

**DISTRACTED DRIVING CLASSIFICATION  
BASED ON DRIVING SKILLS IN FEMALES**

**TANISH A/L RAJOO**

**UNIVERSITI SAINS MALAYSIA**

**2021**

# **DISTRACTED DRIVING CLASSIFICATION BASED ON DRIVING SKILLS IN FEMALES**

**By**

**TANISH A/L RAJOO**

(Matrix no: 137866)

Supervisor:

**Associate Professor Dr. Loh Wei Ping**

This dissertation is submitted to

Universiti Sains Malaysia

As partial fulfilment of the requirement to graduate with honors degree in

**BACHELOR OF ENGINEERING (MECHANICAL ENGINEERING)**



School of Mechanical Engineering

Engineering Campus

Universiti Sains Malaysia

**July 2021**

## DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed..... (Tanish a/l Rajoo)

Date..... (12/7/2021)

### STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by giving explicit references. Bibliography/references are appended. Signed..... (Tanish a/l Rajoo)

Date..... (12/7/2021)

### STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available outside organizations. Signed..... (Tanish a/l Rajoo )

Date..... (12/7/2021)

## **ACKNOWLEDGEMENT**

First and foremost, my praises and thanks goes to God, the Almighty, for his blessings who has given me hope to complete my research successfully throughout this tough time as the world is going through a pandemic era. Furthermore, I would like to acknowledge my deep and sincere gratitude to my supervisor Associate Professor Dr. Loh Wei Ping for providing constant support and guidance throughout this research which has motivated me to do the research in the best possible way. She always helps me whenever I have any doubts or problems and has never given up on me amidst the hard times that the world is facing now due to covid-19 pandemic. I would also like to thank all the technical staff of School of Mechanical Engineering, Universiti Sains Malaysia (USM) for helping me carry out tasks throughout these 4 years in the university. Moreover, I would like to thank my family who have been my pillars of strength since day one enrolling in this university. Without their support I would have not been the person I was today, and I am extremely grateful to them. Last but not least, I am extremely grateful to my friends who have kept motivating me throughout this research and my university life. They have been a great support system in my university life and I am grateful for it. Thank you everyone for your constant care and encouragement towards me.

## TABLE OF CONTENTS

<b>DECLARATION .....</b>	<b>ii</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>1</b>
<b>TABLE OF CONTENTS .....</b>	<b>2</b>
<b>LIST OF TABLES .....</b>	<b>5</b>
<b>LIST OF FIGURES .....</b>	<b>6</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>7</b>
<b>LIST OF APPENDICES.....</b>	<b>8</b>
<b>ABSTRAK.....</b>	<b>10</b>
<b>ABSTRACT .....</b>	<b>12</b>
<b>CHAPTER 1 Introduction .....</b>	<b>13</b>
1.1 Overview.....	13
1.2 Project Background .....	13
1.3 Problem Statement .....	14
1.4 Objective.....	15
1.5 Scope of project.....	15
1.6 Thesis outline .....	15
<b>CHAPTER 2 Literature Review .....</b>	<b>17</b>
2.1 Overview.....	17
2.2 Search strategy .....	18
2.2.1 Extraction from the database .....	18
2.3 Distraction Driving.....	19
2.3.1 Distraction factors.....	20
2.3.2 Driving behavior .....	20
2.3.3 Driving behavior detection.....	21
2.4 Sources of Driving Distraction.....	23

2.5	Summary.....	27
<b>CHAPTER 3 Methodology.....</b>		<b>28</b>
3.1	Overview.....	28
3.2	Data Collection.....	29
3.2.1	US Dataset.....	29
3.2.2	Malaysia Dataset.....	30
3.3	Data Transformation.....	34
3.4	Data Pre-processing.....	39
3.5	Data Classification .....	45
3.6	Classification Error Performance Analysis.....	46
3.7	Summary.....	47
<b>CHAPTER 4 Results and Discussion.....</b>		<b>48</b>
4.1	Overview.....	48
4.2	Preprocessed dataset.....	48
4.3	Classification on WEKA .....	50
4.3.1	Preprocessing effect on classification .....	50
4.3.2	Individual drivers.....	51
4.3.2(a)	US dataset.....	51
4.3.3	Class and driving performance indicator .....	54
4.3.3(a)	US dataset.....	54
4.3.3(b)	Malaysia Dataset.....	56
4.4	Validation case study.....	58
4.5	Summary.....	59
<b>CHAPTER 5 Conclusion.....</b>		<b>60</b>
5.1	Concluding works.....	60
5.2	Study Contribution .....	61
5.3	Future works .....	62

