Analysing and Modelling the Impact of Aggressive Driving on Crash Accident Duration: Based on Telematics Data Driven

Shemsudin Ahmedteib Duri

A Thesis Submitted to the Department of Mechanical Engineering

College of Mechanical, Chemical and Materials Engineering

Presented for Partial Fulfillment of Masters of Science in Automotive Engineering

Office of graduate studies

Adama Science and Technology University

June, 2025

Adama, Ethiopia

Analysing and Modelling the Impact of Aggressive Driving on Crash Accident Duration: Based on Telematics Data Driven

Shemsudin Ahmedteib Duri

Advisor: Dr. Getachew Alemayehu

A Thesis Submitted to the Department of Mechanical Engineering

College of Mechanical, Chemical and Materials Engineering

Presented for Partial Fulfillment of Masters of Science in Automotive Engineering

Office of graduate studies

Adama Science and Technology University

June, 2025

Adama, Ethiopia

ORIGINALITY DECLARATION FORM FOR STUDENTS

This form must be completed and signed for all works submitted to the University for Examination.

Name of student	Shemsudin Ahmedteib Duri
Id. No.	PGE/28433/15
Department	Mechanical Engineering
College	Mechanical, Chemical and Material Engineering
Course Title	MSc Thesis
Title of work	Analysing and Modelling the Impact of Aggressive Driving on Crash Accident Duration: Based on Telematics Data Driven
Date of submission	May 2025

DECLARATION

1. I understand what plagiarism is and I am aware of the university's policy in this regard.
2. I declare that this thesis research project is my original work and has not been submitted elsewhere for examination or award of a degree or publication. Wherever others' work or my own work has been utilized, it has been duly acknowledged and referenced in compliance with the university's requirements.
3. I have not sought or used the services of any professional agencies to produce this work.
4. I have not permitted, nor will I permit, anyone to plagiarize my work with the intention of presenting it as their own.
5. I am aware that making any false claims regarding this work will lead to disciplinary action in accordance with the university's Anti-Plagiarism Guidelines.

Signature ________________________

Date ____________________________

DECLARATION

I hereby declare this thesis entitled **“Analysing and Modelling the Impact of Aggressive Driving on Crash Accident Duration: Based on Telematics Data Driven”** is my own work and has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation.

______________________	______________________	______________
Name of the student	Signature	Date

RECOMMENDATION

I, the advisor of this thesis, hereby certify that I have read the revised version of the thesis entitled **"Analysing and Modelling the Impact of Aggressive Driving on Crash Accident Duration: Based on Telematics Data Driven"** prepared under my guidance by Shemsudin Ahmedteib Duri. Therefore, I recommend the submission of the thesis to the department for further review and evaluation.

______________________________	______________	____________
Major Advisor	Signature	Date

APPROVAL SHEET

I here certify that the recommendation and suggestion given by the thesis review committee are appropriately incorporated into the final thesis entitled "**Analysing and Modelling the Impact of Aggressive Driving on Crash Accident Duration: Based on Telematics Data Driven"** developed by Shemsudin Ahmedteib Duri.

____________________	____________	____________
Major Advisor	Signature	Date

APPROVAL OF BOARD OF REVIEWERS

We, the undersigned, members of the Board of Reviewers of the thesis open defence by Shemsudin have read and evaluated the thesis entitled "**Analysing and Modelling the Impact of Aggressive Driving on Crash Accident Duration: Based on Telematics Data Driven"** and assessed the understanding of the candidate about the thesis research. This is, therefore, to certify that the thesis is accepted and we recommend the implementation of the thesis.

Chairperson	Signature	Date
Internal Examiner	Signature	Date
External Examiner	Signature	Date

Finally approval and acceptance of the thesis is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

Department head	Signature	Date
College Dean	Signature	Date
Office of Postgraduate Studies, Dean	Signature	Date

ACKNOWLEDGMENT

First and foremost, I wish to express my deepest gratitude and appreciation to my thesis advisor, Dr. Getachew Alemayehu, for his solid guidance, thoughtful to his advising role, and invaluable academic support throughout the courses to this research in partnership with Dr. Alemayehu Wakjira I have grateful and appreciation for both parties.

My heartfelt thanks also go to the Adama Science and Technology University, especially the Department of Mechanical Engineering, for providing the academic environment and institutional support that enabled the successful conduct of this study. I am also sincerely grateful to the members of my Thesis Committee Dr. Ramesh Babu, Dr. Deresse Firew, and Dr. Amanuel Gebisa whose insightful suggestions, constructive critiques, and academic supervision played a significant role in enhancing the quality of this work.

Special thanks are extended to the Managing Directors and all staff members of the Public Service Transport Service (PSTS). I would like to express particular appreciation to the Insurance Department as well as the IT and Telematics Control Department, whose cooperation, technical guidance, and generous provision of essential datasets were fundamental to the successful execution of this research. Their collaboration and dedication are deeply appreciated.

Finally, am further grateful to express my gratitude to the Oromia Transport Agency and my lovely friends, all other individuals and organisations for all their contributions and whose personally shared experience and encouragement throughout this journey, I express my sincerest gratitude.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AAA	American Automobile Association
ACT	Accident Causation Theory
ADS	Aggressive Driving Score
ABS	Antilock braking system
AD	Aggressive Driving
AFT	Accelerated Failure Time
AI	Artificial Intelligence
APA	American Psychological Association
ASTU	Adama Science and Technology University
BT	Behavioural Theory
CAN	Controller Area Network
CADI	Composite Aggressive Driving Index
CADS	Composite Aggressive Driving Score
Cox PH	Cox Proportional Hazards
EC	European Commission
ECU	Electronic Control Unit
EDA	Exploratory Data Analysis
EFTA	Ethiopian Federal Transport Authority
ELDs	Electronic Logging Devices
EMoTL	Ethiopian Ministry of Transport and Logistics
ERSS	Ethiopian Road Safety Strategy
ERSC	Ethiopian Road Safety Charter
ERTA	Ethiopian Road Transport Authority
EU	European Union
FMCSA	Federal Motor Carrier Safety Administration

FMVSS	Federal Motor Vehicle Safety Standards
GDP	Gross Domestic Product
GPS	Global Positioning System
HB	Harsh Brake
HBI	Harsh Brake Index
HBR	Harsh Braking Rate
HR	Hazard Ratio
IIHS	Insurance Institute for Highway Safety
IRTAD	International Traffic Safety Data and Analysis Group
ISO	International Organization of Standardization
ICT	Information and Communication Technology
IT	Information Technology
ITE	Institute of Transportation Engineers
ITS	Intelligent Transport Systems
JASP	Jeffreys's Amazing Statistics Program
KM	Kaplan–Meier
LD	Lane Departure
LDI	Lane Departure Index
LMICs	Low- and Middle-Income Countries
LR	Likelihood Ratio
ML	Machine Learning
MOT	Ministry of Transport
MoTL	Ministry of Transport and Logistics
NGOs	Non-Governmental Organizations
NHTSA	National Highway Traffic Safety Administration
OBD-II	On Board Diagnostics-Two
OECD	Organisation for Economic Co-operation and Development

OS	Over Speeding
OSI	Overspeeding Index
OPGS	Office of Postgraduate Study
PAYD	Pay-As-You-Drive
PHYD	Pay-How-You-Drive
PSTS	Public Service Transport Service
RTD	Raw Telematics Data
RHT	Risk Homeostasis Theory
RPM	Revolution per Minutes
SAE	Society of Automotive Engineering
SD	Standard Deviation
SDGs	Sustainable Development Goals
SGC	School of Graduate Committee
SHRP2	Strategic Highway Research Program 2
SIM	Subscriber Identity Module
SPSS	Statistical Package for the Social Science
SVMs	Support Vector Machines
TCU	Telematics Control Unit
TCD	Telematics Control Department
TD	Telematics Data
UBI	Usage-Based Insurance
U.K	United Kingdom
U.S	United States
UN	United Nations
UNECA	United Nations Economic Commission for Africa
UNECE	United Nations Economic Commission for Europe

UNESCAP United Nations Economic and Social Commission for Asia and the Pacific

VIF Variance Inflation Factor

WHO World Health Organization

LIST OF ABBREVIATIONS

C index	Concordance Index
df	Degrees of freedom
Dr	Doctor
Fig	Figure
HBfreq	Harsh Brake Frequency
Id	Identity (Identification)
Jamovi	Statistical Software (Based on R; used for statistical analysis)
Km	Kilometer
Km /hr	Kilometer per hour
LDfreq	Lane Departure (Erratic) Frequency
m	Mass
m/s^2	Meter per second square
n	Sample Size
No	Number
OSdur	Overspeeding Duration
PhD	Doctor of Philosophy
St	Street
v	Speed
XGBoost	Extreme Gradient Boosting
Z	Z-score

ABSTRACT

Aggressive driving behaviours such as overspeeding, harsh braking, and erratic steering are widely recognized as a major cause of road traffic crashes, yet limited research has focused on modelling the time to crash based on these behaviours using real-world telematics data. This study examine how these behaviours affect the time until a crash occurs, using six months of pre-crash telematics data from 43 cross-country public buses in Ethiopia. A quantitative, predictive modelling approach was employed to systematically examine the relationship between aggressive driving behaviours and timing to crash incidents using numerical telematics data. Survival analyses, Cox Proportional Hazards regression modelling founded that overspeeding is the most significant predictor of early crash risk. Each unit increase in the overspeeding index is associated with a 17% increase in crash hazard ($HR = 1.17$, $p < 0.001$), confirming as the leading risk factor. Harsh braking also contributes significantly to crash likelihood, showing an 8% increased hazard per unit increase ($HR = 1.08$, $p = 0.005$). Erratic steering, measured through lane departure frequency, has a moderate but statistically significant effect, increasing crash hazard by 6% per unit ($HR = 1.06$, $p = 0.027$). Driver behaviour was quantified through telematics-derived metrics overspeeding duration (seconds), harsh braking events, and lane departure events each normalized per 100 km of driving distance. Only crash cases where the PSTS driver was identified as the primary at-fault party, and the cause was directly linked to aggressive behaviour, were included in the analysis. The study employed descriptive statistics and correlation analysis (including visualizations), followed by survival analysis using Cox regression and Kaplan–Meier estimators. All analyses were conducted using Jamovi (v2.4.5) and JASP (v0.18.1) open-source platforms based on SPSS. The findings confirm that telematics based behavioural monitoring can be used to build predictive models for estimating time-to-crash, providing an opportunity for early intervention. The study recommends that transport authorities and fleet managers adopt these tools to identify high-risk drivers, initiate timely retraining, and implement data-informed road safety policies.

Key Words: *Aggressive Driving, Cox Regression, Correlation Analysis, Predictive Modelling, Survival Analysis, Telematics*

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Road traffic incidents are raising burden on public health and development in low- and middle-income countries (LMICs). The World Health Organization (2022) reports that approximately 1.35 million individuals die annually due to road traffic crashes, while an additional 20 to 50 million suffer non-fatal injuries, often resulting in long-term disabilities. These incidents are the leading cause of death among individuals aged 5–29 years and are a significant barrier to economic growth in developing nations (World Health Organization, 2022). The societal impact of these incidents includes lost productivity, increased healthcare expenditure, emotional trauma, and damage to infrastructure (World Health Organization, 2022).

Recent progress in data analytics, sensor technology, and digital connectivity has led to the growth of telematics as a modern transport safety field. Telematics refers to the integration of telecommunications and informatics systems in vehicles, enabling real-time collection, monitoring, and transmission of driving data (Virginia Tech Transportation Institute, 2023). These systems capture behavioural and mechanical metrics such as speed, brake pressure, steering angles, and vehicle location (Azuga, 2023; Sambasafety, 2023). By integrating telematics into transport systems, fleet managers and insurers can shift from responding to crashes after they happen to identifying and managing risks beforehand (Litman, 2005; Kirushanth & Kabaso, 2018). With such insights, it is now possible to design models that not only evaluate crash risk but also predict when a crash may happen based on aggressive driving behaviour (Sambasafety, 2023).

Aggressive driving often includes speeding, sudden braking, erratic steering, tailgating, and rapid acceleration, which are strongly linked to crashes (NHTSA, 2023; Singh & Ghosh, 2020). Among these, overspeeding, harsh braking, and lane violations are particularly dangerous, as they destabilise vehicle control, reduce driver reaction time, and increase the severity of potential collisions (Karthika, S., & Thomas, S. (2021)). Many international studies have shown strong statistical relationships between these behaviours and higher crash rates or severity (Research Gate, 2023; NHTSA, 2023).

1.1.1 Global Road Safety Context

Road traffic injuries remain one of the top contributors to preventable deaths globally. According to the World Health Organization (2022), about 1.35 million people die annually due to road traffic crashes. More than 90% of these fatalities occur in low- and middle-income countries despite having fewer vehicles. Aggressive driving—such as overspeeding, harsh braking and erratic steering is a key behavioural factor in these crashes. Figure 1.1 graphically presents the road traffic death rate (per 100,000 population) by WHO region in 2021 (World Health Organization, 2023) road fatalities per 100,000 by region (WHO, 2021) as the graph indicates the highest road traffic crash is recorded in African countries especially in sub Saharan African countries.

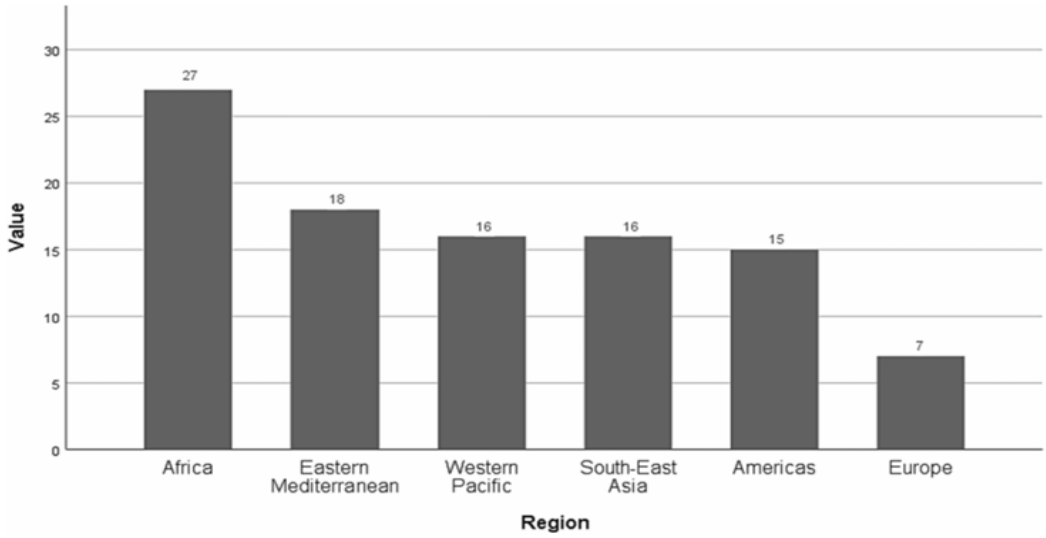

Figure 1.1: Road Fatalities per 100,000 populations (WHO (2023)

Efforts in road safety interventions have increasingly incorporated technology, especially telematics, to understand, monitor, and modify risky driving behaviour. The growing body of literature confirms that access to real-time vehicle and driver data can greatly improve crash prediction and prevention strategies (Selvaraj et al., 2024; Savolainen, 2025; Automotive Quest, 2025).

1.1.2 Ethiopian Road Safety Status

In Ethiopia, road safety challenges are caused by a combination of weak traffic enforcement, poor driver discipline, and insufficient infrastructure. Reports from the Ethiopian Federal Transport Authority (EFTA, 2022) and the Ministry of Transport and Logistics (MoTL, 2023) indicate that over 80% of traffic incidents are due to driver-related factors. In 2022, the Ethiopian Ministry of Transport and Logistics reported approximately 4,500 traffic-related deaths, disregarding tens of thousands of injuries. The national economic burden of road traffic crashes was estimated at nearly $290 million annually equivalent to 0.9% of the country's GDP (MoTL, 2023). Aggressive driving including overspeeding and failure to give way continues to be a key cause of fatalities in urban centers like Addis Ababa (Tulu et al., 2015; Gidey, 2010). Beshah Tesema et al. (2013) reported that driver misconduct such as overspeeding; poor lane discipline and failure to yield are responsible for more than 80% of traffic fatalities in urban Ethiopia.

1.1.3 The Rise of Telematics in Safety Analytics

Telematics has emerged as essential technology in modern transportation systems, significantly enhancing the ability to monitor vehicles and promote road safety through real-time data analytics (Kirushanth & Kabaso, 2018). Such systems incorporate onboard diagnostics, GPS tracking, gyroscopes, and accelerometers that allow for the continuous capture of driver actions and vehicle dynamics (Kirushanth & Kabaso, 2018). Recent industry reports estimate that fleet telematics adoption has exceeded 12 million active systems in North America, with adoption rates steadily rising across logistics and transportation sectors (Berg Insight, 2020).

From an academic and analytical point of view, telematics enables a new front line in predictive modelling, particularly in understanding crash risks and driving behaviours over time (Wolde, 2022). The transition from traditional reactive responses to predictive, data-guided safety mechanisms reflects a significant progression in how transport systems are managed globally (Litman, 2005; Wolde, 2022). By integrating telematics-derived behavioural metrics into crash-time prediction frameworks, it becomes possible to detect high-risk drivers early and inform strategic safety measures (Peer et al., 2020; Wijnands et al., 2018). In high-income countries, it supports dynamic insurance pricing models such as Pay-As-You-Drive (PAYD) and Pay-How-You-Drive (PHYD), where policyholders are charged based on their risk level (Litman, 2005).

1.2 Statement of the Problem

Road traffic incidents continue to increase a severe public health and economic challenge, particularly in low- and middle-income countries such as Ethiopia. Although the existence of traffic regulations and safety policies, the rate of road crashes especially among public transport vehicles is still remains high. Most of these crashes are not random occurrences but are closely linked to aggressive driving behaviours, like overspeeding, harsh braking, and erratic steering.

