An Extensive Exploration of Pedestrian Safety Hazard Associated with Jaywalking on Highway Intersections in Bangladesh

DEWAN TANVIR AHAMMED

NIAZ MAHMUD

AFIA JAHIN PREMA

Department of Civil and Environmental Engineering (CEE)

Islamic University of Technology (IUT)

2020

An Extensive Exploration of Pedestrian Safety Hazard Associated with Jaywalking on Highway Intersections in Bangladesh

Dewan Tanvir Ahammed

Student ID: 160051031

Niaz Mahmud

Student ID: 160051062

Afia Jahin Prema

Student ID: 160051082

A THESIS SUBMITTED FOR THE DEGREE OF SCIENCE

IN CIVIL ENGINEERING

Department of Civil & Environmental Engineering

Islamic University of Technology

2020

APPROVAL

The dissertation entitled "***An Extensive Exploration of Pedestrian Safety Hazard Associated with Jaywalking on Highway Intersections in Bangladesh***", by Dewan Tanvir Ahammed, Niaz Mahmud, and Afia Jahin Prema has been approved fulfilling the requirements for the Bachelor of Science Degree in Civil Engineering.

Supervisor

Moinul Hossain, PhD

Professor
Department of Civil and Environmental Engineering
Islamic University of Technology (IUT)
Board Bazar, Gazipur, Bangladesh.

DECLARATION

We hereby declare that the undergraduate research work reported in this thesis has been performed by us under the supervision of Professor **Dr. Moinul Hossain** and we have taken reasonable care to ensure that this work has not been submitted elsewhere for any purpose.

Niaz Mahmud
Student ID: 160051062
November, 2020

Tanvir

Dewan Tanvir Ahammed
Student ID: 160051031
November, 2020

Afia Jahin Prema

Afia Jahin Prema
Student ID: 160051082
November, 2020

ACKNOWLEDGEMENTS

Alhamdulillah for everything. We can never thank Allah enough for the countless bounties He blessed us with in finishing this thesis project.

The first round of thanks goes to our supervisor Professor Dr. Moinul Hossain for his outstanding style of supervision. His guidance, support and suggestions regarding this research help us so much to improve the work and finish it on time. His most valuable contribution is to make us think widely, largely and make transportation engineering interesting to us.

We would also like to thank Dr. Md Asif Raihan, Assistant Professor, Bangladesh University of Engineering and Technology (BUET), for his valuable comments and information from his experience.

Last of all, our parents deserve a big round of applause for supporting us in the whole year long in finishing the thesis project. Without their support it is impossible for us to complete the thesis project.

ABSTRACT

Pedestrian safety is a major concern around the world as pedestrians are the most vulnerable road users. Pedestrian fatalities account for 22% of all road traffic fatalities around the world. The statistics are even grimmer for the developing countries where jaywalking is predominant. There, along with jaywalking, the use of electronic gazettes, especially cell phones, while crossing the road is considered as a triggering factor in acerbating pedestrian casualties and fatalities. This study takes highway intersections of Bangladesh as the study area and delves into thought processing of jaywalkers and pedestrians using cell phones while crossing roads to devise countermeasures for improving pedestrian safety. The study observes pedestrian behavior at 32 intersections on national and regional highways of Bangladesh through video data and subsequently interviews 2,016 pedestrians found jaywalking and/or using cell phones while crossing the road. During this process, data on their socio-economic and demographic characteristics, various risk perceptions, physical obstructions that may have forced jaywalking, distracting cell phone activities, road crossing behavior as well as their knowledge about basic rules of the road were collected. Next, a Bayesian Belief Networks (BBN) was constructed to answer 'who', 'why' and 'how' related questions regarding jaywalkers and pedestrians who use a cell phone while road crossing. The findings suggest that jaywalking is more predominant among males, aged between 26-40 years who have received secondary education despite having decent knowledge regarding basic rules of the road. The most influential factors concerning risky jaywalking and using cell phone while road crossing are 'Gender', 'Jaywalker Activities', 'Waiting Time', 'Types of Jaywalking', and 'Frequency of phone use while jaywalking'. The results of the study revealed that an appropriate, clean and hygienic way for passengers significantly reduces the amount of jaywalk. In the end decision-makers will develop pragmatic safety policies based on the high impact variables associated with the established jaywalking process and the factors that are caused by mobile jaywalking.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