**REFERENCES .....63**

APPENDICES

## LIST OF TABLES

	<b>Page</b>
Table 2-1	Driving Distraction sources reported..... 25
Table 3-1	Attributes distribution of Case study 1 ..... 31
Table 3-2	Attributes description of Case study 2 ..... 34
Table 3-3	Drivers by Gender and age segregation in US dataset..... 36
Table 3-4	Drivers by Gender and Driving Experience segregation in US dataset ..... 37
Table 3-5	The US dataset layout in .csv format ..... 38
Table 3-6	The Malaysia dataset layout in .csv format..... 38
Table 3-7	List of attributes to filter by criteria set ..... 40
Table 4-1	Driving attributes of US dataset before and after filter ..... 49
Table 4-2	Driving attributes of Malaysia dataset before and after filter ..... 49
Table 4-3	Low accuracy (< 85%) on individual drivers' classification ..... 52
Table 4-4	Classification results of three-stage attribute evaluation analyses for three levels of inconsistent driving exclusion ..... 55



## LIST OF FIGURES

	<b>Page</b>
Figure 2-1	Schematic diagram for search strategy ..... 18
Figure 2-2	Distraction sources of driving ..... 24
Figure 3-1	Research flow chart ..... 28
Figure 3-2	Distribution of data on Boxplot with Interquartile range..... 42
Figure 3-3	Removing data outliers on unsupervised instance filter in WEKA... 43
Figure 3-4	The Initial data on drive T014 ..... 44
Figure 3-5	Instances with replaced numeric values..... 44
Figure 4-1	Percentage of pre-processed US and Malaysia datasets ..... 50
Figure 4-2	Percentage difference before and after for US and Malaysia datasets..... 51
Figure 4-3	Inconsistency drivers' data..... 53
Figure 4-4	Original four classes of DriverDistractionLevel attribute compressed into three classes..... 54
Figure 4-5	Percentage classification accuracies by three-stage attributes evaluation for three levels of inconsistent driving exclusion..... 56
Figure 4-6	Original four classes of DriverDistractionLevel attribute compressed into three classes..... 57
Figure 4-7	Difference of removal a driving indicator..... 58

## **LIST OF ABBREVIATIONS**

US	United State of America
WEKA	Waikato Environment for Knowledge Analysis,

## **LIST OF APPENDICES**

Appendix A      Gantt chart for this study

**KLASIFIKASI MEMANDU BERDASARKAN KEMAHIRAN MEMANDU  
DALAM GOLONGAN WANITA**

## ABSTRAK

Memandu kereta dalam kehidupan nyata adalah contoh pelbagai tugas yang penting bagi keselamatan golongan wanita. Gangguan adalah risiko yang besar terhadap keselamatan jalan raya, bukan hanya untuk pemandu tetapi juga untuk pengguna jalan raya yang rentan. Tingkah laku yang mengganggu semasa memandu sangat popular di seluruh dunia, walaupun pada kenyataannya mereka terbukti mempunyai pengaruh negatif terhadap keselamatan pemandu. Kajian ini mengkaji kemahiran memandu secara tidak aktif pada pemandu wanita. Ini dilakukan dengan menjalankan dua set data yang berbeza, satu dari dataset AS dan satu lagi adalah dataset Malaysia. Pengalaman memandu dengan pemandu lebih dari dan sama dengan satu setengah tahun dipertimbangkan dalam kajian ini. Sebanyak 33 pemandu wanita dan 57 pemandu dipilih masing-masing dalam kedua-dua set data. Selain itu, atribut kelas yang merupakan kelas `DriverDistractionLevel` diperkenalkan dengan 3 perbezaan (Mild Distraction, Moderate Distraction and Severe Distraction). Selain itu, petunjuk prestasi pemanduan juga disenaraikan seperti Speed, Brake, Steering, LaneOffset, Lane.Position dan Time. Analisis antara atribut kajian menunjukkan bahawa pemandu wanita sangat terlibat dalam pemanduan yang tidak aktif berdasarkan kemahiran memandu mereka yang dikenal pasti oleh petunjuk prestasi memandu

