, the *predict()* function used the type argument to define the output as a probability.

predict(model, test_data, type="prob")

where:
model= the trained random forest model, from above
test_data= the testing dataset to assess the modelling technique, and contains all the variables included in the model training

Appendix D: Result Tables for Performance Measures

The results tables are published in Appendix D, which includes individual tables for each of the cross-validation tests (10 in total for each factor combination), which are then averaged to provide a single mean value and associated standard deviation for each performance measure per factor combination, per failure mechanism. In addition to the individual results tables, a summary table of the averaged cross-validation tests (for each factor combination per failure mechanism) is presented in this appendix. It is the data from these summary tables that are used in the density plots and box and whisker plots, also included in this appendix.

The following discusses the difference between the three datasets used in this analysis. This thesis reports on the results from the NAs=0 State Highway LTPP dataset (as detailed below):

- **Full:** The full State Highway LTPP dataset includes all the datapoints from the LTPP database. As part of the modelling process, missing values (such as NAs) are removed from the computational models, reducing the number of datapoints used in the training of the classification models;
- **Limited:** This dataset includes only the sites which have failed, and the data from the year prior to failure, and
- **NAs=0:** A full dataset from the State Highway LTPP database where the missing values were assigned a default value of zero (0) (as per Chapter Five) to ensure a greater number of datapoints available in the training of classification models.

Appendix E: Failure Paths and Causes

A table of successful factor combinations for each failure mechanism (rutting, fatigue cracking, and shear) is presented in this appendix. These combinations were based on the success of each, given the success threshold of 100 % accuracy for rutting and fatigue cracking failure, and 90 % for shear failure. Included in the table is a detailed description of the factor groups (from Table 4-1) for each combination, and the causes associated with such failure paths.

Appendix F: Performance Results of the Prototype System

This appendix includes:

- The raw predictions from the State Highway LTPP dataset (please note that additional data was collected between the years of 2010 – 2011 and added to the research dataset);
- Tables reporting on the performance measures for each failure mechanism (rutting, fatigue cracking, and shear), and box and whisker plots for each, and
- The raw data tables and plots for the sensitivity analysis, for each failure mechanism to test the robustness of the prototype system. These results show how the results of the output can be apportioned to the independent variables.

Appendix G: Case Study Testing

Appendix G contains the result tables from the case study described in Chapter Seven, including:

- The raw predictions from the prototype system evaluating the probability of failure for the Local Authority LTPP rural dataset;
- Tables recording the performance measures for each failure mechanism (rutting, fatigue cracking, and shear), and
- The calculation table for the overall failure probabilities for the Local Authority LTPP rural dataset, based on the two calculation methods discussed in Chapter Seven.