Chapter 1
INTRODUCTION

1.1 Background:

Road transport gives all nations and the public benefits by promoting transport of goods and individuals. It provides improved access to work, industry, education, leisure and health services, which, in turn, directly and indirectly have positive effects on public health. However, increasing road transport has also dramatically affected human wellbeing in the form of car accidents and physical activity loss health consequences. The movement of people and goods has economic, social and environmental consequences that are more harmful.

Road traffic crashes resulting in casualties and fatalities are of major concern around the world and existing patterns demonstrate that this is likely to persist for the foreseeable future. The global problem of road accidents is gradually seen as a significant public health concern. Although in many high-income countries mortality rates have stabilized or declined over recent decades, it is evident that in most of the world's regions the epidemic of global traffic accidents continues to escalate. Approximately 1.35 million people are the victims of traffic-related fatalities each year, and in between 20 to 50 million people throughout the world are the sufferers of road traffic injuries. Traffic tragedy has now become the 8th major cause of mortality all over the world. Vulnerable road users, including pedestrians, pedal cyclists, and motorcyclists, account for more than half of these road traffic fatalities around the world with pedestrians being considered as the most vulnerable road users (WHO, 2018).

Pedestrian road safety has become a major global public health issue that has already gained extensive attention in recent times. Pedestrian deaths constitute nearly a quarter of all road traffic fatalities with the proportion reaching high as one half in African countries (WHO, 2018). Moreover, it has affected developing countries the most where road traffic crashes have become a direct reflection of the combined effect of rapid growth in population, motorization, widespread urbanization, economic expansion along with a lack of resources to enhance overall road safety. Nevertheless, most of the road networks in developing countries are not well-prepared for vulnerable road users. The minimization of vulnerable road users' hazards is, therefore, a growing need in these countries.