**DISTRACTED DRIVING CLASSIFICATION BASED ON DRIVING SKILLS  
IN FEMALES**

## ABSTRACT

Driving a car in a real-life is a safety-critical example of multitasking. Distraction is a significant risk to road safety, not only for drivers but also for vulnerable road users. Distracting behaviours while driving are popular all around the world, despite the fact that they have been shown to have a negative influence on driving safety. This study examine distractive driving skills on female drivers. This is done through conducting two different datasets which one from US dataset and another one is Malaysia dataset. Driving experience with more than and equal to one and half a year drivers were considered in this study. Total of 33 and 57 female drivers were chosen respectively in both datasets. Additionally, a class attribute which is DriverDistractionLevel class is introduced with 3 distinct (Mild Distraction, Moderate Distraction and Severe Distraction). Furthermore, driving performance indicator also been listed which were Speed, Brake, Steering, LaneOffset, Lane.Positon and Time. The analysis among the study attributes showed that female drivers highly engaged in the distractive driving based on their driving skills, which is identified by the driving performance indicator.

# CHAPTER 1

## Introduction

### 1.1 Overview

This project studies the analysis of data concerning distracted driving conditions from two case study datasets. This analysis is done with a data mining approach. Distracted driving could be the cause of physical, visual, auditory, and cognitive distraction from the driver itself. This chapter introduces the concern on distracted driving precisely on female drivers. The driving performance indicator and pattern are discussed. The objectives of the study were presented.

### 1.2 Project Background

Traffic safety is commonly promoted through the standard three-pronged approach using engineering, enforcement, and education. Traffic safety is a primary worldwide challenge. Most traffic safety legislation is designed to prevent risky driving behavior by making it illegal to engage in these behaviors. Driving behaviors or driving performance is defined as the overall performance of the degradation during distracted driving by combining the effects of individual measures [1]. Driving performance is measured by assessing the various indicators to reflect driving behaviors on-road [2].

We are familiar with the fact that it is illegal to drive over the speed limit, fail to stop at a stop sign, or not use a safety belt. This legislation, however, cannot be effective without enforcement. More than 1.25 million fatalities on roads were reported in 2013 by the World Health Organization (WHO) [3]. Distracted driving (DD), particularly cell phone use (texting or speaking), is accounted for 26% of all crashes



and is increasing in frequency [4] (DD), particularly cell phone (texting or speaking) use, is 26% of all crashes were accounted for and the frequency is rising.

Many research works considered the gender impact on distracted driving conditions such as talking on a handheld phone, reading a text message, and checking on social media are more prone to be male whereas females are less engage in these behaviors [5]. It was evident in the past statistics that females are far safer drivers compared to males.

Nevertheless, not much emphasis was put to overcome dangerous driving performances on the road by drastically devoted to female driver's traffic safety problem.

### **1.3 Problem Statement**

There is a growing concern about different driving distractions (texting on mobile phones, listening to music, reaching for devices inside the vehicle, and focusing on events outside the vehicles) posing threat to safe driving. There were many research works on distracted and driving conditions to develop a proper and safe driving model. Not many studies have considered the distracted car driving conditions characterized by driving performance indicators specifically among females.

Besides, driving performances were commonly measured based on different quantitative indicators such as driving simulators and questionnaires. Nevertheless, it remains unclear which indicator best differentiates different driving distraction conditions.

#### **1.4 Objective**

This study aims to identify best driving performance indicators that distinguish driving conditions. Furthermore, enhance driving condition classification by consistency driving patterns.