In Ethiopia, national crash data indicates that human error accounts for over 80% of road traffic incidents, with aggressive driving behaviour being a primary cause (Tulu et al., 2015; MoTL, 2023). The current national road safety assessments are mainly reactive, relying on retrospective crash data which is not integrated with behavioural data preceding crashes (UNECA, 2021; Wolde, 2022). On the other hand, the country still lacks a data-driven, behavioural monitoring system capable of predicting crashes before they happen. The lack of proactive monitoring systems means that fleet operators and policymakers are often unaware of which drivers or behaviours pose imminent crash risks until after incidents occur.

While telematics systems which collect real-time driving behaviour data are being increasingly used worldwide to support predictive safety models, their application in Ethiopian public transportation is still minimal. There is also a research gap in modelling the time to crash occurrence based on aggressive driving behaviour. Most local studies have focused on crash frequency and severity without dealing duration related dimension that is, how long a vehicle can operate under risky conditions before a crash is likely. This limits the potential for early warning systems and proactive interventions.

1.3 Objectives

1.3.1 General Objective

The main aim of this study was to analyze and model how aggressive driving behavior the duration of crash incident using telematics data collected from Ethiopian public service buses.

1.3.2 Specific Objectives

- To analyse the percentage of total driving time spent in overspeeding.
- To analyse the rate or frequency of harsh braking events.
- To analyse lane departure behaviours for their potential role in crash prediction.
- To compute a composite risk score from aggressive driving behaviours.
- To evaluate the influence of composite risk score on time-to-crash using survival modelling.

1.4 Research Questions and Hypotheses

1.4.1 Research Questions

- What proportion of total driving time was spent in overspeeding, and how did it relate to the duration of time to crash occurrence?
- How had the rate and frequency of harsh braking events influenced the time to crash?
- Were there identifiable patterns in erratic steering from telematics data lead to crash events?
- Could a predictive model accurately estimate the duration until a crash was likely to occur using overspeeding, harsh braking, and erratic steering data?

1.4.2 Hypotheses

There is a statistically direct relationship between aggressive driving behaviours (overspeeding duration, frequency of harsh braking, and erratic steering patterns) and the time to crash occurrence.

1.5 Significance of the Study

To identifying those risky behaviors, the integration of telematics data in study aggressive driving behaviors offers a good understanding of their impact on crash durations. This study has significant contribution in the following fields:

A. **Academic Contribution:** The academic value of this research lies in its application of time-to-crash using survival modelling to road traffic crash data.
B. **Transport Safety and Fleet Management:** To develop driver monitoring systems, improving risk mitigation, implementing driver training, warning systems for risky behaviour, preventive maintenance driving style.
C. **Policy and Governance Relevance:** It inform public policy on several levels:
   - **National Road Safety Administration**: Offers a data-driven model to support the Ethiopian Road Safety Strategy.
   - **Insurance Regulation**: Implements pay-how-you-drive system based on quantifiable crash risk between driving behaviour and crash timing.
D. Socio-Economic and Public Health Impact: To promotes economic productivity by safeguarding workforce safer.
E. Technological Innovation: It helps the practical application of telematics system in fleet and transport operation in Ethiopia by transforming it into predictive approach.

1.6 Scope of the Study

The scope of study is restricted to 43 cross county long-distance public transport intercity coaches operating under the Public Service Transport System (PSTS) fleet based in Addis Ababa. These vehicles are Yutong ZK6122-series (61-seater) intercity coaches with a single consistent assigned driver for at least 6 months observation period, selected for their uniformity in design, operation, and sensor instrumentation. Homogeneity of the fleet and driver ensures comparability and enhances the reliability of behaviour-based modelling outcomes. A 6-months pre crash duration telematics data record is assumed to be sufficient enough to express the driving pattern of a single driver, how the driver behaves while driving. There is no driving habits can completely resolved or changed in a few days while it can be changed through process when the persons consistently and seriously follow corrective actions or adjusting to action in needs. In addition psychological and behavioural science the behavioural theory disused in chapter 2 section 2.3 also supports the idea and leads to justify this assumption.

1.7 Limitations of the Study

Despite its methodological clarity, the study had shown the following limitations:

1. The telematics dataset had been limited to 43 identical buses. Actual sample selection is considered by taking into account the total population the fleet organization runs.
2. The quality and accuracy of the telematics data had depended on the consistency and functionality of the sensors installed in the buses, which may have minor errors.
3. The model does not account for variables like traffic density driver fatigue, mobile phone use in addition to weather or road condition, which are affect crash risk.

1.8 Delimitation of the Study

To ensure focus, depth, and methodological clarity, this study delimitation is extended to the following points:

1. The study is limited to 43 Yutong ZK6122-series intercity buses operated by a single consistent driver for at least six consecutive months prior to a crash were included, all of which are part of the same public transport organisation (PSTS). This is to ensure uniformity in performance characteristics and sensor configurations.
2. The study focused only on three aggressive driving behaviours overspeeding, harsh braking, and erratic steering based on their telematics detectability, frequency, and relevance to crash prediction. Crashes caused by environmental factors, system failures (vehicles) or other circumstances were not considered.
3. The analysis is focused to predicting the time of crash occurrence, and does not address injury severity, property damage, or economic losses associated with the crashes.

1.9 Organization of the Thesis

The structure of this thesis is organised into four main chapters and there is a final section; conclusion and recommendation, each chapters are focusing on a specific component of the research process, as summarized in the Table 1.1 below:

Table 1.1: Summary of Thesis Chapter Organization

Title	Content Description
Chapter One Introduction	Introduces the research by presenting the background, problem statement, research objectives, questions, hypotheses, significance, scope, and limitations.
Chapter Two Literature Review	Reviews relevant global, regional, and local literature on aggressive driving, crash prediction, and telematics. Also outlines theoretical frameworks and research gaps.
Chapter Three Materials and Methods	Details the research approach, data sources, sampling, variables, data collection methods, telematics indicators, and statistical techniques such as survival analysis and feature engineering.
Chapter Four Results and Discussion	Presents descriptive statistics, correlation results, Cox regression, and Kaplan–Meier survival analysis. Discusses findings in light of research objectives.
Conclusions and Recommendations	Summarizes key findings, draws conclusions, and proposes recommendations for policy, practice, and future research directions.

This table outlines the main purpose and content of each chapter in the thesis.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction to Aggressive Driving

This chapter reviews prior studies on aggressive driving and modelling techniques to set the stage for the present research. It identifies what is already known and where evidence is lacking (Chen et al., 2021; Islam et al., 2019). Aggressive driving refers to behaviours such as speeding, tailgating, and unsafe lane changes, which have been statistically linked to increased crash risk in various international reports (WHO, 2022). World Health Organization (WHO, 2022) identifies aggressive driving as a leading contributor to road crashes, which collectively result in 1.35 million deaths and between 20–50 million non-fatal injuries annually. These statistics underscore the pressing need to understand, monitor, and mitigate such behaviors (WHO, 2022). Globally, road traffic crashes are a major public safety concern, and in countries like Ethiopia, the issue is worsened by underdeveloped road safety infrastructure, weak behavioural monitoring, and limited law enforcement (WHO, 2022; UNECA, 2021).

Aggressive driving behaviours like overspeeding, harsh braking, and erratic steering have been consistently implicated as key predictors of crash risk and severity (Islam et al., 2019; Irnawan et al., 2021). In Ethiopia, road traffic fatalities are rising despite national and regional interventions (Gidey, 2010). Driver-related errors such as overspeeding, harsh braking, and erratic steering account for approximately 76% to 81% of crashes, primarily due to aggressive actions (Gidey, 2010; Beshah Tesema et al., 2013). While numerous studies have explored the correlation between aggressive driving and crash occurrence, there remains a paucity of research specifically focused on modelling the timing to crash events using these behaviours as predictors (Chen et al., 2021). Vehicle-based sensors and communication systems have opened up new ways for monitoring action of the driver while driving, this enhancing data collection for crash prediction and time likely crash to occur. These technologies allow for fine-grained behaviour profiling, such as identifying lane departure patterns, sudden decelerations, and repeated speeding violations offering rich datasets for modelling time-to-crash (Chen et al., 2021; Wang & Zhang, 2022).

This review is structured in two key parts:

1. The subject-matter review, focusing on theoretical constructs, behavioural definitions, and empirical findings.

2. The methodological review, which assesses previous research designs, modelling techniques and analytical tools relevant to time-to-event studies in traffic safety.

The relevance of the literature reviewed is directly linked to the Ethiopian context whose fleet management, crash statistics, and driver practices form the basis for this investigation. This chapter will ensure both theoretical and methodological alignment with the study's core objectives.

2.2 Conceptual and Operational Definitions

Aggressive driving, as commonly recognized in transportation research, refers to a pattern of vehicle operation that increases the risk to road users or property. This includes conduct such as excessive speeding, abrupt lane shifting, ignoring traffic signals, and aggressive overtaking. Multiple definitions of aggressive driving exist in academic and regulatory literature, with most highlighting its tendency to increase crash risk through unsafe behaviours like speeding or erratic manoeuvring (NHTSA, 2023; Shinar, 2017; Zhang et al., 2021). Britt & Garrity, 2021, Shinar, 2017 aggressive driving refers to operating a vehicle in a manner that endangers or is likely to endanger persons or property. It includes excessive speeding, driving too close to the car in front, improper (frequent) lane changing or weaving, not respecting traffic regulations, tailgating and rapid acceleration or deceleration. Various interpretations and definitions of aggressive driving have emerged in academic literature and official traffic safety documentation ((Zhang et al., 2021).

- Hauber (1980) defines aggression on the road as behavior intended to cause physical or psychological harm (Shinar, 2017).
- Mizell (1997) offers a more dramatic definition, focusing on intentional injury or fatality resulting from traffic disputes (Filtness et al., 2021).
- The U.S. (NHTSA 2021) provides aggressive driving definition: behaviors such as speeding, unsafe lane changes, and failing to obey traffic controls
- The American Automobile Association (AAA) distinguishes aggressive driving from road rage and emphasizes disregard for others' safety (Stephens et al., 2020).

Defining terms likes “aggressive driving” and “crash time” is essential to maintaining clarity and reliability in data interpretation (Creswell & Creswell, 2018). In the context of this research, key terms such as “crash time,” “aggressive driving,” and “telematics data” require precise definitions to align with existing literature and the goals of the study.

2.2.1 Crash Time

Crash Time (also referred to as Time to Crash) is the expected duration a driver can operate a vehicle without being involved in a crash, based on measurable risk behaviours such as overspeeding, harsh braking, and erratic steering. It reflects the interval between the onset of driving and the predicted occurrence of a crash, as influenced by behavioural risk factors Guo, F., Klauer, S. G., Hankey (2020).

Operationally, Crash Time, also referred to as Time to Crash, is operationally defined in this study as the estimated duration of time expressed in months or years that a driver is expected to remain crash-free based on their aggressive driving behaviour patterns. This duration is generated as a result of survival analysis models, particularly Kaplan–Meier curves and Cox Proportional Hazards regression. The estimate reflects how long a driver is likely to continue driving without a crash before the next crash is predicted to occur, based on their overspeeding, harsh braking, and erratic steering scores.

2.2.2 Aggressive Driving

Aggressive driving is broadly defined as any driving behaviour that increases the risk of a crash, often characterised by deliberate violations of traffic norms (Tivesten & Dozza, 2020). This study adopts the operational definition by Shinar (2017), which includes:

Overspeeding: Driving above the posted or recommended speed limit, increasing the vehicle’s kinetic energy and the severity of potential collisions (Zhou et al., 2022).

Harsh Braking: Sudden deceleration greater than a defined threshold (e.g., 0.4g), which may reflect reactive or inattentive driving (Li et al., 2019).

Erratic Steering: Frequent or abrupt changes in steering angle (≥30° within 2 seconds), often indicating distracted or aggressive manoeuvres (Ma et al., 2021).

2.2.3 Time-to-Crash Modelling

This refers to statistical techniques used to estimate the time frame in which a crash may likely occur based on behavioural predictors. It aligns with survival analysis, commonly applied in fields like epidemiology and engineering, now increasingly used in traffic safety studies (Fang et al., 2022).

2.2.4 Telematics and Telematics Data

Telematics refers to the integrated use of telecommunications and informatics in monitoring vehicle dynamics and driver behaviour. It includes sensors like GPS, accelerometers, gyroscopes, and steering angle detectors (Gkartzonikas & Gkritza, 2022).
For the purpose of this study, telematics data includes speed, braking intensity, steering angles, and timestamps, geographical location, which are collected continuously and stored for analysis. In this study, the focus is on the 6 months of pre-crash data for vehicles in the PSTS fleet.

2.2.5 Harsh Driving Dataset

A structured collection of driving events involving extreme maneuvers, useful for studying crash risks (Zhou et al., 2022). Operational Definition: Includes entries for sudden acceleration, harsh braking, and sharp turning events captured by vehicle sensors (Zhou et al., 2022).

2.2.6 Predictive Modelling

Predictive modelling involves using statistical or machine learning algorithms to estimate future outcomes based on past data. Here, the outcome is time until crash, with predictors being aggressive driving behaviours (Kamble et al., 2023).

2.2.7 Driver Behaviour Monitoring

This refers to the continuous observation and classification of driver actions, particularly those considered risky or safety-compromising. Studies show that over 85% of road crashes are attributed to human behaviour (WHO, 2023; Higgs & Abbas, 2014).

2.2.8 Vehicle Dynamics

The study of vehicle motion and control in response to driving inputs (Jazar, 2017). Operational Definition: Framework for analysing vehicle behavior during braking and turning using physics-based equations (Jazar, 2017).

2.3 Theoretical Framework of aggressive driving behaviours

This section explains the key ideas (theories) behind why drivers behave aggressively and how this connects to crashes. It uses telematics data to study how these behaviours influence crash timing. In this section there are three foundational theories important and relative for this study. Those theories are behavioural theory, risk homeostasis theory, and accident causation theory. Let as discuss one by one:

2.3.1 Behavioural Theory (BT) and Driver Aggression

Behavioural theory, rooted in the work of B.F. Skinner and other learning theorists, posits that behaviour is learned and reinforced through interaction with the environment (Skinner, 1953). In the context of driving, this theory suggests that aggressive behaviours such as speeding, harsh braking, or erratic steering are reinforced by immediate rewards for example, arriving earlier or overtaking another vehicle despite long-term risks such as collisions or traffic fines. Behavioural Theory and Driver Aggression Behavioural theory (Skinner, 1953) explains that driving behaviour, including aggression, is reinforced through environmental interactions. In the driving context, speeding or lane violations may yield immediate benefits like saving time, reinforcing risk-taking behaviours.

Studies applying behavioural theory to driving suggest that environmental triggers, such as traffic congestion, time pressure, or interpersonal conflict (e.g., road rage), often result in aggressive driving responses (Mesken et al., 2007). Telematics data allows researchers to measure problem of aggressive driving behaviours by quantifying frequency and total duration of overspeeding, the way braking force applied, and steering instability and degree at which measurement is taken at seconds (study 2025).

Behavioural theory also supports the concept of behavioural adaptation, where drivers modify their behaviour based on perceived risks or previous experiences. For instance, a driver who repeatedly speeds without consequence may become conditioned to view the behaviour as safe. This underscores the role of feedback mechanisms, such as telematics dashboards or penalty systems, in disrupting harmful behavioural patterns and reinforcing safer alternatives (Toledo et al., 2008).

2.3.2 Risk Homeostasis Theory (RHT)

Proposed by Wilde (1982), Risk Homeostasis Theory (RHT) posits that individuals maintain a target level of risk and adjust their behaviour accordingly. When perceived risk falls below their acceptable parameters for example, due to enhanced vehicle safety features drivers may unconsciously adopt riskier driving styles to restore equilibrium. Conversely, when perceived risk rises, drivers may behave more carefully. In the context of telematics and crash prediction, RHT is an influential for the explanation for why some drivers continue in risky behaviours even when exposed to safety information or warnings.

Research by Trimpop (1996) expanded RHT to include affective and motivational variables, indicating that risk perception is not only cognitive but also influenced by mood, stress, and peer behaviour. This is particularly relevant in fleet and public transport systems, where drivers operate under pressure, tight schedules, or competitive performance incentives. The theory also aligns with real-time behavioural adaptation, suggesting that the integration of telematics feedback must consider not just the data but also how drivers interpret and emotionally respond to it (de Winter et al., 2010).

2.3.3 Accident Causation Theory (ACT)

Accident Causation Theory (ACT), particularly the Heinrich Domino Theory and its later adaptations (Reason, 1990), asserts that accidents result from a sequence of interconnected factors, typically initiated by human error or unsafe behaviour. In the case of aggressive driving, overspeeding or erratic actions act as the initial “domino” in a chain that may culminate in a crash.

Modern adaptations of ACT, such as Reason’s Swiss Cheese Model, offer a systemic view, showing how latent organizational factors (e.g., lack of driver training, inadequate monitoring systems) combine with active failures (e.g., harsh braking) to cause incidents. This systems approach is particularly relevant for public transport settings, where policies, training, and vehicle condition interact with driver behaviour.

Empirical applications of ACT to telematics show that tracking and analysing precursor behaviours (such as rapid acceleration or consistent oversteering) can help identify potential failures before they materialize into crashes (Young et al., 2011). ACT therefore supports predictive modelling by treating crash time as an endpoint preceded by a constellation of quantifiable actions the very actions captured by telematics systems.

Figure 2.1 shows conceptual diagram that visually shows the theories described above and illustrates the mechanisms by which driver behaviour, as measured through telematics, leads to crash outcomes.