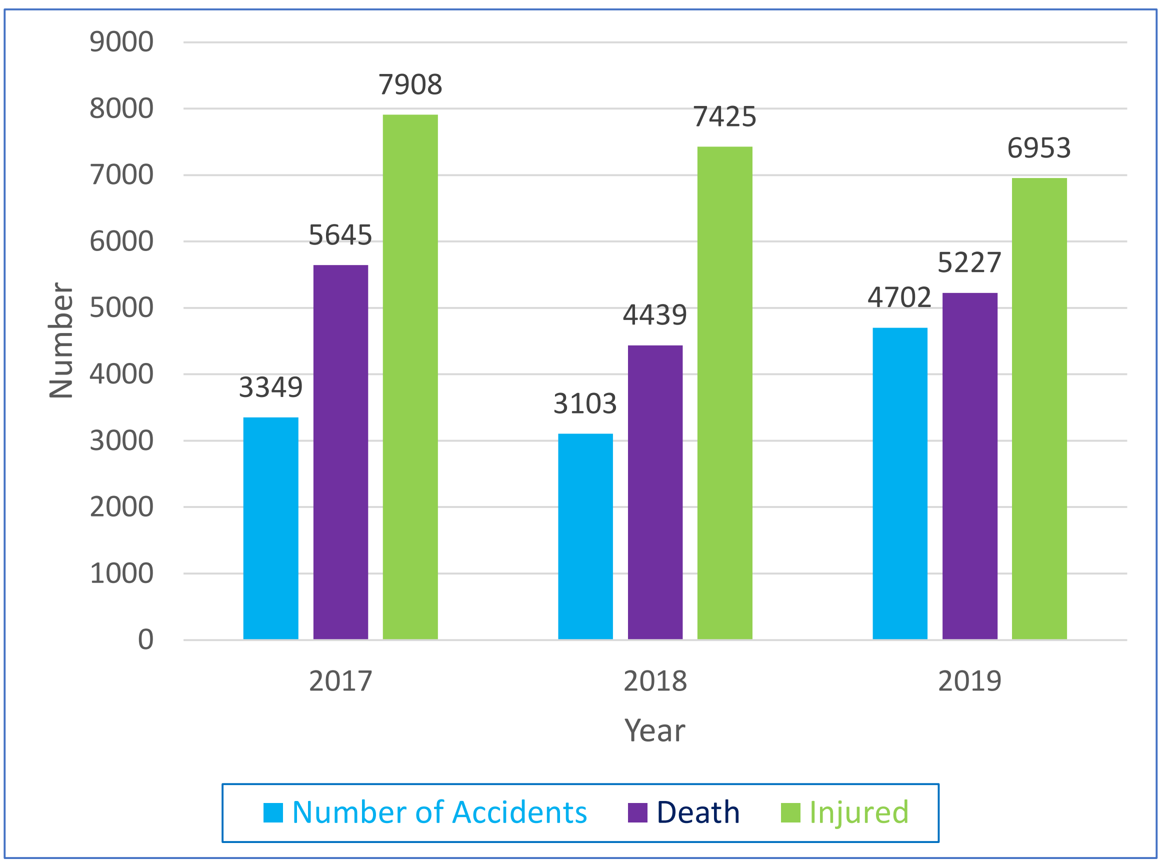

Figure 1.1: Road Crashes in Bangladesh.

The average fatality rate of low-income countries due to road traffic crashes is 27.5 for each 100,000 population, whereas in high-income countries the average rate of death is 8.3 per 100,000 population (WHO, 2018). However, while there has been a significant improvement in the reduction of road traffic injuries and fatalities in developing countries over the last few years, more than 21,000 people in Bangladesh have been victims of traffic-related fatalities every year. In Bangladesh, statistics indicate that vulnerable road users account for about 45% of all road traffic fatalities. According to the Global Status Report on Road Safety, approximately 32% of the victims of road traffic fatalities in Bangladesh are pedestrians which is around 10% higher than the global average (WHO, 2015). Figure 1.1 indicates that, while death and injury have decreased over the years, the number of accidents has increased. From Figure 1.2, pedestrians have come across the largest amount of fatalities (WHO, 2019). The consequences of such road traffic catastrophe led to increased spending in health and social care, and loss in economic prosperity due to unexpected deaths and physical disabilities. From an economic point of view, road crashes in Bangladesh are costing the society approximately 2% of gross domestic product (GDP) (Hoque, Anowar, & Raihan, 2008).