#### **1.5 Scope of project**

The study involves an analysis of female driving behaviors experimented on a driving simulator. The drivers included have more than one and half years of driving experience. Gender and age categories are set as constant. The driving performance is measured on Speed, Brake, Steering, LaneOffset, Lane. Position, Time, Drive, and Distance attributes. Several driving conditions will be considered which include DriverDistractionLevel. These conditions will be labeled as the attribute classes. The data will be analyzed using the data mining technique using WEKA software. Classification analysis will be performed on J48, NaiveBayes, Ibk, AdaBoost, and Bagging algorithms.

#### **1.6 Thesis outline**

This thesis is structured into 5 chapters. The first chapter discusses the introduction of distracted driving that begins with the project background. In this chapter, the problem statement, objective, and scope of the project are presented.

Chapter 2 presents the literature review based on the published information from the previous study. The topics that are related to this study are extracted to explore the existing studies and the gap of analysis. The issues and challenges encountered in previous studies are considered.

Chapter 3 is related to the entire methodology processes involved in four levels of processes. The methodology processes involve data collection, data transformation, data pre-processing, data classification, and classification error performance analysis.

Chapter 4 delivers the results obtained from both global and local datasets beginning from the pre-processing of the dataset to the classification on WEKA till obtaining the best accuracy. The external validation of dataset also been done.

Lastly, Chapter 5 summarises the overall conclusion from findings obtained in this study. Study contributions and how the objectives of this study are achieved as well as the future directions of the study are discussed.

## CHAPTER 2

### Literature Review

#### 2.1 Overview

Traffic safety, a primary worldwide challenge is commonly promoted through the standard three-pronged approach using engineering, enforcement, and education. Most traffic safety legislations are designed to prevent risky driving behaviors. The common illegal driving behaviors include exceeding the speed limit, fail to stop at a stop sign, or not wearing a safety belt. More than 1.25 million roads fatalities were reported in 2013 by the World Health Organization (WHO) [6]. The majority of road accidents were related to driving behaviors of the drivers itself where usage of cell phone are most prevalent. Past studies have shown that distracted driving (DD), particularly cell phone use (texting or speaking) was accounted for 26% of all crashes and is expected to keep increasing in frequency [7]. Furthermore, statistics showed that the females are regarded as safer driver on road compared to males in that they report significantly fewer driving abbreviations. Moreover, female drivers with higher levels of education wore their seatbelts more and committed driving aberrations much less frequently than male [3].

Nevertheless, not much research emphasis was put to overcome dangerous driving behaviors. Few works were devoted to investigate the drivers safety specifically among the female drivers. Thus this chapter presents the state-of-the-art review based on the female driving patterns under distracted and non-distracted conditions.

## 2.2 Search strategy

### 2.2.1 Extraction from the database

The Science Direct platform were consulted from year 2021-2017 on search keywords: ‘Distracted female driving’, ‘Safe car driving model’, and ‘female drivers’. Figure 2.1 shows the schematic diagram of the articles search strategy. There was a total of 3068 journals discovered following the keywords. A further filtration by studying the abstract and article titles on related works based on the following criteria was performed.