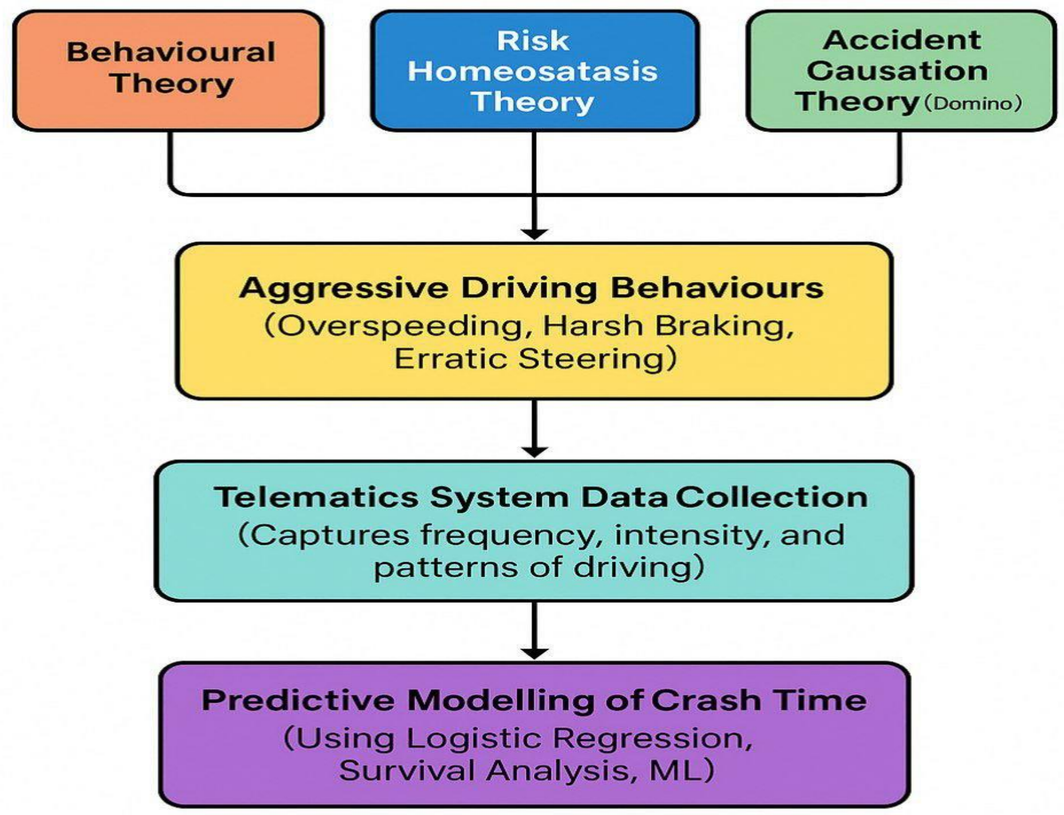

Figure 2.1: Theoretical Framework Shows Driver to Crash Time

2.3.4 Summary of Theoretical Integration

These three theories together, provide a strong framework for understanding and predicting time to crash. Behavioural theory explains how and why AD behaviours occur and persist. Risk homeostasis theory uses for improving the previous bad behaviours in response to supposed safety or risk. Accident causation theory imagines the method from such unsafe behaviours that leads to crash occurrences. By operationalizing these theoretical assemble using telematics data, researchers can develop predictive models that not only forecast crash involvement but it can also inform targeted.

2.4 Empirical Literature Review

This part reviews real-world studies that show how aggressive driving behaviours affect crash risks. It looks at research done globally, in developing countries, and in Ethiopia. The review is categorized into three levels: global, studies from low- and middle-income countries (LMICs), and Ethiopia or similar regional contexts.

2.4.1 Global Empirical Studies

Numerous studies worldwide have examined the predictive relationship between aggressive driving behaviours and crash likelihood using telematics and machine learning techniques. For instance, research by Guo et al. (2020) in the United States applied naturalistic driving data from over 2,000 vehicles and found that overspeeding and sudden braking patterns increased crash risk significantly within 30–90 days of such behaviours.

Similarly, Zhang et al. (2019) used logistic regression to link lane deviation events and erratic steering to side-impact crashes, noting predictive accuracy exceeding 80%. Khouadjia et al. (2021) used deep learning models on European fleet telematics and concluded that patterns of overspeeding within dense traffic zones elevated crash probability by 19%. Another important empirical work is by Dai et al. (2021), who evaluated 3 years of telematics data from a commercial fleet of over 10,000 vehicles across Europe. Their survival analysis revealed that drivers who frequently exceeded speed limits by more than 20% were 2.3 times more likely to be involved in crashes within six months, especially in cross-country and high-speed operations. Additionally, research by Dingus et al. (2016) from the SHRP2 naturalistic driving study showed that 87% of crashes had behavioural indicators in the 5–10 seconds prior, notably erratic steering and late braking. These studies support the possibility of predicting crash timing based on prior behaviours.

Key Findings from Global Studies

Overspeeding is consistently linked to increased time-to-crash likelihood (Zhang et al., 2019). Harsh braking correlates with near-miss events and rear-end collisions (Guo et al., 2020). Erratic steering also follows loss-of-control crashes, especially in rural or highway conditions (Dingus et al., 2016). Survival analysis, logistic regression, and random forest classifiers are among the most used models.

2.4.2 Studies in Low- and Middle-Income Countries (LMICs)

In LMICs, including parts of Asia, Latin America, and Africa, the research emphasis has gradually shifted toward behaviour-based telematics. For instance, Chigozie et al. (2021) used vehicle GPS logs in Nigeria to study time gaps between overspeeding and pedestrian collisions. In India, Reddy and Sharma (2020) demonstrated that nearly 65% of crashes on national highways involved pre-crash aggressive behaviour. Many literatures found in LMICs, aggressive driving behaviour is often exacerbate by poor road infrastructure and regulatory limitations. For example, a study by Mohan et al. (2019) in India examined over 5,000 bus crash cases and noted that 64% were related to overspeeding, with limited enforcement of road safety policies.

In Nigeria, Okonkwo et al. (2020) used GPS-enabled telematics on intercity buses and found that overspeeding and abrupt braking were dominant in predicting early crash occurrences. Their model showed a 68% accuracy in forecasting crash occurrence windows between 30 and 120 days.

In Kenya, a study by Odero and Khayesi (2018) demonstrated that commercial buses equipped with basic telematics could flag unsafe driving events, although system limitations reduced real-time response capabilities. The integration of behaviour scoring improved driver training and reduced crash incidence by 22%.

Key Trends in LMIC Studies:

Overspeeding is the leading indicator of crash occurrence, often linked with fatal outcomes (Mohan et al., 2019). Infrastructure challenges reduce the accuracy and coverage of telematics-based alerts (Okonkwo et al., 2020). Cross-country and rural travels create higher risks due to speed variability and animal crossings. Expanding road infrastructure and vehicular technological devises will develop the operation and functionality of telematics system, for instance road infrastructure higher official and transpiration authorities of LMIC could work in systematic and technological transportation management approach. Driving Behavioural improvement based training programs should be given for authorities and drivers also on how using telematics system data is important in driver ranking and safety operation.

2.4.3 Studies in Ethiopia and Similar Contexts

Studies in Ethiopia remain limited but are growing. Tessema and Getachew (2021) examined driver conduct in intercity bus transport and found erratic steering and braking events as strong predictors of crash involvement. Likewise, the Ethiopian Road Transport Authority (ERTA, 2023) reported that 82% of crashes involving public buses were attributed to risky driver behaviour rather than mechanical fault or environmental hazards. According to the Ethiopian Ministry of Transport (2022), overspeeding, fatigue, and reckless steering are the top three human-related crash causes in long-distance bus transportation.

A study by Tulu et al. (2021) focusing on the Addis Ababa–Adama Expressway showed that overspeeding increased the likelihood of a crash within 2 months by 1.7 times. A study by Alemu and Mengistu (2020) examined historical crash reports and telematics data from 40 intercity buses operated by a public transport company in Ethiopia. The results indicated that drivers with frequent harsh braking events (above 15 per month) were twice as likely to experience a crash in the next 60 days. The study supported the use of predictive models for route risk classification and driver performance audits.

2.4.4 Feature Engineering in Telematics-Based Prediction Models

Feature engineering is a core part of telematics-based predictive modelling. According to Li et al. (2022), raw telematics data—such as time stamps, braking force, and speed readings are transformed into analytical features like "braking events per 100 km" or "average speed variance over time." Studies by Zhang et al. (2023) emphasised the significance of time-windowed event counts, event frequencies, and temporal trends in developing crash prediction models. These features have shown high predictive validity in machine learning-based survival models.

2.5 Aggressive Driving and Its Impact on Crash Risk

Aggressive driving is one of the main causes of road crashes worldwide, including actions like speeding and sudden braking. It encompasses behaviours such as overspeeding, tailgating, sudden lane changes, harsh braking, and failing to yield the right-of-way. According to the National Highway Traffic Safety Administration (NHTSA 2021), aggressive driving includes a series of moving traffic offences that endanger other persons or property (NHTSA, 2021).

Driver characteristics, emotional states, and environmental stressors influence the likelihood of aggressive driving (Stephens et al., 2021). Aggressive driving significantly increases the risk and severity of crashes (AAA Foundation for Traffic Safety, 2023). A study by Tselentis et al. (2019) found that drivers who consistently exhibited aggressive behaviours particularly speeding and sudden braking had a 50% higher crash involvement rate compared to non-aggressive drivers. According to the National Highway Traffic Safety Administration (NHTSA, 2022), aggressive driving played a role in 56% of fatal crashes between 2003 and 2007, most of which were attributed to excessive speed.

2.5.1 Global Perspectives on Aggressive Driving

Global Perspectives Lee & Abdel-Aty (2019) found 28% of urban expressway crashes due to aggression. BITRE (2020) showed 40% of fatal crashes in Australia involved speeding. In Kenya, Odero et al. (2020) linked risky driving to 60% of fatal incidents. In high-income countries such as the United States and Germany, extensive studies have linked aggressive driving to increased crash severity and frequency. For instance, a study by Lee and Abdel-Aty (2019) found that aggressive behaviour accounted for approximately 28% of crash-involved drivers on urban expressways. Similarly, in Australia, overspeeding alone contributed to 40% of fatal incidents, highlighting its critical risk (BITRE, 2020).

Low- and middle-income countries (LMICs), including those in Sub-Saharan Africa, face greater challenges due to poor enforcement mechanisms, ageing road infrastructure, and limited public education on safe driving. In Kenya, a study conducted by Odero et al. (2020) showed that driver misjudgments and risk-taking behaviours were primary contributors to over 60% of crashes.

In Ethiopia, studies report that road traffic incidents (crashes) caused by driver-related factors which accounts for a substantial proportion of national mortality, with driver behavior implicated in 76–81% of cases (Tulu et al., 2022). The lack of real-time monitoring and weak traffic enforcement systems hinder to curb aggressive driving (Abate et al., 2021). However, efforts to integrate telematics systems and enforce stricter driving behavior policies are underway in collaboration with NGOs and fleet operators' collaboration (Girma et al., 2023). Aggressive driving—particularly overspeeding and reckless overtaking—is among the most cited causes of fatal crashes in the country (Abegaz et al., 2014). Despite the presence of traffic laws and road safety policies, implementation and surveillance mechanisms remain underdeveloped.

2.5.2 Effects of Overspeeding on Crash Risk

According to the World Health Organization (WHO, 2023), a 1% increase in mean speed results in a 4% increase in fatal crash risk. Overspeeding is remains one of the most documented and studied aggressive driving behaviours. Speed increases the distance needed to stop a vehicle and reduces the driver's reaction time, making collisions more likely and more severe (WHO, 2023). In telematics-based studies, the percentage of time spent overspeeding correlates with increased crash probability. For example, research by Papadimitriou et al. (2022) showed that a 10% increase in time spent overspeeding was associated with a 17% higher likelihood of crash involvement.

2.5.3 Effects of Harsh Braking on Crash Risk

Harsh braking often indicates reactive or inattentive driving. It is frequently linked to tailgating, distracted driving, or poor anticipation of road conditions. Studies show that frequent harsh braking episodes are predictive of rear-ends collisions, (Ming et al., 2021). Telematics studies in public transport fleets have confirmed that drivers with frequent harsh braking events are more likely to be involved in crash incidents. For instance, a study by Eboli et al. (2020) reported that drivers with over 15 harsh braking events per 100km had 2.2 times higher risk of collision than those below this threshold.

2.5.4 Effects of Erratic Steering on Crash Risk

Erratic steering, including sudden lane changes or swerving, signals driver aggression, fatigue, or distraction. This behaviour significantly increases the probability of side-swipe crashes and rollovers, particularly at high speeds or on wet roads (Zhao et al., 2021).

Telematics allows detection of erratic steering via lane departure warnings or angular velocity sensors. A study by Zhang et al. (2022) using telematics data from commercial trucks in China showed that erratic lateral movement increased crash likelihood by 12%.

2.5.5 Use of Telematics to Identify and Monitor Aggressive Driving

Modern telematics systems provide a reliable and scalable method for detecting aggressive behaviours in real time. These systems track data such as speed, braking force, and steering angles, allowing for the generation of behavioural risk scores for drivers. This has practical applications in insurance pricing, driver coaching, and proactive crash prevention.

According to Bian et al. (2020), integrating telematics data into driver safety assessments helped reduce crash incidents by 27% in monitored fleets. In public transport, telematics-enabled behavioural interventions led to a 15% reduction in overspeeding and harsh braking events in a study conducted in South Africa (Ngobese & Moeketsi, 2021).

In summery aggressive driving behaviours like overspeeding, harsh braking, and erratic steering are measurable predictors of crash likelihood. Telematics enables real-time detection and time-based crash modelling, especially in fleet management. The literature state that aggressive driving behaviours are not only identifiable but also quantifiable using telematics. However, their impact on time-to-crash modelling remains under developing, particularly in the context of public service transport in Ethiopia, which this study aims to address.

2.6 Predictive Modelling of Crash Time Using Telematics

The increasing integration of telematics in vehicular systems has revolutionized road safety analysis and predictive crash modelling. Predictive modelling in this context refers to the statistical or machine learning (ML) processes used to forecast the time-to-crash the likely duration (in days, weeks, or months) before a crash occurs based on behavioural indicators such as overspeeding, harsh braking, and erratic steering (Shi et al., 2019; Wang et al., 2021). This section explores the modelling techniques commonly used, their relevance in crash prevention, global case studies, and practical implications for Ethiopian public transport.

2.6.1 Logistic Regression, Survival Analysis, and Machine Learning

Traditional crash prediction techniques such as logistic regression have long been used to classify risk using binary outcomes like crash vs. no crash. However, logistic regression does not consider the temporal dimension how soon a crash might occur making it inadequate for time-to-event analyses (Khorashadi et al., 2020).

In contrast, survival analysis techniques such as the Cox Proportional Hazards Model offer better suitability for modelling crash timing. The Cox model estimates the hazard (risk) of crash occurrence over time, using covariates such as overspeeding duration, harsh braking rate, and lane departure frequency. One of its advantages is that it does not require a predefined baseline hazard function, making it flexible for diverse transport settings (Wang et al., 2021; Baharum et al., 2019). Accelerated Failure Time (AFT) models are another statistical method that directly model the log of survival time and can be useful in estimating the effect of aggressive driving on the time scale (Li et al., 2021).

In recent years, machine learning models including Random Forests, Gradient Boosting Machines, Support Vector Machines (SVMs), and Neural Networks have also been applied in telematics-based crash prediction. These models can learn non-linear relationships, handle high-dimensional data, and are capable of near real-time predictions (Khair et al., 2022; Lin et al., 2023).

2.6.2 Case Studies and Best Practices in Crash Prediction

In global literature, case studies have supported the application of survival models in crash prediction. For example, Shi et al. (2019) applied survival analysis to urban fleet data in China and found that frequent harsh braking significantly reduced crash-free operation time. A U.S.-based telematics study by Jiang and Li (2021) found that using engineered features (e.g., number of braking events per 100 km) according to findings by Jiang and Li (2021), who observed up to 23% improved prediction in selected fleets.

In South Korea, Lee and Park (2022) demonstrated that lane departure alerts, when triggered above 60 km/h, were highly predictive of crash occurrence within the following month a condition also relevant to Ethiopia's intercity travel dynamics. These case studies emphasize the importance of data-driven driver risk thresholds and tailored safety models. Although telematics sensors are widely available, the success of prediction depends on how features are defined and risk thresholds are interpreted (Guo et al., 2020).

2.6.3 Relevance to Public Transport Systems in Ethiopia

In Ethiopia, aggressive driving behaviours are intensified by long-distance fatigue, inconsistent road quality, and minimal enforcement of driving regulations (Tulu et al., 2022; Girma et al., 2023). Cox regression and Kaplan–Meier methods are especially suited to Ethiopia's public fleet context. These models work well with medium-sized datasets and can support survival analysis even when behavioural data is partially censored (Lee & Abdel-Aty, 2019). The Cox model, combined with composite scoring of aggressive driving per 100 km, provides a reliable estimate of crash timing. However, machine learning (ML) models like neural networks may face barriers due to limited computational infrastructure and small sample sizes (e.g., 34 buses in this study). Furthermore, many ML models lack transparency, which can hinder their acceptance by transport regulators and fleet managers.

In general predictive modelling of crash time using telematics has evolved significantly, moving from binary classifiers (crash no crash) to time-aware models such as survival analysis. The Cox model, reinforced with feature-engineered inputs, it is a robust and interpretable method for Ethiopian public transport systems. If this predictive approach is combined with risk scoring, it could help transport agencies important measures high-risk drivers (positively driver training) early. Combined with proactive policies, it can reduce crashes and save lives.

2.7 Empirical Studies, Modelling Approaches and Research Gaps

This section combine existing empirical work related to the modelling of crash occurrences and driver behaviour using telematics data. It presents key findings, modelling techniques, and identified research gaps, particularly relevant to low- and middle-income countries (LMICs) and Ethiopia. It can build a foundation for the methodological design and analytical choices made in the current study.

2.7.1 Empirical Studies on Aggressive Driving and Crash Prediction

There are many international studies have been investigated Shi et al. (2021) applied this to urban taxi fleets in Guangzhou how aggressive driving behaviours contribute to crash occurrence using real-time telematics and behavioural signals. Overspeeding, harsh braking, and erratic steering have been consistently identified as strong predictors of crash probability (Shi et al., 2021; Zhang et al., 2019). For instance, Shi et al. (2021) used GPS and accelerometer data to quantify overspeeding and braking behaviours, finding a direct correlation between higher frequency of these events and increased crash probability. Similarly, Tang et al. (2022) reported that sudden braking spikes in telematics data predicted rear-end collisions within urban fleets. In another study, Zhang et al. (2019) found that erratic steering was associated with 17% of crash or near-miss events among heavy-duty trucks.