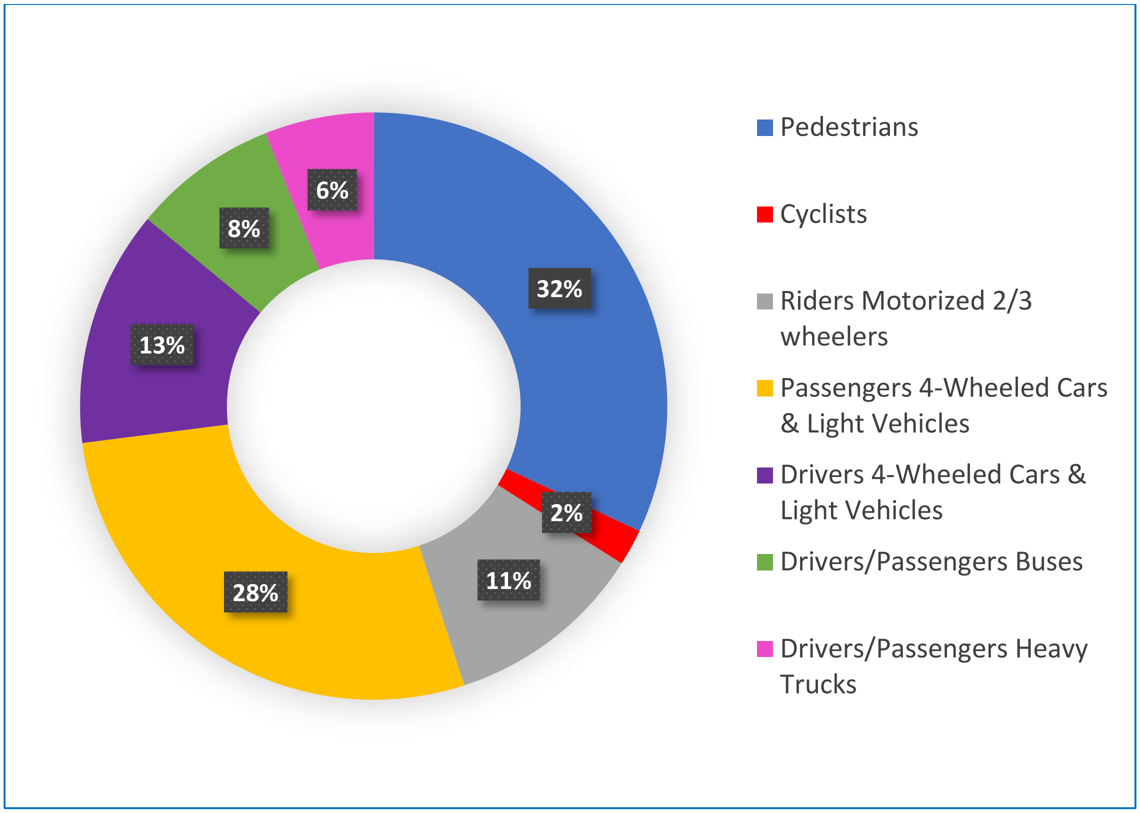

Figure 1.2: Road Accident Fatalities.

There have been notable variations in the figures concerning road accidents and casualties for Bangladesh in the last 13 years (1998-2010). From Table 1.1, accidents rose by about 12% between 1998 and 1999. After they fell rapidly in 2001 by 26% compared to 2000, but again peaked in 2003 (4114 accidents). The information is generally available and the compilation of precise data is incomplete (e.g., exact location, road user movement involved) (ARI, 2007).

Table 1.1: Number of Accidents and Casualties.

Year	Number of Accidents					Number of Casualties			
	Fatal	Grievous	Simple	Collision	Total	Fatal	Grievous	Simple	Total
1998	2000	1137	193	203	3533	2358	2313	984	5655
1999	2437	986	305	220	3948	2893	2168	1301	6362
2000	2523	1029	209	209	3970	3058	2270	1215	6543
2001	2029	642	137	117	2925	2388	1661	904	4953
2002	2599	904	200	238	3941	3053	2155	1130	6338
2003	2752	921	239	202	4114	3334	2421	1319	7074
2004	2509	683	216	158	3566	3150	2027	999	6176
2005	2424	631	142	125	3322	2960	1830	740	5530
2006	2695	602	124	145	3566	3250	1705	707	5662
2007	2923	705	166	160	3954	3341	1783	648	5772
2008	2842	676	154	128	3800	3570	1752	664	5986
2009	2161	474	71	109	2815	2703	1438	308	4449
2010	1911	387	62	77	2437	2443	1271	435	4149
Total	31805	9777	2218	2091	45891	38501	24794	11354	74649

From Table 1.2, 49% of all deaths recorded in the accident database were pedestrians. Pedestrians accounts for 62% of traffic accidents in urban areas. Inclination in road crash deaths of pedestrians can be seen from 43% in 1986-87 to 74% in 1998-2010. 50% pedestrians' causalities occur in urban areas during road crossing whereas 52% pedestrians' causalities can be seen while walking roadside in rural areas. In comparison, the greatest injuries to passengers are grievous and simple injuries (63%) (ARI, 2007).

Table 1.2: Driver, Passenger and Pedestrian Casualties by Year.