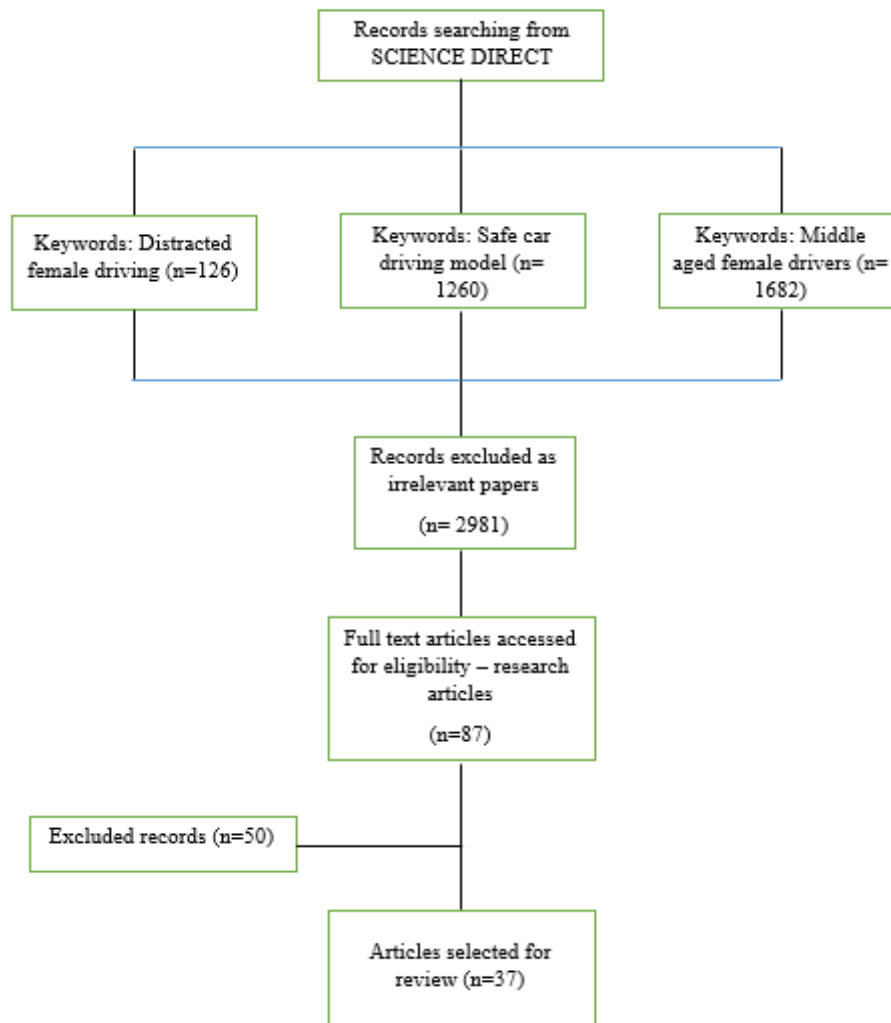


Figure 2-1 Schematic diagram for search strategy

The articles' abstracts was selected based on any of the following criteria: Containing the keyword 'female drivers'. If the article's abstract contains this keyword, then the article would be taken. Keywords such the distracted driving factors of distracted driving such as phone usage and intoxicated drivers. If keywords listed included in the abstract, then it would be taken. Furthermore, the article Described about the safe car driving method would be chosen.

The screening based on abstracts resulted in 2981 journal records. After full text articles accessed for eligibility, 87 articles remained. Meanwhile, a total of 50 articles were removed as they were not related to the main keyword search. The remaining articles were selected based on the following inclusion and exclusion criteria yielding 37 articles suitable for analysis.

**Inclusion criteria:**

- Articles were written in English language
- Research articles
- Open access full-text articles

**Exclusion criteria**

- Articles written in other languages.
- Review articles.
- Only abstracts

### **2.3 Distraction Driving**

Distracted driving is defined as a specific type of inattention. Based on [4], Distraction happens when a driver's attention is diverted from the task of driving to another activity. According to [8], the diversion of attention from activities critical for safe driving to a competing activity knows as distracted driving whereas [9] said that completing a secondary task that draws focus away from the primary task of driving could be categorize as distracted driving.

### **2.3.1 Distraction factors**

Apart from distraction driving, the surrounding and car condition also play the role to distract the drivers[8] and [10] had studied signal obstruction such as blind spot, inattention, and distraction along and inside the passenger vehicle which may cause fatal crashes. In their study, one-third of inattention-related crashes involved pedestrians whom were hit by a motor vehicle, typically because the drivers have detected the pedestrian too late. Statistics reported that the distraction by mobile phones contributed only 2-4% of all fatal crashes, while other sources of distraction within or outside the vehicle, contributed to about 10%.

According to [11], female drivers were sometimes distracted by their co-passengers. Female drivers were involved in twice as many fatal crashes alone with child passengers as compared to the male drivers. Drivers with only child passengers were more often reported as being inattentive whether female or males. This is due to potential source of distraction is the child passengers. This is because some actions or care need to be taken towards the child such as feed the child passenger or pick up the child's toys while driving). In that case, for female drivers, risk-taking behaviours were less typical than for other drivers without child passengers.