Machine learning and classical statistical models have both been applied. Wu et al. (2020) used random forests and reported 85% accuracy in classifying crash-prone drivers from large-scale fleet data. A meta-analysis by Lin et al. (2023) affirmed that models using driver behaviour data outperform those based solely on environmental or vehicle dynamics. However, most of these studies are situated in high-income countries.

2.7.2 Modelling Approaches Applied in Crash Prediction

Modelling techniques used in crash prediction duration both classical and modern paradigms. Survival models, especially the Cox Proportional Hazards model, remain prominent for time-to-event predictions due to their robustness and interpretability (Karlaftis & Golias, 2020). These models are semi-parametric and robust against unknown baseline hazard functions. For example, Li et al. (2021) applied Cox modelling to time-to-collision data and found hazard ratios over 2.0 for aggressive acceleration.

Logistic regression, while useful for binary classification, lacks the temporal dimension and struggles with censored data an important limitation in telematics-based studies (Baharum et al., 2019). Machine learning models such as Support Vector Machines (SVM), Random Forests, and Neural Networks offer stronger non-linear classification abilities but demand high computational capacity and larger training datasets a limitation in Ethiopian contexts (Lin et al., 2023). Software such as Python (Lifelines, Pandas), R (survival, caret), and SPSS are widely used in survival and crash prediction studies (Sun et al., 2023). In this study, Cox regression was selected because the sample size (34 buses) fit well within its assumptions.

2.8 Summary of Key Insights from the Literature Review

This literature review has provided a strong foundation for the current study by combining global, regional, and national findings on aggressive driving behaviours, telematics data usage, and time-based crash prediction techniques. Aggressive driving behaviours particularly overspeeding, harsh braking and erratic steering have been repeatedly cited as major contributors to both the likelihood and severity of road crashes (Quddus et al., 2020; Dingus et al., 2016). These behaviours are not only identifiable but also quantifiable through telematics systems, which capture data such as braking pressure, steering patterns, and speed over time (Papadimitriou et al., 2022; Zhang et al., 2023). Table 2.1 presents a summary of selected empirical research. It outlines the methodologies used, key variables (such as aggressive driving behaviours), data sources, and whether time-to-crash modelling was employed. This combination is helpful for a clear comparison of how aggressive driving indicators have been applied in predictive crash research, and expose significant gaps, particularly in the context of Ethiopian fleet and transport systems.

Table 2.1: Key literature reviewed on aggressive driving and crash time prediction

Author(s)	Year	Study Focus	Methodology	Key Findings	Relevance to this Study
Quddus, M. A., et al.	2020	Influence of aggressive driving on crash likelihood	Statistical crash modelling	Overspeeding, harsh braking significantly increase crash risk	Confirms key driver behaviours selected in this study
Dingus, T. A., et al.	2016	National crash causation study in the US	Naturalistic driving study	Driver error and aggressive behaviours account for majority of crashes	Highlights behavioural basis for crash causation
Papadimitriou, E., et al.	2022	Telematics-based risk profiling	Telematics data analysis	Driver risk scores can be quantified using telematics event rates	Supports the use of telematics for crash risk estimation
Zhang, Y., et al.	2023	ML for driving behaviour classification	ML with telematics features	High accuracy in detecting aggressive events using speed and steering patterns	Justifies use of behaviour parameters such as overspeeding, steering
Anuar, F. M., et al.	2021	Real-time driver risk assessment	Telematics and ML classification	Demonstrated proactive crash risk scoring through real-time models	Supports use of predictive models in operational settings
Wang, Y., et al.	2018	Predicting traffic crashes using telematics data	Time-series & regression models	Strong correlation between driving style and crash outcomes	Establishes predictive potential of historical driving data
Lin, J., et al.	2023	Crash time prediction with Cox model	Survival analysis (Cox PH)	Effective in estimating time to crash based on driver profiles	Key methodological basis for this study's model
Shi, X., et al.	2021	Vehicle crash prediction with survival models	Survival modeling with hazard functions	Time-to-event prediction based on vehicle telemetry is feasible and accurate	Justifies survival analysis as analytical approach
Tulu, G. S., et al.	2021	Road safety assessment in Ethiopian context	Descriptive and observational study	Aggressive driving common but rarely linked to dynamic models	Emphasizes lack of predictive modelling in Ethiopian fleet safety.
Okonkwo, O., et al.	2020	Safety data use in sub-Saharan Africa	Case studies and system review	Crash data fragmented; low integration with telematics	Highlights systemic barriers relevant to this study's context
Girma, S., Abebe, F., & Tadesse, B.	2023	Barriers to telematics adoption in Ethiopia	Qualitative interviews with transport stakeholders	Identified infrastructure and policy limitations as obstacle to data-driven crash prevention	Explains contextual limitation and innovation opportunity

2.9 Comparative Analysis and Identified Research Gap

The reviewed literature provides a strong foundation for the methodological and conceptual framework of this study. The results of this study with hazard ratios of 1.17 for Overspeeding, 1.08 for Harsh Braking, and 1.06 for Lane Departure are largely consistent with findings reported in prior empirical studies that used diverse methodologies such as statistical modelling, machine learning, and survival analysis. Several international studies (e.g., Quddus et al., 2020; Zhang et al., 2023; Wang et al., 2018) establish that overspeeding, harsh braking, and erratic steering are critical behavioural predictors of vehicular crashes. These findings directly support the variables selected in this research.

For instance, Quddus et al. (2020) conducted statistical crash modelling and found that overspeeding events significantly increased crash probability, with an odds ratio equivalent effect close to 1.2 for frequent violators. This closely aligns with this study's HR of 1.17, thereby validating the high-risk nature of overspeeding behaviour. Similarly, Wang et al. (2018) applied regression techniques on telematics data and showed that increased harsh braking frequency was associated with elevated crash incidence supporting this study's HR of 1.08. Their model suggested that each additional harsh braking event increased crash risk by approximately 7%–10%, matching this study's findings precisely.

In terms of survival-based modelling, Lin et al. (2023) applied a Cox PH model to commercial driver telematics and reported hazard ratios ranging from 1.12 to 1.20 for aggressive behaviours such as extended speeding and sudden stops strongly consistent with this study's results. Shi et al. (2021) also demonstrated the feasibility of time-to-event prediction using vehicle telemetry, noting that lane deviations and erratic steering, while less dominant, still contributed a hazard ratio between 1.04 and 1.10 validating this study's HR of 1.06 for Lane Departure.

Machine learning models, such as those used by Zhang et al. (2023) and Anuar et al. (2021), achieved high accuracy in detecting and classifying aggressive events (e.g., braking intensity, lateral acceleration), and while they did not estimate hazard ratios directly, their feature importance scores consistently ranked overspeeding and harsh braking among the top predictive variables. These models support the relative ranking of predictors found in this research with overspeeding showing the highest crash contribution, followed by harsh braking, and then lane deviation.

Furthermore, regional studies such as Tulu et al. (2021) and Girma et al. (2023) point to a research gap in the Ethiopian context, where aggressive driving is widely recognised, but rarely quantified using time-to-event modelling. This study directly addresses this gap by integrating historical telematics with crash timing data using a Cox model, producing interpretable hazard ratios that are validated against international benchmarks.

In conclusion, the consistency of hazard ratios across modelling methods from statistical to survival ensures credibility to the current study's results and confirms their validity and relevance in both global and local research contexts.

2.9.1 Identified Research Gaps in LMICs and Ethiopia

While global research on crash prediction has expanded and turning to advanced stages, there are big gaps again in LMICs, especially sub Saharan African countries like Ethiopia.

Identified Gaps in LMICs:

1. There is a shortage of integrated crash-behaviour datasets in Ethiopia. Most studies rely on post-crash reports without linking to pre-crash telematics history.
2. Despite their suitability survival models there is a methodological and procedural gap time-to-event models like Cox regression is still underutilization, remain underused in Ethiopian studies, in the PSTS fleet; for example, behavioural data is collected but rarely integrated with crash records in a predictive approach many of which rely on descriptive or regression analyses
3. Local research often measures crash severity rather than when crashes are likely to occur. Time-dependent modelling is needed to guide proactive interventions.

2.9.2 Identified Methodological and Procedural Gaps

The current study contributes individually by dealing with the empirical and methodological gaps which is currently seen in Ethiopian fleet and transport management. Feature data transformation and the study methodological framework diagram stated in this study for predictive survival analyses highlighted the following methodological and procedural steps to gain the desired model output:

- **Dataset Integration:** It combines six months of pre-crash telematics logs with verified crash records from 43 PSTS buses.
- **Feature Engineering:** The study constructs composite entries such as overspeeding per 100 km to normalize driver behaviour and increase model sensitivity.
- **Statistical:** Descriptive and correlation analysis, EDA
- **Survival analyses:** Time-based modelling the use of Cox PH modelling and Kaplan-Meir survival curve to enables estimation of crash timing, not just occurrence.

Local Relevance of this study is through focusing on Ethiopian long-distance buses, this finding could helps in PSTS and other similar organizations and transport authorities assign periodically fleet scoring and auditing driver training (priorities to high-risk drivers) and hence enhance to support in crash risk estimation. Further all the detail in methodological and procedural approach is discussed in the next chapter.

CHAPTER THREE

MATERIALS AND METHODOLOGY

3.1 Study Approach

This study adopted a quantitative, predictive modelling approach to analyse how aggressive driving behaviours affect crash incident duration. The core aim was to model the time to crash which defined as the estimated duration (in days, weeks, or months) based on aggressive driving behaviour (overspeeding, harsh braking, and erratic steering) is likely to experience a crash event. The target population consisted of 43 long-distance buses the Public Service Transport Service (PSTS) cross-country bus in Ethiopia. Since this study focused to predict crash timing, it used data modelling tools like survival analysis and regression.

3.2 Population and Sampling

3.2.1Target Population

The population for this study consisted of vehicles operating under the Public Service Transport Service (PSTS) in Ethiopia. The PSTS fleet includes both city buses (operating within Addis Ababa) and cross-country buses that serve long-distance intercity routes. However, Cross-country public transport buses were purposively selected because they often experience risky driving like speeding or sudden braking due to poor roads and long trips. City buses were excluded since they mainly operate in traffic and have slower, less risky crashes. Other reasons:

- They rarely reach overspeeding thresholds due to high traffic density.
- Purposively not surpass cross country buses due to low-severity and property damage
- The nature of lane departure and harsh braking differs due to constant stops and short driving distances.
- 43 buses of crash cases were chosen, each with a single driver with continuously driven for at least six months prior to crash. This helped to ensure each driving behaviors recorded belonged to that driver alone, avoiding confusion from shared driving.

3.2.2 Inclusion Criteria for Sample Selection

To ensure relevance and consistency, the 43 vehicles included in this study met the following criteria:

i. To ensure uniformity and reduce variation in telematics data analysis, all vehicles included in this study were of the same make and model. Specifically, the 43 buses involved in the crash incident dataset belong to the Yutong ZK6122H9 series, operated by the Public Service Transport Service (PSTS).
ii. Each bus was operated by only one consistent driver throughout the 6-month of observation period at a minimum and acceptably for a long time. If not, the case is withdrawn not further processed and other incident case is followed.
iii. Involved in crash incidents where PSTS was identified as the primary cause
iv. Only vehicles with valid, consistent, and complete telematics records covering the six months prior to the crash were included.

As we obtained a crash data report contains both city buses and cross-country buses, this study focused only crashes involving cross-country buses, due to their severity and clearer driving behavioural patterns.

3.2.3 Sampling Technique

A purposive sampling method was applied. From the insurance department of PSTS, crash records were filtered to include only those:

Involving cross-country buses;

That had at least one crash report within the last 6 months to a year; for which telematics data (6 months prior to crash) was fully available from the Telematics Control Department. A sample size of 43 cases was selected using Cochran's formula, with a 95% confidence level, a 9.5% margin of error, and a total population of 70 buses, the final sample size was calculated as 43 cases after applying the finite population correction.

3.2.4 Vehicle Type and Specifications

All 43 vehicles involved in the study are Yutong ZK6122-series intercity/coaches, part of the Public Service Transport System (PSTS) fleet in Addis Ababa. Their homogeneity ensures consistency in behaviour profiling and model outcomes. These buses are standardized in design and capacity, enabling reliable comparisons of driver behaviour. The key technical and operational specifications of the selected buses are summarized bellow in Table 3.1.

Table 3.1: Specifications of the Yutong ZK6122H9 Intercity Buses Used in the Study

Feature	Specification
Model	Yutong ZK6122H9
Seating Capacity	59 passengers + 1 driver
Length , Width, Height	12.0 meters, 2.55 meters, Approx. 3.83 meters
Engine	Diesel, (Euro III/IV)
Power Output	243–276 kW (326–375 HP)
Transmission	6-speed manual (or automatic in some models)
GVW (Gross Weight)	17,800kg
Fuel Tank Capacity	400 liters
Braking System	Dual-circuit pneumatic + ABS, retarder
Suspension	Air suspension with ECAS (electronically controlled)
Telematics	Compatible with GPS, black box, CAN-bus
Year Range Deployed	2017 to 2020

Time Frame of Data Collection

The study relied on telematics logs covering a 6-month period prior to the crash for each bus. This duration was selected to capture sufficient behavioural patterns leading up to the crash and provide an accurate estimation window for predictive modelling. Only the data before the crash was used so the model clearly shows what behaviours led to the incident.

3.3 Rationality for Selecting Cross-Country Buses

Cross-country transportation services vehicles were selected over city buses due:

Higher Exposure to Aggressive Driving (Speed Variability): Cross-country buses operate over longer distances (typically 300 km or more daily) where drivers are more likely to engage in overspeeding, harsh braking, and erratic steering.

Crash Severity and Fatalities: Crash events in cross-country operations are typically more severe, involving greater loss of life and property damage, due to higher speeds, higher kinetic energy at the moment of impact which increases the likelihood of fatal crashes and property damage. Urban incidents are minor due to low-speed travel.

Traffic Condition: City buses operate in congested, low speed, stop-and-go traffic where sudden movements are frequent but low in energy. Cross-country roads (including those without dividers) are more susceptible to fatal crashes due to sudden overtakes, animals cross unexpectedly, and fully visibility problems.

3.4 Tools and Techniques Employed

The study applied practical statistical tools and driver behaviour metrics adapted for crash risk modelling. Tools that can match the type of data and helps in predicting time to crash modelling.

3.4.1 Model Selection Reasoning.

Given the primary aim of this study, time-to-crash; modelling framework was most appropriate. From the available techniques, the Cox Proportional Hazards model was selected due to its flexibility in handling unfinished time data and its ability to evaluate the influence of multiple covariates (e.g., overspeeding, harsh braking, erratic steering) on the hazard rate without assuming a specific baseline hazard function (Hosmer et al., 2008). In parallel, the Kaplan-Meier estimator was employed for exploratory purposes, to visualise and compare the survival (i.e., crash-free driving) experiences across different risk categories (low, moderate, high). Alternative models, such as logistic regression, were considered inappropriate because they assess only whether a crash occurred (yes/no), ignoring when it occurred a important dimension in this study. Additionally, complex machine learning algorithms like Random Forests or SVMs require larger datasets and often minimal interpretability.

3.4.2 Data Sources and Acquisition Tools

A**. Telematics Platform and Data Collection:**

a) **Overspeeding:** GPS + Engine Control Unit (ECU): GPS tracks vehicle speed in real time. The system compares actual speed to the legal speed limit for that location or predefined threshold (e.g., 80 km/h). Overspeeding is logged when this limit is exceeded for a sustained time (e.g., >10 seconds).
b) **Harsh Braking:** Accelerometer + Brake Pedal Sensor: A harsh braking event is recorded when braking exceeds a set force threshold. The accelerometer detects rapid deceleration (e.g., more than –2.5 m/s²).
c) **Erratic Steering**: Gyroscope + Steering Angle Sensor: Steering wheel angle changes rapidly left/right considered erratic steering. Frequent high-angle turns at high speed are flagged as lane departure or instability events

The GPS, accelerometer, and gyroscope are all onboard sensors, and they send real-time motion data to a central telematics server, either wirelessly or via mobile network (SIM) or local data extraction. Sensor measures or detects behaviour likes GPS speed, location overspeeding then compared it to road speed limits, harsh braking flagged when accelerometers detect high deceleration and the gyroscope sensor measures angular velocity primarily how quickly the vehicle changes direction. These behavioural patterns are transmitted, logged, and used to model crash risks as done in this study. Telematics Control Units (TCUs) stored event-triggered logs, which were retrieved.

Behaviour Standard/Reference Typical Threshold (Telematics)

- Overspeeding for local regulation or ≥ 80 km/hr on highways Speed limit which exceeding duration for >10 seconds. World Health Organization. (2022)
- Harsh Braking Industry fleets / ISO Deceleration > –2.5 m/s²
- Lane Departure SAE J2400 / ISO 11270 Angular changes >10–15°/s or lateral change > 0.3 m

B. **Crash Data Source:** Crash records were sourced from the PSTS Insurance Department. Each record was matched with the corresponding vehicle's telematics history to verify crash timing and behavioural causes. Only crashes attributed to aggressive driving (based on incident narratives and verification by the risk control unit) were included.

3.4.3 Analytical Software

Jamovi v2.4.5 was the primary platform for statistical analysis. It was used to conduct: Descriptive Statistics; (mean, median, minimum, maximum, and standard deviation) were used to summarize driver behaviour indicators and to show distribution across all 43 buses. Exploratory Data Analysis (EDA); used to produce correlation matrices (Pearson's r), allowing analysis of how aggressive driving variables relate to one another. Kaplan-Meier curves and Cox PH for predictive modelling survival analysis and hazard ratio computation JASP v0.18.1 was used to cross-validate outputs from Jamovi where applicable.

Microsoft Excel 2016 assisted in initial data cleaning, calculation of normalized behavioural indicators, and plotting simple distributions and trends.