Year	Driver Casualty				Passenger Casualty				Pedestrian Casualty			
	Fatal	Grievous	Simple	Total	Fatal	Grievous	Simple	Total	Fatal	Grievous	Simple	Total
1998	315	537	263	1115	848	1204	645	2697	1195	572	76	1843
1999	397	401	304	1102	1079	1266	863	3208	1417	501	134	2052
2000	414	432	287	1133	1212	1341	826	3379	1432	497	102	2031
2001	344	298	185	827	859	1016	659	2534	1185	347	60	1592
2002	421	391	239	1051	1059	1272	821	3152	1573	492	70	2135
2003	499	455	267	1221	1243	1525	980	3748	1592	441	72	2105
2004	443	322	249	1014	1168	1338	676	3182	1539	367	74	1980
2005	392	311	161	864	1077	1126	501	2704	1491	393	78	1962
2006	482	286	160	928	1121	1056	486	2663	1647	363	61	2071
2007	477	310	166	953	1001	1101	405	2507	1863	372	77	2312
2008	557	321	224	1102	1308	1084	360	2752	1705	347	80	2132
2009	487	278	109	874	1019	891	163	2073	1197	269	36	1502
2010	392	234	108	734	964	820	300	2084	1087	217	27	1331
Total	5620	4576	2722	12918	13958	15040	7685	36683	18923	5178	947	25048

The minimization of pedestrian safety hazards founded on a comprehensive understanding of the potential causative factors is of growing importance. Pedestrian fatalities and casualties are perhaps predictable and preventable with a number of established strategies because of the existing studies have been enthusiastic to investigate the pedestrian safety related issues. Pedestrian behavioral patterns of road crossing potentially accountable for traffic crashes which mostly jeopardize safety (Shaaban, Muley, & Mohammed, 2018), (Rosenbloom, 2009), (Lord, Cloutie, Garnie, & Christoforou, 2018). Additionally, several recent studies have particularly emphasized on the influential factors of illegal road crossings at intersections (Marisamynathan & Vedagiri, 2018), (Shiwakoti, Tay, & Stasinopoulos, 2017), (Dommes, Granié , Cloutier, Coquelet, & Huguenin-Richard, 2015), (Mukherjee & Mitra, 2019). Furthermore, in recent years, the extensive use of electronic devices, in particular the use of mobile phones resulting in distracted crossing behavior which exhibits a potential threat to pedestrian safety (Hatfield & Murphy, 2007), (Alsaleh, Sayed, & Zaki, 2018), (Zhang, Zhang, Wei, & Chen, 2017).

1.2 Problem Statement:

Despite the extensive literature on pedestrian road crossing behavior and pedestrian safety, very few of these studies have explored the potential triggering factors of pedestrian safety associated with jaywalking and mobile phone induced distractions at intersections in developing countries. Most of the existing studies have emphasized on pedestrians' traffic law violations due to illegal road crossing (crossing at red light and crossing outside of crosswalks). The pedestrian walkway is generally found to be congested, unhygienic, and unsafe in developing countries like Bangladesh. Hence, pedestrians are sometimes forced to jaywalk along the road.

In most of the developed countries, people are aware of jaywalking and its consequences. Also, the level of consequences of jaywalking while engaged in various mobile phone induced distracting activities is high, therefore, the number of indifferent people towards the risk of it is very less. On the other hand, the people of developing countries like

Bangladesh, have a lack of knowledge regarding jaywalking and pedestrians are often found to jaywalk while engaged in various mobile phone induced distracting activities.

Hence, within this study, an attempt has been executed to study pedestrians' safety perceptions towards the danger of jaywalking and cell phone induced distracted jaywalking behavior together.

1.3 Purpose and Objectives:

The present study attempts to identify the potential risk factors of both illegal road crossing by the pedestrians as well as jaywalking along the road at national and regional highway intersections. The explicit objectives of this study are as follows:

i. To explore the pedestrians' illegal road crossing behavior and jaywalking along the road at intersections.

ii. To unveil underlying determinants and their interconnections associated with jaywalking and cell phone induced distractions at highway intersections in the context of developing countries.

iii. To identify the socio-demographic characteristics of pedestrians based on their level of awareness regarding jaywalking.