### **2.3.2 Driving behavior**

Driving behaviour reflects the past studies on driving behaviours commonly focused on developing a safe car driving model by gender and age. Recent works also studied the causes, effects, and solutions to overcome this dangerous driving behaviour.

According to [12], driving distractions are most of the times due to the driving styles. The styles considered include, for instance, the Dissociative, Angry, Reckless, Anxious, and Distress Reduction are the factors that caused the distraction specifically

in the female drivers group. This is supported by a study in Jannusch et al.[5] in which female drivers are prompt to distracted and fatigued driving.

Considering the age group, is the younger category drivers are often involved in using the handheld devices (HH) while driving as compared to the elderly drivers [5]. The young drivers are likely to be fatigued after midnight into the early morning hours, while elderly drivers are more likely to be fatigued in the afternoon. This is evident in [13]and[14] who examined drivers aged '30–60 years' and found this group to be more associated with cell phone use compared to '<30 years' and '>60 years' age group. The effects of simple conversation, complex conversation, simple texting and complex texting task while driving poses a significant collision risks on road.

Automated driving has also been one of the causes in distracted condition. Non-driving related tasks (NDRT) causes of drowsiness to the drivers [15]. On the findings, young drivers develop more drowsiness than old and middle-aged drivers. Young drivers tend to engage in secondary tasks easily and being distracted while driving and also increase the crash risk very often when compared the old-age drivers [16].

According to [17], middle-aged female drivers have higher reaction times when distracted but lower reaction in non-distracted conditions on different traffic characteristics and driving performance parameters. Moreover, when considering the type of areas, young drivers in rural areas have low reaction time in non-distracted conditions regardless of talking to a passenger or using on the cell phone [8].

### **2.3.3 Driving behavior detection**

Various applications and instruments were developed to detect driving behaviours. For instance [18] used the Multidimensional Driving Style Inventory (MDSI) to the Spanish spoken in Spain and studied the driving habits of Spaniards.



MDSI is an instrument that allows identifying not only “maladaptive” Driving Styles (DS) in order to modify them, but also “adaptive” DS to encourage safe driving. Maladaptive refers to Reckless, Angry, Anxious, Dissociative, and Distress Reduction. Adaptive is more to encourage a safer driving such as Careful Driving style. Their analyses reaffirmed that the MDSI can serve as a diagnostic instrument for sketching the profiles of the safe vs. unsafe drivers, based on the different adaptive and maladaptive driving styles. Findings from their study can be used for evaluation as well as for planning interventions directed at mitigating the deficits detected.

Meanwhile, Park[19] mainly focused on indicators to quantify the driving characteristics. The difference between observed data and the reconstructed values that indicate the representative values of low risk driving was evaluated on quantitative analysis for safe driving feedback.

Statistics have also indicated that Thailand ranks top globally in terms of automobile crashes, human factors account for most crashes. The Driving Behaviour Questionnaire (DBQ) was used as a tool to analyse the factor of crashes in Thailand [7]. In [7], two factors: texting and social media, as well as drug and alcohol consumption were used to create a measurement model for risky behaviour. The outcome of DBQ was that middle aged female drivers were prone to read and reply messages while driving.

The internet and machine learning based software called the Connected Assistant for Driving Safe (CADS) software was developed in [20]. The concept of this software is to continuously monitor the driver, road ahead and alert drivers of threats ahead. The setting up of CADS should be done on windscreen of the car. Through having The internet connection the CADS able to send the collected data to the program to process and read it as a readable data.

Only lateral performance parameters were considered in the study by [1] for detecting the distracted state of drivers. Those parameters include the number of lane excursions, steering wheel reversal rates, SDLP (Standard deviation of lane positioning), the mean and standard deviation of lateral acceleration, and mean and standard deviation of steering wheel angle.

The outcome driving behaviour detection tools developed in the pasts could help to decrease the fatal crashes by enhancing both the drivers and passengers' awareness on the road conditions. It should be noteworthy that considering the behaviour of surrounding vehicles and various factors such as road surface condition, weather, and behaviour of driving can be additions to construct a more precise model with an expanded scale.