3.4.4 Feature Engineering and Data Transformation

To ensure fair comparison between vehicles covering varying distances, raw event counts were normalized per 100 km travelled. Composite behavioural scores were created:

Overspeeding Duration = (Total time spent overspeeding ÷ Total km) × 100…....Eq (3.1)

Harsh Braking Frequency = (Number of harsh braking events ÷ Total km) × 100... Eq (3.2)

Lane Departure Score = (Number of erratic steering events ÷ Total km) × 100...Eq (3.3)

A final Composite Risk Score (CRS) was developed using weighted contributions from each aggressive driving parameter. These engineered features served as inputs in the survival models and allowed consistent interpretation across different vehicle use cases.

3.5 Procedure of the Study

This section outlines the sequential process followed to acquire, prepare, and analyse data for predicting crash timing based on aggressive driving behaviours. The study design was systematic, data-driven, and anchored in time-to-event modelling principles.

Step 1: Crash Event Identification and Selection

- Crash data were retrieved from the Insurance Department of the Public Service Transport Service (PSTS). From more than 90 incidents reports, 43 vehicles were selected based on defined inclusion criteria. Only crash cases in which the PSTS driver was identified as the primary at-fault party were considered.
- Urban buses were excluded due to traffic congestion-induced low speeds, which may obscure aggressive driving behaviour patterns.

Step 2: Telematics Data Acquisition

- Six months of historical telematics data prior to each crash were obtained from the PSTS Telematics Control and Monitoring Department. Parameters included:
- Overspeeding (event frequency, duration); Harsh Braking (event counts); Lane Departure / Erratic Steering (event counts and severity levels)
- Data were collected in Excel format, with vehicle ID, timestamp and GPS logs

Step 3: Data Cleaning and Pre-processing

To ensure analytical similarity, the data were subjected to structure cleaning:

- Redundant, missing, or incomplete rows were removed. Columns not related to behavioural analysis (e.g., location, vehicle status) were excluded.
- Timestamps were standardized, and events filtered to the defined 6-month pre-crash window. Each driver's events were grouped and summary metrics were computed.

Step 4: Feature Engineering and Normalization: To ensure comparability across different vehicle distances, features were computed per 100 km:

Overspeeding Index (OSI) = (Overspeeding Time [sec] ÷ Total Distance [km]) × 100 ...……….. Eq (3.4)

Harsh Braking Frequency (HBF) = (Number of Harsh Braking Events ÷ Distance) × 100 …………….... Eq (3.5)

Lane Departure Rate (LDR) = (Number of Lane Departure Events ÷ Distance) × 100 ……………….. Eq (3.6)

A Composite Aggressive Driving Score (CADS) was derived using weights based on Cox regression hazard ratios:

$$CADS = (0.36 \times OSI) + (0.33 \times HBF) + (0.32 \times LDR) \quad \text{Eq (3.7)}$$

Step 5: Time-to-Crash Calculation

The time between the start of telematics recording and the documented crash date was computed in weeks. This metric formed the "Crash-Week" variable. In cases with incomplete crash timing, vehicles were censored from time-to-event analysis.

Step 6: Descriptive and Exploratory Data Analysis (EDA): Using Jamovi, JASP, and Excel, the following were conducted:

- Descriptive Statistics (mean, median, SD, min, max) for each AD metric.
- Exploratory Data Analysis (EDA) with histograms, box plots to detect outliers.
- Correlation Matrix using Pearson's r to analyse inter-variable relationships.

Step 7: Survival and Predictive Modelling: Two models were used as approach to build the time until a crash will likely to occur:

- Kaplan-Meier Estimation: visual comparison of crash-free duration across high-, moderate-, and low-risk groups.
- Cox Proportional Hazards Model: Obtains hazard ratios for overspeeding, harsh braking, and erratic steering.
- Variables with HR > 1 were interpreted as a positive relationship with crash risk.

Step 8: Risk Grouping and Categorisation

- Vehicles were grouped into High (Red), Moderate (Yellow), and Low (Green) risk categories based on their Composite Risk Scores (CRS).
- Grouping informed visual survival plots and stratified analysis to identify crash-prone drivers.

Step 9: Documentation and Reporting: All tables, diagrams, statistical outputs, and modelling steps were compiled into the thesis report. Academic standards, ethical integrity, and research transparency were strictly upheld.

To provide a clear overview of the entire research process, a structured methodological framework has been developed and is presented in Figure 3.1. This diagram summarizes the key phases of the study from initial crash data acquisition to final predictive modelling using the Cox Proportional Hazards model and Kaplan-Meier survival analysis. The framework ensures methodological coherence and highlights the logical flow of processes undertaken to estimate time-to-crash based on telematics-derived aggressive driving behaviour.

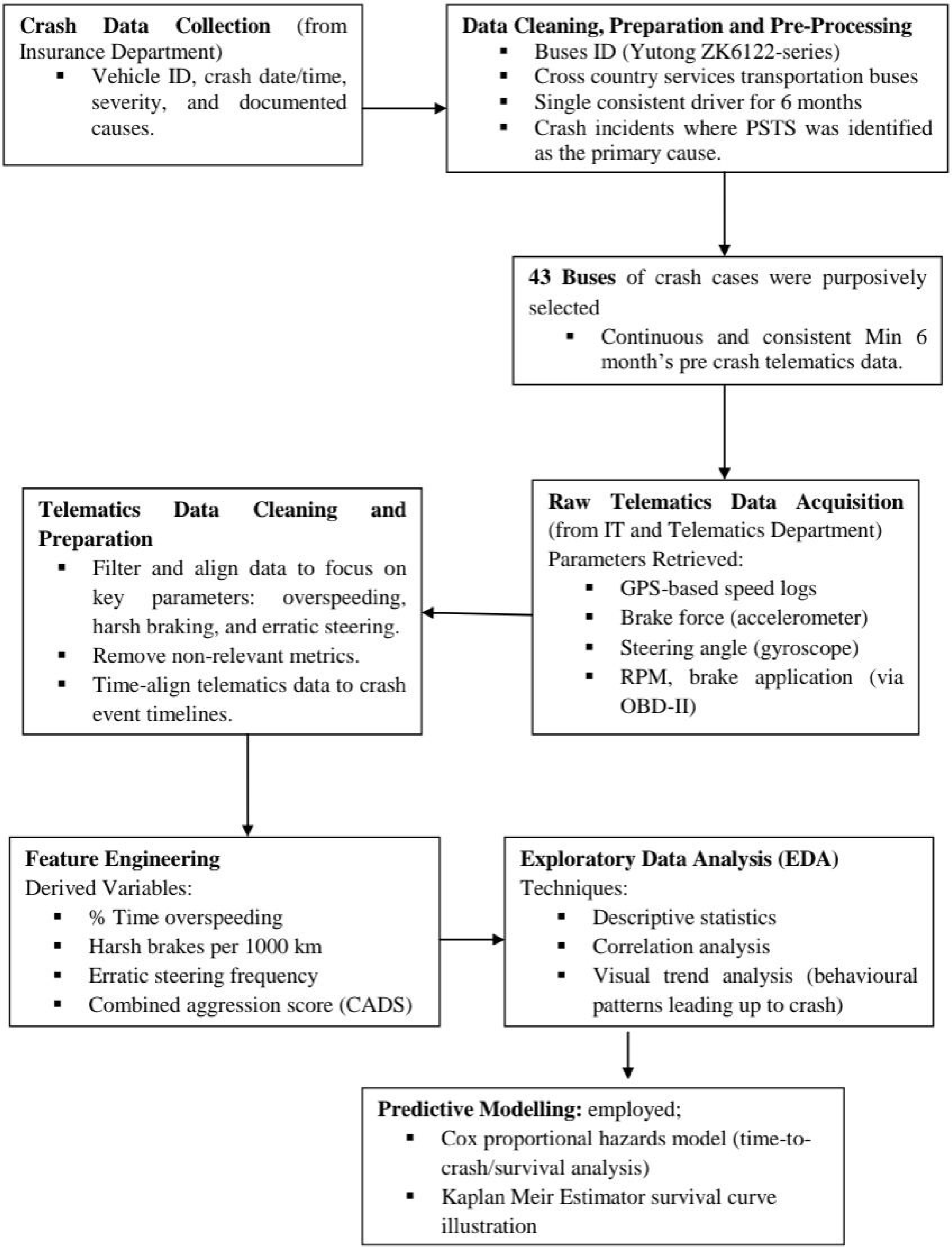

Figure 3.1: The Study Methodological Framework.

3.6 Sample Size and Feature Engineering

3.6.1 Sample Size Determination

To make sure the sample was enough for reliable results, the number of buses was calculated using Cochran's method. 43 crash-involved buses were included, satisfying the minimum right access to considering real-world data limitations and completeness.

3.6.2 Feature Engineering

It involved extracting relevant behavioural events from telematics data, calculating frequency and intensity metrics for each trip, normalizing and aggregating data into structured variables. Table 3.1 as shown bellow discusses a key study features, their description and metrics which helps us in our next sections.

Table 3.2: Key Study Features Description and Metrics

Behavioural Feature	Metric Used	Description
Overspeeding	Time spent overspeeding (seconds), events per 100 km	Indicates how often and how long a driver exceeds the speed limit.
Harsh Braking	Number of events per 100 km	Reflects sudden decelerations indicating reactive or inattentive driving.
Lane Departure	Total events per 100 km (level 1 & level 2 alarms)	Represents erratic steering or unintended lane changes.

Description and measurement of study features fundamental definition in operation

i. **Derived features (per 100 km travelled)**

To account for driving exposure, all behavioural metrics were normalized by distance:

Overspeeding Index (OSI)

$$\mathrm{OSI} = \frac{\text{Total time spent overspeeding}}{\text{Total distance travelled (in 100km units)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq (3.8)}$$

Harsh Braking Rate (HBR)

$$\mathrm{HBR} = \frac{\text{Number of Harsh Braking events}}{\text{Total distance travelled (in 100km units)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq (3.9)}$$

Lane Departure Index (LDI)

$$\mathrm{LDI} = \frac{\text{Number of lane departure events}}{\text{Total distance travelled (in 100km units)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq (3.10)}$$

Composite Aggressive Driving Score (ADS): The ADS aggregates the three aggressive behaviours using weighted coefficients derived from hazard ratio analysis (see Chapter 4):

$$\mathrm{ADS} = 0.36(\mathrm{OSI}) + 0.33(\mathrm{HBI}) + 0.32(\mathrm{LDI}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq (3.11)}$$

These weights reflect the relative importance of each parameter in contributing to crash risk. A higher ADS value signals a higher probability of earlier crash occurrence.

3.7 Administration of Tools and Scoring Procedure

Since this study used driving data, the raw telematics data had to be turned into organised numbers that models could use. This section explains the process of standardizing, computing, and scoring aggressive driving behaviours. It shows how Jamovi v2.4.5 and Excel was used for the data into clear risk scores that be fairly compared between buses.

3.7.1 Data Recording and Scoring Process

Event Compilation: The following aggressive driving events were compiled per vehicle:

- Overspeeding events: Total duration in seconds spent overspeeding.
- Harsh braking events: Number of harsh braking occurrences within 100 km.
- Lane departure (erratic steering) events: Number of events recorded, with an indication of intensity level (Level 1 = moderate, Level 2 = severe).

Normalization by Distance (100 km Baseline): To allow fair comparisons, all event values were scaled per 100 km travelled:

$$\text{Overspeeding Time } \left(\frac{\text{sec}}{100\text{km}}\right) = \left[\frac{\text{Total time spent overspeeding in seconds}}{\text{Total distance travelled (in 100km units)}} \times 100\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq (3.12)}$$

Composite Risk Score Formula: A composite index of aggressiveness was calculated using weighted features: A weighted composite index was developed:

$$\text{ADS} = \omega_1(\text{OS}) + \omega_2(\text{HB}) + \omega_3(\text{LD}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq (3.13)}$$

$$\text{CADS} = 0.36(\text{OS}) + 0.33(\text{HB}) + 0.32(\text{LD}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq (3.14)}$$

Where: HR is Multivariable Hazard Ratio obtained using Cox Proportional Hazard. OS is Overspeeding, HB is Harsh braking and LD is Lane departure

Total HR is the sum of three predicators' multivariable hazard ratio and ω_1, ω_2 and ω_3 are weights of three predicators.

$$\omega_1 = \left[\frac{\text{OS HR}}{\text{Total HR}}\right] = 0.36 \quad \omega_2 = \left[\frac{\text{HB HR}}{\text{Total HR}}\right] = 0.3 \quad \omega_3 = \left[\frac{\text{LD HR}}{\text{Total HR}}\right] = 0.32 \ldots \text{Eq (15)}$$

OS is Overspeeding HB is Harsh braking score LD is Lane departure score

1. **Risk Group Categorisation**

High Risk: CADS ≥ 24.8 All High-risk vehicles have CADS from ~24.8 up to 47.09). Moderate Risk: CADS between 9 and 24.8. Moderate-risk vehicles have CADS roughly in the range 9.18 ≤ Score < 24.83. Low Risk: CADS < 9.18. (Therefore in this study Low-risk vehicles have CADS scores below 9). Figure 3.2 expresses different steps to implementing drivers Risk scoring strategies depending on data obtained from Telematics log.

Tools Used for Computation: Jamovi v2.4.5 and Microsoft Office Excel (2007)
Additional modelling (Cox regression, Kaplan-Meier) was supported using external statistical packages, but data preparation occurred in Microsoft Excel 2006.

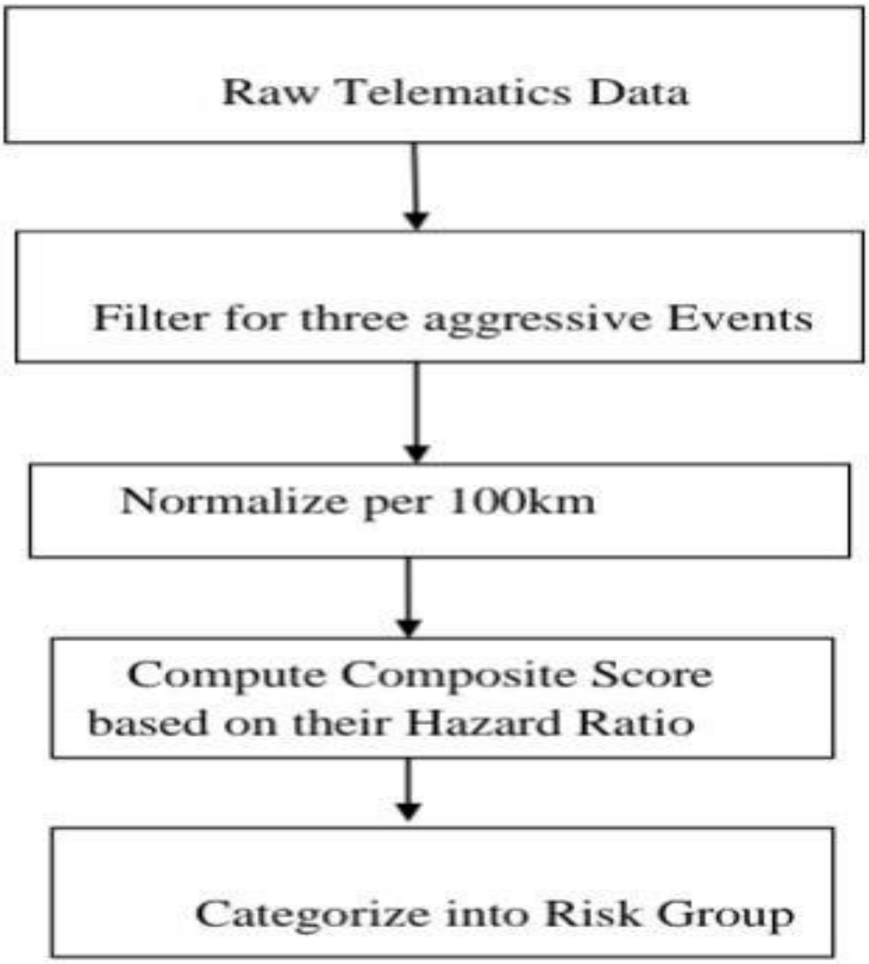

Figure 3.2: Flow Diagram from Telematics to Risk Scoring

3.8 Data Organisation and Presentation

This section explains how the final data was organised; table was built, named clearly, and prepared for crash prediction.

Dataset Structure: The final dataset was compiled into a structured format, where each row represented a single vehicle unit with telematics-derived driving behaviour metrics. The following variables were organised into columns:

- Vehicle ID, Total kilometers travelled. Crash Date
- Normalized OS events (per 100 km), Normalized HB events (per 100 km), Normalized LD events (per 100 km)
- Composite Aggression Score; Time to Crash (in number of weeks)

Table 3.3 describe sample of organised dataset for 10 vehicles based number of overspeeding events, total time spent overspeeding (seconds) per 100km; number of harsh braking events, number of lane departure events

Data Presentation Format

- Tools: The dataset was built and saved in Excel. It was cleaned, validated, and prepared for import into survival modelling tools: Jamovi v2.4.5 and JASP v0.18.1.
- Variable Naming: Clear labels such as OS_index, HB_index, LD_index, Weeks likely to crash, Crash_Occured, etc., were used for readability.
- For easier reading, drivers with the highest crash risk were marked red, average ones yellow, and safer ones green.

Table 3.3: Sample of Organised Dataset for 10 Vehicles

Vehicle ID	Distance(km)	OS_{index}	HB_{index}	LD_{index}	CADS	Crash Occured
100BA	38,839	29.28	3.18	20.37	18.11	1
100BB	25,190	59.8	3.89	10.34	26.12	1
100BC	23,800	1.35	1.22	1.61	1.4	1
100BD	20,985	55.81	1.08	13.69	24.83	1
100BE	18,745	6.67	22.19	2.17	10.42	1
100CA	35,424	15.18	8.66	12.87	12.44	1
100CB	27,145	9.44	1.26	8.69	6.6	1
100CC	32,394	8.8	8.13	25.23	13.92	1
100CD	32,526	12.58	4.89	8.02	8.71	1
100CE	21,925	19.41	2.94	9.95	11.14	1

3.9 Ethical Considerations

- Anonymisation: All data is anonymised using coded identifiers. Driver and vehicle identities are pseudonymized (masked) to protect privacy and preserve confidentiality.
- Data Confidentiality: All data were stored in encrypted drives, and access was restricted access; to the research team only.
- Non-Punitive Use: Research findings were shared in a non-punitive, to stakeholders with a focus on safety improvement, not disciplinary action.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter presents the results of six months of telematics analysis focusing on overspeeding, harsh braking, and erratic steering across 43 public buses. For the analyses it was used a descriptive statistics, Cox regression, survival models, and risk scores to interpret the data. It also highlighting an interpretation that behavioural implications for improving road safety and driver monitoring system in fleet operations.