1.4 Scope of the Study:

Number of intersections need to be increased from 32 to much more to get better insight about pedestrians' behavior. Men and women participant ration needs to be equal. Some other questions need to be added in the questionnaire i.e., number of trips, income etc.

1.5 Organization of the Thesis:

The thesis is organized into five chapters and each chapter consists of several sub-chapters. These chapters were divided based on the various activities that were done during this study. The sequence of the chapters is commensurate with the sequence in which the activities were performed.

Chapter 1: Introduction

This Chapter contains the main idea of this thesis. It is divided into six sub-chapters which are as follows:

1.1 Background

1.2 Problem Statement

1.3 Purpose and Objectives

1.4 Scope of the Study

1.5 Organization of the Thesis

Chapter 2: Literature Review

This chapter is the written overview of major writings and other sources on the selected topic. The sub-chapters are-

2.1 Introduction

2.2 Illegal Road Crossing

2.3 Mobile Phone Induced Distraction

2.4 Summary

Chapter 3: Methodology

Methodology is a key part of this dissertation. It describes the broad philosophical underpinning to the chosen research method. It has following sub-chapters-

3.1 Introduction

3.2 Study Area and Data Collection

3.3 Descriptive Statistics

3.4 Bayesian Network

Chapter 4: Analysis and Results

This chapter deals with the statistical methods used to evaluate the comfort model and check the validation of the created model. The sub-chapters are-

4.1 Introduction

4.2 Model Preparation

4.3 Model Validation

4.4 Analysis and Result of Developed BBN

Chapter 5: Conclusion

This is the final chapter that contains the findings and future scope of the research. It's divided into four sub-chapters-

5.1 Introductions

5.2 Policy Implementation

5.3 Limitations and Future Scope

Chapter 2
LITERATURE REVIEW

2.1 Introduction

In the last few decades, due to the enormous number of deaths of pedestrians, the need for a proactive road safety system has increased a significant amount. The behavior of pedestrians has been given significant importance by transport researchers as a new method that promotes protection for pedestrians. A large number of investigations have also been carried out with the primary objective of understanding the cause of jaywalking. Most analysis is carried out in the developed world. As the main purpose of this study is to understand the triggers of jaywalking in the developing world, the current studies in this area of road safety are discussed in this chapter. The chapter also provides a summary of their use for road safety.

2.2 Illegal Road Crossing:

Ma et al. (2020) introduced a new methodology to explore pedestrian illegal road crossing strategies as well as the influential factors of traffic law violations at signalized intersections. The authors developed two Bayesian networks. Data on pedestrian demographic and socio-economic characteristics along with traffic and road conditions were collected using an online questionnaire survey. Video-graphic surveillance was conducted at three signalized intersections to observe illegal crossing characteristics. Age (31~45 years), income (<3000 Yuan), education (high school), medium waiting time (20~60 s), medium vehicle volume (250~550pcu/h∗lane), moderately hurried condition, medium pedestrian volume (35~60ped/circle), crossing distances (3~4 lanes) were found to be the most significant factors of crossing at a red light.

Dommes et al. (2015) performed a study among the adult pedestrians combining observational and questionnaire data to explore the potential contributing factors of traffic law violations associated with safety at signalized intersections. The authors established a stepwise logistic regression model to identify the behavioral indicators of safety at road crossings. Age, driving experience, crowd size, and traffic density as well as nearby parked vehicles were found to be the most influential factors of safety. The number of companions at the intersection and parked vehicles also accelerated the likelihood of illegal street crossing.