#### **2.4 Sources of Driving Distraction**

Driving distraction can be divided into 4 types: Physical, cognitive, visual, and auditory. Physical distraction is when the driver has use one or both hands to handle the phone in order to dial a number, answer, or finish a call rather than concentrating on the physical demands of driving. Next, cognitive distraction can be explain as something that causes drivers to lose focus or takes their mind off driving such activities are conversing with passengers, focusing on an engaging news report, or reprimanding children. Furthermore, visual distraction where the drivers have taken their eyes off the road and glanced elsewhere. Those possible activities could be rubbernecking, texting, reading billboards, and looking to see what the kids are doing in the back seat are all examples of visual distractions. In addition, auditory distraction can be define as noises and sounds inside the car can distract you. For an example, if

the baby is crying or the kids are fighting in the back seat, the noise can divert the drivers' attention.

Figure 2.1 shows the distraction sources of female drivers while driving extracted from the 37 articles reviewed. There is a total of 14 factors being the major sources of driving distractions as listed in Table 2.1.

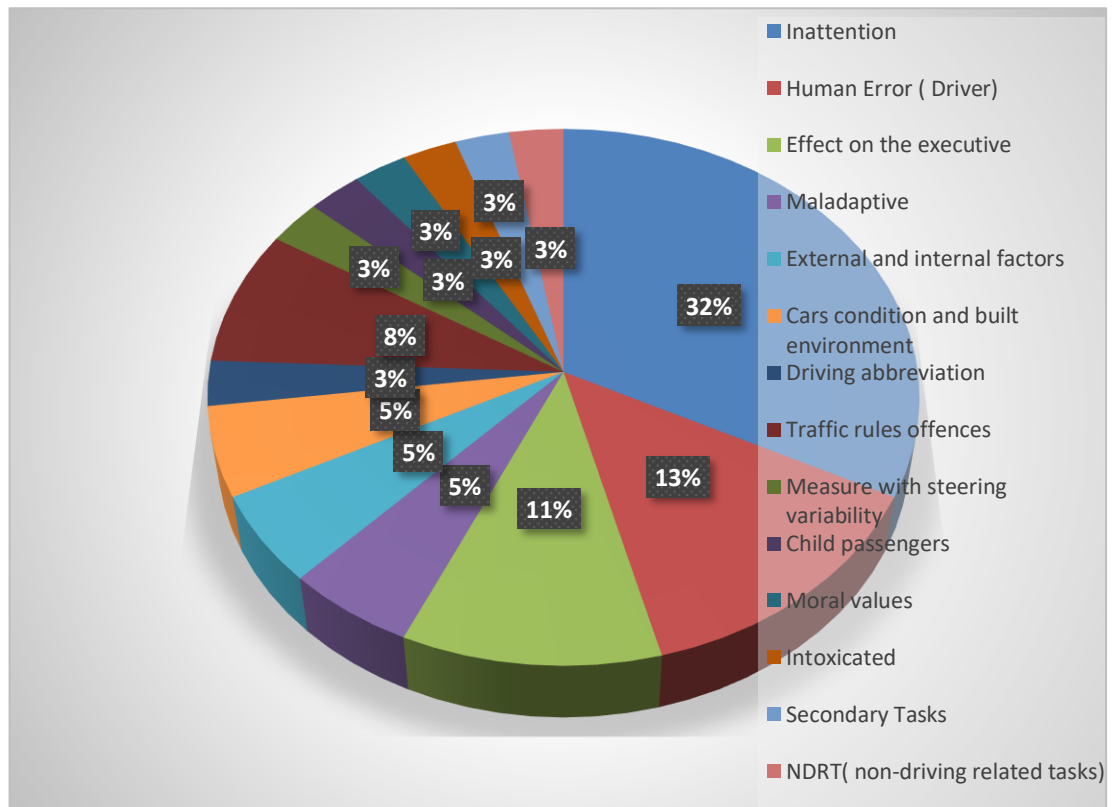


Figure 2-2 Distraction sources of driving