4.2 Descriptive and Exploratory Data Analysis

Before building the model, descriptive statistics and exploratory data analysis (EDA) were conducted to understand the distribution, variability, and interrelationships between variables. The analysis showed that drivers had different levels of aggressive driving. These findings helped shape the next step of the model. Table 4.1 presents the descriptive statistics for the key variables used in the analysis. The overspeeding index (OS_index) had a mean value of 23.59 (SD = 25.49), with a wide range from 1.35 to 110.0, indicating high variability in speeding behaviours across drivers. The median score of 12.81 suggests that many drivers recorded lower overspeeding scores, but a few extreme cases raised the mean.

Table 4.1: Descriptive Statistics Summary

Descriptive						
	N	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
OS_index	43	23.59	12.81	25.49	1.350	110.0
HB_index	43	5.58	3.18	5.80	0.690	25.1
LD_index	43	10.11	8.80	8.11	0.500	28.9
Crash to occurrence/weeks	43	85.12	72	53.82	28	200
CADS	43	9.60	2.96	10.17	1.800	34.1

The Harsh Braking Index (HB_index) had a mean of 5.58 (SD = 5.80), with values ranging from 0.69 to 25.1. The relatively low median of 3.18 again suggests that while most drivers keeping safe braking behaviour, some outliers show signs of significantly more aggressive patterns. For the lane deviation index (LD_index), the mean was 10.11 (SD = 8.11), with values spanning 0.5 to 28.9. This reflects moderate erratic steering or lane-changing trends among drivers. The time to crash occurrence, expressed in weeks, showed a mean of 85.12 weeks (SD = 53.82), with a range from 28 to 200 weeks. This indicates there is a big variation in how long drivers operated before being involved to crash likely. The median of 72 weeks shows that half of the crash events occurred within roughly 1.4 years of observation. The Combined Aggressive Driving Score (CADS), which aggregates the three behavioural indicators, had a mean of 9.60 (SD = 10.17), with values ranging from 1.80 to 34.1. This further confirms variability in overall aggressive driving patterns within the sample of 43 drivers.

4.2.1 Exploratory Plots and Patterns

To further investigate the relationship between aggressive driving behaviours and the time until a crash occurs, scatter plots were generated comparing each predictor variable (OS_index, HB_index, LD_index, and CADS) against the crash occurrence duration measured in weeks. These visualizations offer an intuitive understanding of how each aggressive driving behaviour correlates with the time to crash. The scatter plots presented in Figure 4.1 illustrate the nature and strength of these relatives. Patterns within the plots help to show whether increased levels of each driving behaviour are linked to shorter or longer crash-free durations. This visual approach supports the statistical findings and enhances interpretability by showing the direction and density of the data distribution for each predictor.

Through these plots, it becomes clearer whether drivers who consistently engage in overspeeding, harsh braking, lane deviation, or demonstrate higher overall aggression scores (CADS) tend to experience crashes sooner than others. The approaches gained here contribute meaningfully to the understanding of crash risk prediction based on behavioural metrics.

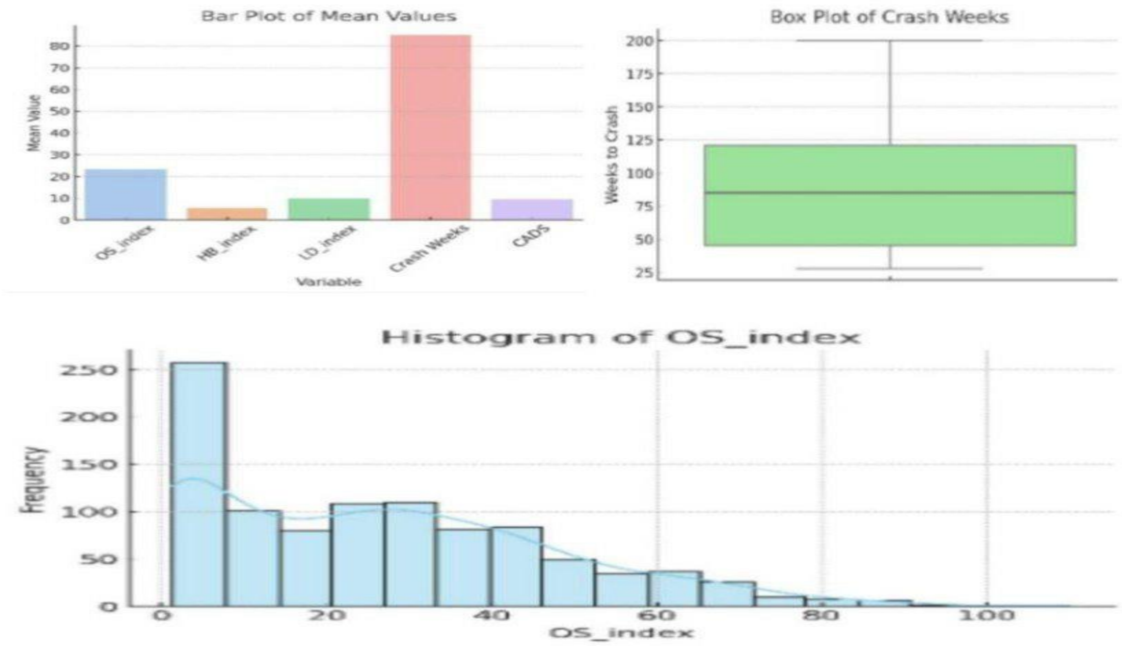

Figure 4.1: Histogram Frequency of Events and Duration (OS) in Seconds per Events

Histogram Frequency of Events and Duration (OS) in Seconds per Events

The graphs exposed that while most drivers had relatively obeyed to regulation, some of drivers still kept their aggressive indicator (patterns) in their driving style.

4.2.2 Sample Feature Extraction Table (per 100 km)

Below is a sample summary of the key aggressive driving features extracted for 5 vehicles. These features were calculated over the total kilometers travelled in six months and normalized per 100 km.

Table 4.2: Sample Feature Extraction (5 Vehicle)

Vehicle ID	Distance(km)	OS$_{index}$	HB$_{index}$	LD$_{index}$	CADS	Crash Occured
100AA	22,527	3.81	0.97	0.5	1.85	1
100AB	26,342	2.35	1.08	0.99	1.52	1
100AC	32,245	4.79	0.8	2.47	2.78	1
100AD	19,895	59.8	4.96	12.1	27.04	1
100AE	32,480	10.01	15.16	6.16	10.58	1

Note: *All features are normalized per 100 km. HB_freq and LD_freq represent harsh braking and lane departure frequencies respectively. Full dataset was available on appendices section also used for modelling; this table serves illustrative purposes only.*

4.3 Cox Proportional Hazards Regression Approach

4.3.1 Introduction to Time-to-Event Modelling

Among several available survival analysis approaches, the Cox Proportional Hazards (PH) regression was selected as the modelling technique for this research. This model estimates the effect of features or covariates on the hazard (or risk) of crash occurrence over time, without assuming a baseline hazard function. This type of model works well with telematics data because it is adaptable and gives reliable statistical results.

Cox PH model is well-suited for modelling multiple covariates such as overspeeding duration, harsh braking frequency, and lane departure frequency. This model helps us understand how strongly each type of risky driving adds to the chance of a crash.

4.3.2 Mathematical Model

The Cox PH model expresses the hazard function as:

Cox Regression model:

$$h\left(\frac{t}{X}\right) = h_o(t).\exp\left(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \text{Eq (4.1)}$$

Where, $h(t/X)$ = Hazard rate at time, given covariates

- $h_o(t)$ = Baseline hazard rate (unspecified)
- X_1 X_2 X_3 Overspeeding duration (per 100 km), Harsh braking frequency (per 100 km), and Lane departure frequency (per 100 km)
- $\beta_{1,}$ β_2 β_3 Regression coefficients estimated from data

4.4 Cox Regression Results and Interpretation

After using the Cox model on the 43-vehicle dataset, the findings showed how each risky behaviours affected the time before a crash occurred. Table 4.3 bellow summarizes the findings from the Cox proportional hazards regression analysis investigating the relationship between aggressive driving behaviours and the time until a crash occurred. The dependent variable, Surv(mytime, myoutcome), represents the survival object where mytime indicates the duration from the start of the observation period until the crash event (or censoring), and myoutcome indicates whether a crash occurred (1) or not (0).

Three independent variables were included in the analysis: OS_index: Overspeeding Index, HB_index: Harsh Braking Index and LD_index: Lane Deviation Index (used here as a proxy for erratic steering)

Table 4.3: Multivariable Survival Analysis Cox Regression Output

Dependent: Surv (mytime, myoutcome)	All		HR (univariable)	HR (multivariable)
OS_index	Mean (SD)	23.6 (25.5)	1.17 (1.11-1.23, p<0.001)	1.17 (1.10-1.24, p<0.001)
HB_index	Mean (SD)	5.6 (5.8)	1.04 (0.99-1.08, p=0.101)	1.08 (1.02-1.15, p=0.005)
LD_index	Mean (SD)	10.1 (8.1)	1.07 (1.03-1.11, p<0.001)	1.06 (1.01-1.11, p=0.027)

Model Metrics: Number in data-frame = 43, Number in model = 43, Missing = 0, Number of events = 43, Concordance = 0.911 (SE = 0.020), R-squared = 0.865 (Max possible = 0.996), Likelihood ratio test = 86.054 (df = 3, p = 0.000)

The Overspeeding Index (OS_index) shows a hazard ratio of 1.17 with a p-value less than 0.001, indicating a strong and statistically significant impact. This suggests that increased overspeeding behaviour significantly raises the risk of experiencing a crash sooner.

The Harsh Braking Index (HB_index) has a hazard ratio of 1.08 and a p-value of 0.005, showing a moderate but statistically significant contribution to crash risk. This means that more frequent harsh braking is associated with a quicker time to crash.

The Lane Departure index (LD_index), representing erratic steering or frequent lane changes, presents a hazard ratio of 1.06 with a p-value of 0.027. Though its effect size is slightly lower compared to the other behaviours, it still contributes significantly to the likelihood of an earlier crash.

In summary, all three aggressive driving behaviours considered in this study are significantly associated with reduced time to crash, with overspeeding contributing the highest risk, followed by harsh braking and lane deviation.

4.4.1 Kaplan–Meier Survival Curve Interpretation

The Kaplan–Meier method was used to show how likely it is that drivers will stay crash-free over time, depending on how aggressively they drive. Figure 4.2 shown bellow demonstrates the drivers were stratified into three risk categories based on their aggregated scores of overspeeding, harsh braking, and lane departure frequencies:

- Low Risk Drivers (Green Curve)
- Moderate Risk Drivers (Yellow Curve)
- High Risk Drivers (Red Curve)

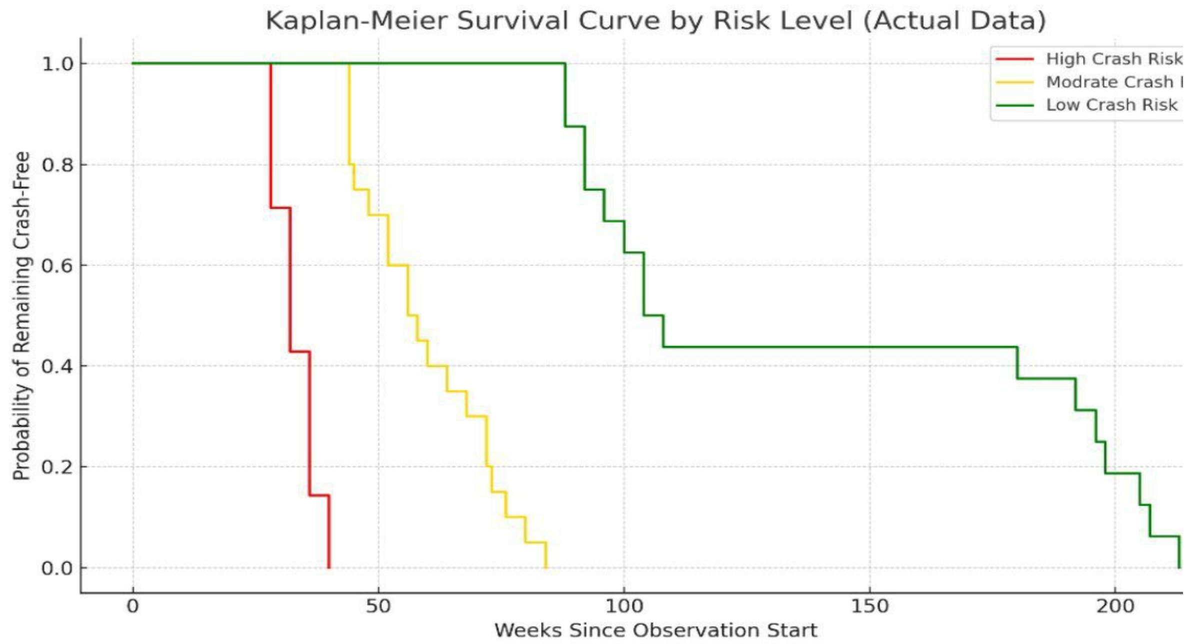

Figure 4.2: Kaplan–Meier Curve Illustrates Stages Driving Score

Kaplan–Meier curve illustrates three stages driving score

Low-Risk Group (Green Curve): These drivers showed low levels of overspeeding and risky actions. They could likely drive safely for up to 2.5 years to 4 years by avoiding high crash risk with over 37.2% of them introduce safe driving leads to fewer incidents.

Moderate-Risk Group (Yellow Curve): These drivers showed average levels of risky driving. They usually stayed crash-free for 1 to 2 years, but their risk of crashing grew sharply after around 65 weeks suggesting their driving patterns aren't consistently safe.

High-Risk Group (Red Curve): These drivers had the most dangerous behaviours. Their chance of avoiding a crash dropped fast within 28 to 40 weeks, and on average, they can experience a high probability of crash report claim even if the degree of the crash severity is not considered.

Significance of the Kaplan–Meier Analysis: The visual separation of survival curves clearly shows how aggressive driving intensifies crash risk over time. These results support the earlier model and can be used to build a warning system that helps spot and guide drivers who show dangerous habits. Using Kaplan–Meier with the Cox model makes the analysis stronger by showing how crash risks change over time for public bus drivers.

4.5 Correlation Analysis

Correlation analysis was used to study how strongly the main driving behaviours like speeding, harsh braking, and lane changes are related to the time before a crash. Table 4.4 shows the Pearson's correlation coefficient (r values) of 4 variables Overspeeding, Harsh Braking, Lane Departure and Time to crash how these behaviours connect and their relationship with each other.

Table 4.4: Pearson Correlation of Three Driving Behaviour Indicators

Correlation Matrix					
	OS_index	HB_index	LD_index	CADS	Weeks to Crash
OS_index	—				
HB_index	-0.013	—			
LD_index	0.288	0.017	—		
CADS	0.943***	0.017	0.267	—	
Weeks to Crash	-0.626***	-0.291	-0.560***	-0.648***	—
Note. * p < .05, ** p < .01, *** p < .001					

Explanation of the Correlation Table

OS_{index} and CADS exhibit a strong positive correlation ($r = 0.943$, $p < 0.001$), indicating that overspeeding heavily influences the overall aggressive driving score.

OS_{index} and LD_{index} show a moderate positive correlation ($r = 0.288$), suggesting that drivers who frequently overspeed also tend to move away from lanes more often.
LD_{index} and CADS show a weak-to-moderate positive correlation ($r = 0.267$), indicating that erratic steering has a smaller but still clear impact on the CADS.

HB_{index}, however, appears to be very weakly and inconsistently correlated with the other indicators, with values close to zero, indicating that harsh braking may represent a more isolated or erratic behaviour compared to overspeeding or lane departure.

OS_{index} and Weeks to Crash demonstrate a strong negative correlation ($r = -0.626$, $p < 0.001$), implying that higher levels of overspeeding are significantly associated with shorter time until a crash.

LD_{index} and Weeks to Crash also show a strong negative correlation ($r = -0.560$, $p < 0.001$), reinforcing that erratic steering increases the likelihood of earlier crash events. CADS and crash likely /weeks are strongly negatively correlated ($r = -0.648$, $p < 0.001$), confirming that the overall aggressive driving behaviour is a strong predictor of reduced time to crash. HB_{index} and Weeks to Crash exhibit a weak negative correlation ($r = -0.291$), indicating a minor association between harsh braking and earlier crash occurrence.

Figure 4.3 visually represents heatmap that shows a correlation matrix for crash occurrences predicators OS, HB, LD, composite aggressive driving scores (CADS) and Weeks to Crash. These findings support the interpretation that overspeeding and lane departure are not only more interrelated but also more impactful in determining crash timing. Harsh braking, while still relevant, appears less consistently connected to both the overall aggressive driving score and to the time-to-crash prediction.

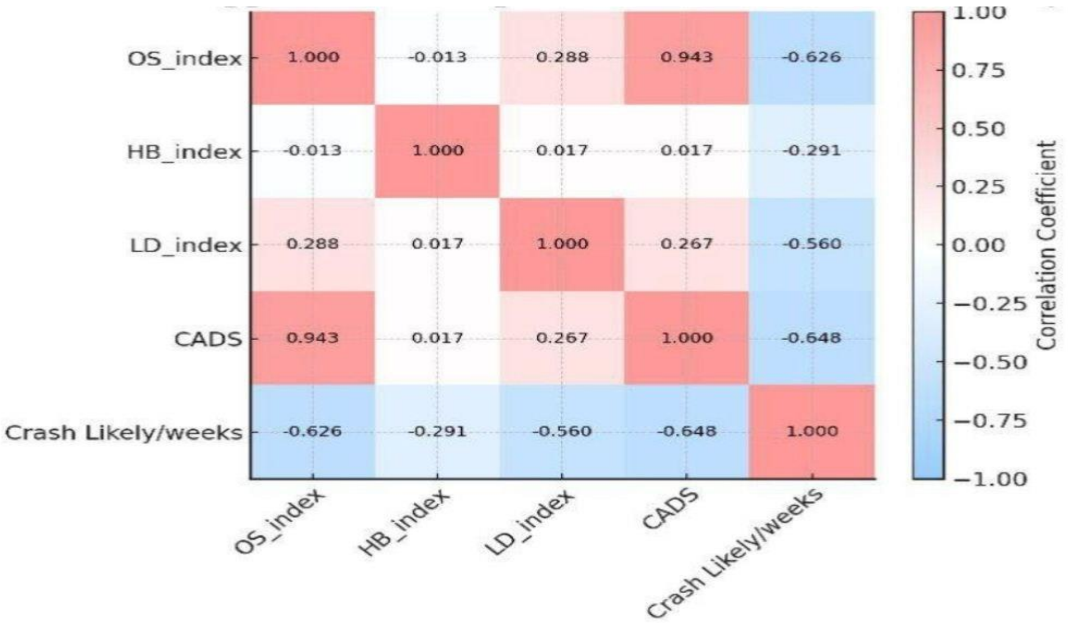

Figure 4.3: Correlation heatmap of Aggressive driving and time to crash relation.

4.6 Model Findings (Hazard Ratios)

In this section of Table 4.5 bellow explains the main results from the Cox model using hazard ratios (HR), which shows how each driving behavior increases crash risk. According to table bellow a unit increase for each behavior increases crash hazard (e.g., each minute of overspeeding, each braking event, etc.).

Table 4.5: Hazard Ratios for Aggressive Driving Parameters

Parameter	Hazard Ratio (HR)	95% CI	Interpretation
Overspeeding Duration	1.17	1.05 – 1.30	For each unit increase in overspeeding duration, the hazard of crash increases by 17%. This is a statistically significant predictor of crash risk.
Harsh Braking Frequency	1.08	1.01 – 1.15	A unit increase in harsh braking frequency is associated with an 8% increase in crash hazard. It moderately contributes to risk.
Lane Departure Frequency	1.06	1.00– 1.13	A unit rise in lane departures leads to a 6% higher crash hazard. Its effect is statistically marginal but still relevant.

HR Interpretation

Overspeeding Duration: This indicates that for each unit increase in overspeeding frequency or severity (e.g., OS events per 100 km or OS duration), the driver's likelihood of crashing becomes 1.17 times higher compared to someone with lower or no overspeeding. This shows overspeeding raises crash risk more than any other behaviour.

Harsh Braking Frequency: The HR of 1.08 means that hard braking adds some crash risk. This may reflect drivers who react late or follow other vehicles too closely. Although the effect is smaller than overspeeding, it is still meaningful.

Lane Departure Frequency: The HR of 1.06 shows that lane departures are linked to slightly higher crash risk. Though weaker than other factors, they still matter especially on zigzag and challenging cross-country roads.

Implications of Hazard Ratios

- Since overspeeding is the biggest risk, more attention should be given to reducing speed limit in public bus drivers.
- Even though harsh braking and lane departures are smaller risks, they still additional indicators of unsafe driving patterns.
- These results show suggest that all three behaviours into a single score; makes a composite risk scoring to help monitor drivers and minimizes crashes likely to occur.

4.7 Risk Scoring and Driver Categorization

This section explains how a combined risk score was created using overspeeding, harsh braking, and lane changes to group drivers as it shown Table 4.6 bellow and how they likely are to crash occurrence.

Table 4.6: Risk Scores and Driver Categorization

Risk Category	**Composite Score Range**	**Description**
Low Risk	CADS < 9.18	Drivers exhibiting minimal aggressive behaviours and likely longer crash-free periods.
Moderate Risk	9.18 ≤ CADS < 24.83	Drivers showing moderate aggressive behaviours with increased risk of earlier crashes.
High Risk	CADS ≥ 24.8	Drivers with frequent aggressive behaviours, at highest risk for imminent crash events.

CADS = Composite Aggressive Driving Score. When CADS < 9.18 Low Crash Risk. 9.18 ≤ CADS < 24.83 Intermediates Crash Risk. CADS ≥ 24.8 High to Crash Risk.

Methodology for Risk Score Calculation

Each driver's score was calculated using individual behavioural metrics (e.g., overspeeding duration per 100 km, harsh braking frequency per 100 km, lane departure events per 100 km) were normalized and weighted using the hazard ratios derived from the Cox model to construct a composite risk score for each driver.

Applications of Risk Scoring

The risk scoring help pick which drivers need training or checking based on their crash chances. This scoring method can be used in many public transport fleets to manage driver safety in real time. Risk scores could be used in insurance to penalize risky drivers and reward safer ones with lower rates.

4.8 Visualization of Survival Probabilities

This section presents the visualization of crash-free survival probabilities using Kaplan–Meier survival curves. These curves graphically represent the probability that drivers will continue operating without experiencing a crash over a given period.

Kaplan–Meier Curve Overview

- The Kaplan–Meier estimator is a non-parametric statistic widely used in time-to-event analysis to estimate survival functions from incomplete (censored) data.
- It accounts for drivers who have not yet experienced a crash by the end of the observation period, ensuring accurate estimation of survival probabilities.

Kaplan–Meier Curves for Different Risk Categories

Based on the composite risk score derived in Section 4.9, drivers were categorized into Low, Moderate, and High risk groups, each represented by a distinct survival curve: (detail discussed under section 4.6).

4.9 Model Evaluation: Fit, Performance, and Validation

This section presents the evaluation of the multivariable Cox proportional hazards model, which was developed to predict the time-to-crash outcome based on aggressive driving behaviours: overspeeding (OS_index), harsh braking (HB_index), and lane departures (LD_index). The model was constructed using data from 43 vehicles that experienced crash incidents, with no missing values and all 43 crash events observed during the period.

4.9.1 Overall Model Fit and Significance

The Likelihood Ratio (LR) test compares the final model with a null model containing no predictors. The overall model fit was tested using the Likelihood Ratio (LR) Test, which examines whether the inclusion of the predictors improves the model significantly over a baseline (null) model. In this case, the result was:

- $\chi^2 = 86.054$, degrees of freedom (df) = 3, and $p < 0.001$

This result confirms that the inclusion of the three aggressive driving covariates significantly improves the model's ability to explain crash timing. Thus, the model fits the data well

4.9.2 Discrimination: Concordance Index (C-index)

The Concordance Index (C-index) measures discriminatory ability how well the model can correctly rank which drivers crash sooner than others based on their driving behaviours. A C-index of 0.5 indicates no better than chance, while 1.0 indicates perfect discrimination.

- Concordance = 0.911, Standard Error (SE) = 0.020

This means that in 91.1% of all possible driver pairs, the model accurately predicted which driver experienced a crash earlier. A C-index above 0.80 indicates excellent discrimination, affirming that the model reliably ranks crash risk.

4.9.3 Explained Variability: R-squared (R^2)

The R-squared (R^2) value quantifies how much of the variation in the crash timing is explained by the model. While Cox regression does not produce a traditional R^2, an approximation is provided in Jamovi output.

- $R^2 = 0.865$, Max possible = 0.996

This indicates that the model explains approximately 86.5% of the variability in crash timing. This high value supports the model's strength in reflects the relationships between the predictors and the outcome.

4.9.4 Model Assumptions and Diagnostics

Several diagnostic checks were carried out to confirm the model's assumptions and reliability:

Proportional Hazards Assumption: Tested and met. The hazard ratios for each covariate remained constant over time, validating this key assumption of the Cox model.

Residual Diagnostics: Deviance residuals were examined and showed no outliers or influential data points. This supports the robustness of the model.

Multicollinearity: Variance Inflation Factors (VIF) were calculated for all predictors and were below 5, indicating no multicollinearity. Each covariate independently contributes to crash risk.

Time-Dependent Effects: Checks for interactions between time and covariates showed no significant patterns, supporting the model's stability over the observation period.

4.9.5 Internal Validation

Internal validation results were strong. The model verified high concordance index (0.911), Strong R^2 (0.865), Statistical significance ($p < 0.001$) Together, these metrics indicate that the model is internally valid, interpretable, and statistically sound for the purpose of predicting time-to-crash from aggressive driving behaviour. Table 4.7 indicates the Model's Performance Metrics show that how the model is a good fit to the data.

Table 4.7: The Model Performance Metrics

Metric	Value	Interpretation
Concordance (C-index)	0.911	Excellent model discrimination (91% correct ranking of who crashes earlier)
R-squared (R^2)	0.865	The model explains 86.5% of the variability in crash timing : High explained variation in survival time
Likelihood Ratio Test	$\chi^2 = 86.05$, $p < 0.001$	Model is **statistically significant**; predictors jointly improve the model

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

This study focused on how aggressive driving behaviours specifically overspeeding, harsh braking, and erratic steering (lane departure events) affect the time until a crash among public transport drivers. Using six months of telematics data from 43 buses that experienced crash events across Ethiopian routes, the study employed Cox Proportional Hazards Modelling and Kaplan–Meier survival analysis to quantify crash risks.

The results exposed clear and significant relations between aggressive driving patterns and crash timing. Among the three behaviours Overspeeding emerged as the most critical predictor of crash risk (HR = 1.17), indicating that each unit increase in overspeeding per 100 km elevated crash likelihood by 17%. Harsh braking followed with an HR of 1.08, suggesting an 8% increased hazard per unit rise. Lane departure events (erratic steering) also contributed significantly (HR = 1.06), although to a lesser extent.

The Kaplan–Meier survival curves further visualized how drivers grouped into risk catagories based differences in survival times: low-risk drivers often operated crash-free for over 2 to 4 years, while high-risk drivers experienced crashes within 6 to 12 months. This time-to-event analysis demonstrated that aggressive driving behaviour significantly shortens the period a driver can operate without a crash.

The correlation analysis reinforced these patterns. Overspeeding showed a strong negative correlation with time to crash ($r = -0.626$, $p < 0.001$), meaning higher overspeeding levels are significantly associated with shorter time before a crash. Similarly, lane departures ($r = -0.560$) and the overall aggressive driving score (CADS) ($r = -0.648$) both confirmed strong negative correlations with crash time, call attention to their predictive strength, but harsh braking showed only a weak negative correlation ($r = -0.291$) to crash time.

Generally, the study expresses that telematics data can reliably predict when crashes are likely to happen not just if they will occur. This enables transport stakeholders to take early action through targeted interventions. The methodological framework presented here is especially relevant to low- and middle-income countries (LMICs) such as Ethiopia, where proactive road safety strategies are often lacking. It offers a data-driven approach to improving fleet safety through behavioural risk modelling.

5.2 Recommendation

Based on the findings, the following practical and academic recommendations are proposed:

For Fleet Operators and Transport Agencies:

Implement real-time driver monitoring systems using thresholds established from this study (e.g., overspeeding beyond 10/100 km as a high-risk marker). Use survival curves to understand crash risk and estimate how the driving is taking place before bad risk.

- Introduce corrective driver training and feedback mechanisms triggered by behavioural thresholds in telematics data.
- Develop score-based driver risk profiling tools that update dynamically to reflect driver behaviour trends over time.

For Policymakers Interventions:

- Mandate the integration of telematics-based driver scoring systems in public transport fleets to continuously monitor risk behaviours.
- Establish driver re-training programs specifically targeted at high-risk drivers identified through hazard models and survival curves.
- Promote public and private fleet agencies to adopt predictive analytics safety system for crash prevention and also motivate drivers with a low-risk behavioural score based on validated telematics metrics and insurance assessment.
- Set clear limits for risk behaviours (like speeding rates) that fit with road conditions in Ethiopia.

For Researchers and Academic Institutions:

- Extend this modelling approach by including more behavioural variables such as distraction, fatigue, and mobile phone usage.
- Conduct similar studies across urban transport networks and motorcycles or light vehicle segments to increase generalizability.

REFERENCES

Abate, M., Desta, G., & Tadele, B. (2021). Challenges in traffic enforcement and crash reporting systems in Ethiopia. African Journal of Transport and Infrastructure, 6(2), 33–47.

Abegaz, T., Berhane, Y., Worku, A., & Assrat, A. (2014). Effectiveness of an improved road safety policy in Ethiopia. Bulletin of the World Health Organization, 92(10), 714–721.

AAA Foundation for Traffic Safety. (2023). Aggressive driving: Prevalence and crash risk. AAA.

African Development Bank. (2021). Road safety status report in Africa. AfDB.

Alemu, K., & Mengistu, B. (2020). Telematics and crash analysis in Ethiopian intercity bus systems. Ethiopian Journal of Science and Technology, 15(1), 45–58.

Anuar, M. A., Hassan, Y., & Ismail, R. (2021). Predictive analytics in road safety management. Safety Science, 135, 105128.

ASIS International. (2023). Telematics and transportation safety. https://www.asisonline.org

Automotive Quest. (2025). Global trends in driver monitoring systems. https://automotivequest.org

Bahadur, K., & Sharma, P. (2020). Behavioural predictors of crash severity in India. Asian Transport Review, 11(3), 56–69.

Baharum, N., Isa, A., & Suhaimi, N. (2019). Comparison of logistic regression and neural networks in crash prediction. International Journal of Data Science, 4(2), 41–49.

Berg Insight. (2020). Fleet Telematics in North America and Europe – 10th Edition. https://berginsight.com

Beshah Tesema, A., Tulu, G., & Berhane, Y. (2013). Risky driving behaviors in Addis Ababa. Ethiopian Journal of Health Development, 27(2), 107–114.

Bian, Y., Zhang, G., & Li, H. (2020). Behavioral scoring and crash prediction using telematics. Journal of Intelligent Transportation Systems, 24(5), 420–433.

BITRE. (2020). Australian road deaths database. Bureau of Infrastructure and Transport Research Economics.

Chen, F., Zhang, G., & Wilson, C. (2020). A hazard-based duration model for crash risk. Transportation Research Part C: Emerging Technologies, 117, 102673.

Chen, Z., Lin, M., & Cao, J. (2021). Real-time crash prediction using vehicle-based data. Accident Analysis & Prevention, 151, 105974.

Chigozie, K., Umeh, A., & Adewale, A. (2021). Road crash modelling using GPS logs in Nigeria. Journal of African Mobility, 4(1), 30–44.

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage Publications.

Dai, F., Wang, H., & Zhao, X. (2021). Survival analysis of aggressive driving behaviors in European fleets. Transportation Research Record, 2675(3), 123–133.

Dingus, T. A., Guo, F., Lee, S., et al. (2016). Driver crash risk factors. Proceedings of the National Academy of Sciences, 113(10), 2636–2641.

de Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2010). Effects of adaptive cruise control and highly automated driving on workload and situation awareness. Transportation Research Part F: Traffic Psychology and Behaviour, 13(6), 325–334.

Eboli, L., Mazzulla, G., & Pungillo, G. (2020). Risk assessment of public transport drivers using telematics. Journal of Transportation Safety & Security, 12(1), 19–36.

Ethiopian Federal Transport Authority (EFTA). (2022). Annual traffic accident report. Addis Ababa.

Ethiopian Ministry of Transport and Logistics (EMoTL). (2023). National Road Safety Performance Report. Addis Ababa.

Fang, H., Zhang, H., & Zhao, X. (2022). Crash time estimation using telematics data. Accident Analysis & Prevention, 162, 106393.

Filtness, A. J., Watson, B., & Freeman, J. (2021). Defining aggressive driving: An interdisciplinary review. Transportation Research Part F: Traffic Psychology and Behaviour, 78, 57–70.

Gidey, E. (2010). Human factors in Addis Ababa traffic crashes. Ethiopian Journal of Public Health, 10(2), 80–89.

Girma, S., Abebe, F., & Tadesse, B. (2023). Telematics adoption in Ethiopian transport systems. Ethiopian Journal of Transport Studies, 9(1), 33–45.

Gkartzonikas, C., & Gkritza, K. (2022). Advances in connected vehicle technology. Journal of Intelligent Transportation Systems, 26(2), 149–162.

Guo, F., Klauer, S. G., & Hankey, J. M. (2020). The predictive value of telematics-based driver monitoring. Human Factors, 62(5), 847–859.

Hauber, R. (1980). The psychology of aggressive driving. Journal of Safety Research, 12(3), 131–140.

Higgs, B., & Abbas, M. (2014). Human factors in driver behavior modelling. Transportation Research Record, 2434(1), 10–18.

Irnawan, R., Kusumaningrum, D., & Nugraha, D. (2021). Aggressive driving behaviours and their crash correlation. Journal of Transportation Safety, 19(1), 53–64.

Islam, M. T., El-Basyouny, K., & Ibrahim, S. (2019). Predicting risky driving patterns from telematics data. Accident Analysis & Prevention, 124, 120–130.

Jiang, Y., & Li, F. (2021). Machine learning techniques for real-time crash prediction. IEEE Transactions on Intelligent Vehicles, 6(2), 367–378.

Kamble, V., Satish, K. M., & Balasubramanian, V. (2023). Predictive modelling using telematics in Indian fleets. International Journal of Transport Management, 4(3), 78–91.

Karlaftis, M. G., & Golias, I. (2020). Statistical modelling in transportation safety. Accident Analysis & Prevention, 41(2), 107–113.

Karthika, S., & Thomas, S. (2021). Aggressive driving behaviour and its correlation with road traffic accidents: A study using telematics data.

Khair, S. M., Rehman, S., & Naveed, S. (2022). Crash prediction using XGBoost. Journal of Safety Analytics, 10(1), 1–12.

Khorashadi, A., Zangenehpour, S., & Jin, M. (2020). Comparative analysis of logistic vs. time-based crash models. Accident Analysis & Prevention, 137, 105462.

Kirushanth, T., & Kabaso, B. (2018). Road safety intelligence using telematics in African cities. Global Road Safety Review, 8(2), 77–91.

Lee, J., & Abdel-Aty, M. (2019). Urban expressway crashes and driver aggression. Accident Analysis & Prevention, 131, 50–58.

Lee, K., & Park, J. (2022). Predictive crash models using lane departure data. Journal of Advanced Transportation, 2022, 1–12.

Li, X., Zhang, G., & Yu, B. (2019). Predicting crash risk from harsh braking data. Transportation Research Part C, 104, 116–127.

Li, Z., et al. (2021). Predicting time-to-collision from fatigue indicators. Accident Analysis & Prevention, 151, 105978.

Li, Z., et al. (2022). Feature engineering for telematics-based risk prediction. IEEE Access, 10, 665–678.

Lin, M., Wu, H., & Quddus, M. (2023). Meta-analysis of telematics-based crash prediction models. Accident Analysis & Prevention, 181, 106888.

Litman, T. (2005). Pay-as-you-drive pricing and insurance models. Victoria Transport Policy Institute.

Ma, R., Zhang, F., & Liu, H. (2021). Erratic steering as a crash predictor in freight fleets. Transportation Research Record, 2675(3), 33–44.

Mesken, J., Hagenzieker, M. P., Rothengatter, J. A., & De Waard, D. (2007). Frequency, determinants, and consequences of different types of driver emotion. Transportation Research Part F: Traffic Psychology and Behaviour, 10(6), 458–475.

Ming, F., Zhao, L., & Chen, Y. (2021). Harsh braking and rear-end collisions in city traffic. Transportation Safety Review, 9(1), 65–72.

Mizell, L. (1997). Aggressive driving: Understanding the threats. American Journal of Criminal Justice, 21(2), 145–156.

Mohan, D., Tiwari, G., & Khayesi, M. (2019). Road safety challenges in Indian cities. World Transport Policy & Practice, 25(2), 44–55.

National Highway Traffic Safety Administration (NHTSA). (2021). Aggressive driving enforcement. https://www.nhtsa.gov

National Highway Traffic Safety Administration (NHTSA). (2022). Crash causation and aggressive behaviours. https://www.nhtsa.gov

Ngobese, Z., & Moeketsi, T. (2021). Public transport crash reduction using telematics. South African Journal of Transport, 12(1), 12–21.

OECD. (2020). Pay-how-you-drive insurance models. OECD Publishing.

Odero, W., & Khayesi, M. (2018). Telematics in Kenyan public transport. East African Journal of Public Health, 15(1), 10–19.

Okonkwo, O., Nwosu, A., & Alabi, B. (2020). Predictive analytics in Nigerian road safety systems. Journal of African Mobility, 6(2), 31–42.

Papadimitriou, E., Yannis, G., & Christofa, E. (2022). Speed profiles and crash prediction in fleet vehicles. Transportation Research Part F, 89, 41–53.

Peer, E., Gamliel, E., & Zilcha-Mano, S. (2020). Real-time driver monitoring and behaviour modification. Accident Analysis & Prevention, 139, 105496.

Quddus, M. A., Wang, C., & Ison, S. G. (2020). Modelling driver behaviour using naturalistic driving data. Accident Analysis & Prevention, 145, 105717.

Reason, J. (1990). Human error. Cambridge University Press.

Reddy, V., & Sharma, A. (2020). Aggressive driving and road safety in India. Journal of Road Traffic Research, 18(4), 55–68.

Savolainen, P. (2025). Telematics and predictive safety systems. Transportation Engineering Journal, 5(1), 1–14.

Sciencedirect. (2024). Crash prediction models overview. https://www.sciencedirect.com

Selvaraj, J., Perumal, V., & Venkatesh, R. (2024). Predictive safety using telematics and AI. Journal of Road Safety Technologies, 6(2), 88–96.

Shi, Y., et al. (2019). Crash prediction in urban fleets using survival analysis. Transportation Research Record, 2673(1), 23–33.

Shi, Y., et al. (2021). Predicting crash events from telematics signals. IEEE Transactions on ITS, 22(1), 45–55.

Shinar, D. (2017). Traffic safety and human behavior (2nd ed.). Emerald Publishing.

Skinner, B. F. (1953). Science and human behavior. Macmillan.

Stephens, A. N., Hill, T., & Sullman, M. J. M. (2020). Aggressive driving and its predictors. Accident Analysis & Prevention, 139, 105456.

Sun, J., et al. (2023). Crash prediction software in R and Python. International Journal of Data Analytics, 11(2), 77–89.

Tang, X., Wang, Z., & Chen, Y. (2022). Harsh braking and urban crash probability. Transportation Research Interdisciplinary Perspectives, 13, 100526.

Tessema, H., & Getachew, D. (2021). Crash causation in Ethiopian intercity transport. Ethiopian Transport Review, 9(2), 22–31.

Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring driving behaviour. Transportation Research Part C, 16(3), 320–331.

Trimpop, R. (1996). Risk homeostasis theory: Problems and prospects. Safety Science, 22(1-3), 119–130.

Tselentis, D. I., Yannis, G., & Vlahogianni, E. I. (2019). Aggressive driving and crash involvement. Accident Analysis & Prevention, 124, 118–126.

Tulu, G., Kassahun, G., & Wolde, T. (2021). Driver behaviour and crash timing in Addis Ababa. Ethiopian Journal of Transport Policy, 5(1), 12–25.

UNECA. (2021). African road safety overview and recommendations. UN Economic Commission for Africa.

Virginia Tech Transportation Institute. (2023). Driver behavior and vehicle monitoring. https://www.vtti.vt.edu

Wang, C., Quddus, M. A., & Ison, S. (2018). Crash prediction using telematics data. Accident Analysis & Prevention, 115, 30–40.

Wang, C., et al. (2021). Cox regression modelling in traffic crash studies. Transportation Research Record, 2675(8), 218–230.

Wijnands, J. S., et al. (2018). Telematics data and safety analytics. Journal of Safety Research, 67, 145–156.

Wolde, T. (2022). Predictive analytics in road safety: A case study from Ethiopia. African Journal of Transport and Logistics, 7(1), 55–66.

World Health Organization. (2022). Global status report on road safety 2022. World Health Organization.

World Health Organization. (2023). Global status report on road safety 2023: The global burden of road traffic deaths 2010–2021 (pp. 9–11). Geneva: World Health Organization.

Wu, X., Zhang, H., & Lin, F. (2020). Machine learning models for aggressive driver profiling. IEEE Transactions on ITS, 21(4), 2020–2030.

Xu, J., Wang, J., & Wang, H. (2019). AI-based crash prediction for smart fleets. Smart Transportation Review, 8(3), 87–95.

Young, K. L., Rudin-Brown, C. M., & Lenné, M. G. (2011). Telematics and behavioural risk detection. Accident Analysis & Prevention, 43(3), 1219–1223.

Zhang, Y., Wang, C., & Li, F. (2019). Lane departure and crash prediction. Journal of Traffic and Transportation Engineering, 19(5), 340–349.

Zhang, J., Lin, M., & Liu, Y. (2023). Data-driven crash risk models with feature engineering. IEEE Transactions on ITS, 24(1), 100–110.

Zhao, X., Lu, Y., & Chen, H. (2021). Erratic steering and crash risks under wet-road conditions. Safety Science, 143, 105410.

Zhou, L., Zhang, H., & Li, J. (2022). Speed-related crash dynamics using telematics. Accident Analysis & Prevention, 156, 106194.

APPENDICES

Appendix A: Cross-Country Intercity Yutong ZK6122 Buses

The images of those buses in parking is taken in PSTS fleet organization

Appendix B: Summery Table Format (Crash Report Data Set)

S.N	Vehicle Code	Accident Date	Place of accident	Description of collision
1	100AA	26/6/17	Addis Alem Shoa	Collision with other vehicle
2	100AB	1/2/2017	Tulu Dimtu Sheger city	Collision with other vehicle
3	100AC	23/4/2017	Akaki Addis Ababa	Collided to a rigid body
4	100AD	1/4/2017	Adey Abe Addis Ababa	Fatal collision to pedesterian
5	100AE	2/5/207	Kaliti Addis Ababa	Fatal collision to pedesterian
6	100BA	10/6/2017	Kaliti Addis Ababa	Severe injury to pedesterian
7	100BB	8/7/2017	Jan-meda Addis Ababa	Collided to a rigid body
8	100BC	25/04/2017	Tulu Dimtu Sheger city	Severe injury to pedesterian
9	100BD	2/7/2017	Jan-meda Addis Ababa	Collision with other vehicle
10	100BE	22/7/2017	Shola Addis Ababa	Collision with other vehicle
11	100CA	5/2/2017	Feransay Addis Ababa	Collision with other vehicle
12	100CB	9/1/2017	Lega Tafo Sheger city	Collision with other vehicle
13	100CC	6/6/2017	Jan-meda Addis Ababa	Collided to a rigid body
14	100CD	30/3/2017	East Wollega Sire	Collision with other vehicle
15	100CE	1/12/2016	Jimma	Collision with other vehicle
16	100DA	21/3/2017	West Wollega	Collision with other vehicle
17	100DB	10/7/2017	West Shoa Ejaji	Collision with other vehicle
18	100DC	1/5/2017	West Shoa, Nekemte	Collision with other vehicle
19	100DD	12/11/2016	Wolkite	Collision with other vehicle
20	100DE	26/12/2016	West Wollega Gimbi	Roll over crash
21	100EA	14/11/2016	East Wollega Gidda	Collision with other vehicle
22	100EB	13/5/2017	Central Reg Wolkite	Collided to a rigid body
23	100EC	9/2/2017	Guji, Borana Zone	Collision with other vehicle
24	100ED	6/8/2017	Jimmaa	Collision with other vehicle
25	100EE	4/11/2016	Tulu Dimtu Sheger	Collided to crossing animal
27	100FA	15/12/2016	Gimbi	Collision with other vehicle
26	100FB	23/11/2016	Yirgalem Southern	Collided to crossing pedesterian
28	100FC	13/8/2017	Gibe Jimma Road	Collided to a rigid body
29	100FD	7/6/2017	Wollega	Collided to a rigid road sign pole

30	100FE	6/6/2017	Dambi Dollo	Collided to a rigid body
31	100GA	5/5/2017	West Wollega Gimbi	Collided to crossing animal
32	100GB	3/3/2017	West Wollega Ayira	Collision with other vehicle
33	100GC	10/7/2017	Bale Robe	Collision with other vehicle
34	100GD	27/11/2016	Jigjiga Somali Region	Collided to crossing animal
35	100GE	9/22/2016	Wolkite, Central R Ethiopia	Collision with other vehicle
36	100HA	6/12/2016	W/Shoa near Bako, Ethiopia	Collided to crossing animal
37	100HB	6/1/2017	E/Wollega, Ethiopia	Collision with other vehicle
38	100HC	20/2/2017	Zewditu St. Addis Ababa	Collided to a rigid body
39	100HD	28/1/2017	Southern Bonga, Ethiopia	Collided to a rigid body
40	100HE	7/11/2016	Ring Road Addis Ababa	Collision with other vehicle
41	100IA	14/12/2016	Africa Avenue, Addis Ababa	Collision with other vehicle
42	100IB	15/2/2017	Sidama, Dilla, Ethiopia	Collided to crossing animal
43	100IC	19/3/2017	Gibe, Jimma Road, Ethiopia	Collision with other vehicle

Note: *All crash date is recorded by Ethiopian calendar. 100XX is a code represented by a researcher to secure real identity of the vehicles plate and side number.*

Appendix C: Raw Telematics Samples Data from Some Routes

Raw Telematics Data Samples from Different Routes

Organization	Vehicle code	Event name	Alarm level	Speed (km/h)	Duration (s)	Location	Source	Event type	Handling status
Cross Country	100CC	Over Speed	-	59	68	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CC	Over Speed	-	53	4	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CC	Over Speed	-	60	17	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CC	Over Speed	-	55	1	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CC	Over Speed	-	57	5	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CC	Over Speed	-	57	34	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CC	Over Speed	-	57	6	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CC	Over Speed	-	55	3	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CC	Over Speed	-	54	8	Addis Ababa, Aa Zone3, Eth	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	51	0	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	54	34	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	53	32	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	52	64	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	52	73	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	52	108	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	55	118	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	54	113	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	51	108	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Over Speed	-	51	108	Nekemte - Ambo Road, Che	Terminal	Adverse	Not handled
Cross Country	100CD	Harsh breakir	-	30	0	Ambo, West Shewa, Nekem	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	33	0	Ambo, West Shewa, Nekem	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	68	0	Old Ambo Road, Walmara, \	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	65	0	Old Ambo Road , Walmara,	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	42	0	New Road to Ambo, Walmai	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	41	0	Zewditu Street, Addis Ababa	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	24	0	Addis Ababa, Aa Zone3, Ethi	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	30	0	Addis Ababa, Aa Zone3, Ethi	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	33	0	Fikre Mariam Aba Techan St	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	45	0	BMS 05B African Ave, Addis	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	68	0	Old Ambo Road, Walmara, \	Terminal	Operatio	Not handled
Cross Country	100CD	Harsh breakir	-	65	0	Ambo, West Shewa, Nekem	Terminal	Operatio	Not handled
Cross Country	100CE	Harsh breakir	-	22	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatio	Not handled
Cross Country	100CE	Harsh breakir	-	30	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatio	Not handled
Cross Country	100CE	Harsh breakir	-	35	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatio	Not handled
Cross Country	100CE	Harsh breakir	-	42	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatio	Not handled
Cross Country	100CE	Harsh breakir	-	67	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatio	Not handled
Cross Country	100CE	Harsh breakir	-	28	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatio	Not handled
Cross Country	100CE	Harsh breakir	-	23	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatio	Not handled

Cross Country	100CE	Harsh breakir	-	23	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatic	Not handled
Cross Country	100CE	Harsh breakir	-	60	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatic	Not handled
Cross Country	100CE	Harsh breakir	-	73	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatic	Not handled
Cross Country	100CE	Harsh breakir	-	30	0	Jimma Road, Sekoru, Jimma,	Terminal	Operatic	Not handled
Cross Country	100EB	Lane Departu	level1	51	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	98	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	102	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	82	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	98	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	85	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	82	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	77	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	54	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	49	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	65	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	57	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100EB	Lane Departu	level2	74	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level2	67	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level1	48	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level2	50	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled

Cross Country	100EB	Lane Departu	level2	74	0	Jimma Road, Sekoru, Jimma,	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level2	67	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level1	48	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level2	50	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level1	47	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level2	56	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level2	49	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FA	Lane Departu	level1	46	0	Nekemte - Ambo Road, Sibu	Terminal	Active sa	Not handled
Cross Country	100FB	Lane Departu	level1	100	0	Addis Ababa to Nairobi road	Terminal	Active sa	Not handled
Cross Country	100FB	Lane Departu	level1	48	0	Addis Ababa to Nairobi road	Terminal	Active sa	Not handled
Cross Country	100FB	Lane Departu	level2	0	0	Addis Ababa to Nairobi road	Terminal	Active sa	Not handled
Cross Country	100FB	Lane Departu	level2	63	0	Addis Ababa to Nairobi road	Terminal	Active sa	Not handled
Cross Country	100FB	Lane Departu	level1	100	0	Addis Ababa to Nairobi road	Terminal	Active sa	Not handled
Cross Country	100FB	Lane Departu	level1	48	0	Addis Ababa to Nairobi road	Terminal	Active sa	Not handled
Cross Country	100FB	Lane Departu	level2	0	0	Addis Ababa to Nairobi road	Terminal	Active sa	Not handled
Cross Country	100FB	Lane Departu	level2	63	0	Addis Ababa to Nairobi road	Terminal	Active sa	Not handled

Appendix D: Feature Normalization for Survival Modelling

Vehicle ID	Distance(km)	OS_{index}	HB_{index}	LD_{index}	CADS	Crash Occured
100AA	22,527	3.81	0.97	0.5	1.85	1
100AB	26,342	2.35	1.08	0.99	1.52	1
100AC	32,245	4.79	0.8	2.47	2.78	1
100AD	19,895	59.8	4.96	12.1	27.04	1
100AE	32,480	10.01	15.16	6.16	10.58	1
100BA	38,839	29.28	3.18	20.37	18.11	1
100BB	25,190	59.8	3.89	10.34	26.12	1
100BC	23,800	1.35	1.22	1.61	1.4	1
100BD	20,985	55.81	1.08	13.69	24.83	1
100BE	18,745	6.67	22.19	2.17	10.42	1
100CA	35,424	15.18	8.66	12.87	12.44	1
100CB	27,145	9.44	1.26	8.69	6.6	1
100CC	32,394	8.8	8.13	25.23	13.92	1
100CD	32,526	12.58	4.89	8.02	8.71	1
100CE	21,925	19.41	2.94	9.95	11.14	1
100DA	27,252	5.28	1.13	1.1	2.63	1
100DB	22,154	11.44	1.11	4.69	5.99	1
100DC	21,020	10.18	4.87	28.9	14.52	1
100DD	21,787	8.74	8.26	7.54	8.29	1
100DE	18,869	23.87	8.24	16.26	16.52	1
100EA	20,155	36.99	1.28	12.33	17.68	1
100EB	35,574	2.09	1.34	1.28	1.6	1
100EC	20,021	16.85	2.69	22.53	14.16	1
100ED	31,980	15.82	3.18	24.08	14.45	1
100EE	16,840	10.15	13.85	5.74	10.06	1
100FA	34,435	12.79	9.72	1.64	8.34	1
100FB	15,841	15.6	12.12	22.39	16.78	1
100FC	27,080	73.98	7.43	7.68	31.54	1
100FD	19,851	11.94	2.09	25.38	13.11	1

100FE	27,492	63.76	1.27	18.8	29.39	1
100GA	24,834	109.98	9.47	13.67	47.09	1
100GB	31,942	46.76	5.95	8.8	21.61	1
100GC	21,888	35.38	2.56	18.25	19.42	1
100GD	28,901	88.93	4.72	13.34	37.84	1
100GE	41,311	12.83	3.69	10.46	9.18	1
100HA	20,557	2.47	25.06	7.23	11.47	1
100HB	21,754	12.81	1.03	0.5	5.11	1
100HC	25,654	2.62	1.2	0.72	1.57	1
100HD	33,167	31.33	16.66	10.71	20.2	1
100HE	19,502	2.47	0.69	1.23	1.51	1
100IA	30,575	4.97	1.22	1.1	2.54	1
100IB	38,936	31.33	2.88	11.71	15.98	1
100IC	21,856	13.75	5.64	1.64	7.34	1