Marisamynathan and Vedagiri (2018) studied pedestrian road crossing strategies at six signalized intersections to explore the influences of key determining factors on illegal road crossing behavior associated with pedestrian safety in the context of Mumbai in India. Pedestrians' demographic characteristics, various risk perceptions, crossing volumes, crossing performance, crossing time, compliance behavior, crossing location, red light duration, group size, mobile phone usage while crossing, type of crossing (single stage or two-stage crossing) were collected from video-graphic surveillance along with questionnaire surveys. In this study, most of the pedestrian violated traffic law to save time. Combined analysis of Pearson correlation test and odd ratio (OR) statistics publicized that education level, occupation, as well as trip purpose potentially affected pedestrians' crossing behavior. The authors also developed a binary logit model to predict the illegal road crossing strategies at signalized intersections which had 65% predictive abilities.

Mukherjee and Mitra (2019) investigated pedestrian safety-related crossing behavior with more precision and to identify the influential factors of traffic law violations at signalized intersections. Information on road conditions, traffic conditions, pedestrians' socio-demographic characteristics, and various risk perceptions were extracted from questionnaires, field surveys, and video surveillance. The outcomes of the Multinomial Logit Model (MNL) revealed that pedestrians' age, education level, purpose of trip, home location, waiting time at the intersection promoted the likelihood of traffic law violation.

Ren et al. (2011) utilized ANOVA and OR statistics to analyze pedestrian crossing performances and triggering factors associated with traffic law violations at signalized intersections. Data on pedestrian demographic characteristics, various risk perceptions, group size, crossing location, crossing strategies, and crossing time were collected by video surveillance and questionnaire survey. The outcomes revealed that gender, age, and crowd size potentially accelerated the rate of unlawful crossing behavior at signalized intersections.

Shiwakoti et al. (2017) investigated the potential influencing factors of jaywalking at an intersection in Australia. Video observational data didn't show any significant gender differences in the mean number of jaywalkers per signal cycle. Social interactions and mobile phone-induced distractions were found to be the most contributing factors of jaywalking at an intersection.

Brosseau et al. (2013) used a logistic regression model to explore the influencing factors as such waiting time, arrival time, and the presence of a pedestrian signal on the pedestrian propensity of traffic law violation and risky crossing at a signalized intersection. The study revealed that maximum waiting time directly sped up the likelihood of traffic law violations.

Priyadarshini and Mitra (2018) applied a Poisson Regression Model to investigate fatal pedestrian crashes at unsignalized intersections. The findings exposed that wider carriageway, narrow zebra-crossing, higher pedestrian volume, lower post encroachment time were the most significant influencing factors associated with pedestrian-vehicle conflict.

Mamun et al. (2018) explored the road crossing strategies with a wide range of demographic data. Women, age group below 25 years, married people, people who drive often, people who seldom walk, and people who have kids exposed safer crossing behavior before intervention at intersections.

Gong et al. (2019) extracted seven influencing factors (age, gender, crosswalk length, crossing time, headway, red light duration, and countdown display) of illegal road crossing behavior at signalized intersections in the context of Lanzhou City in China. In this study, male and elderly pedestrian were more involved in illegal crossings. Longer waiting time along with no countdown display revealed an adverse effect on pedestrian traffic law violation.

Rosenbloom et al. (2009) compared road crossing behavior between individuals and groups of pedestrians at an intersection. Data on pedestrian demographic characteristics, traffic flow characteristics as such traffic volume, duration of red light, duration of green light was collected using an interactive environment. In this study, male pedestrians exhibited more tendency to violate traffic law than female pedestrians. The outcomes of logistic regression also indicated that the intention to cross at red light increased with the presence of fewer waiting people at the curb.

Zhang et al. (2016) examined the influencing factors associated with pedestrians' red-light traffic law violations at intersections in China. In this study, data on pedestrians' socioeconomic characteristics, trip characteristics such as trip purpose, time requirement, waiting time, and pedestrian's attitudes towards traffic law violations were collected using an interactive questionnaire survey. A binomial logistic model demonstrated that trip purposes, time period in a day, and pedestrians' attitudes towards traffic law violations were responsible for pedestrians' red-light crossing behavior.

Zhang et al. (2017) used a logistic regression model to extract the motivating factors associated with pedestrian safety at unsignalized intersections in China. Age, mobile phone-induced distracting activities, and traffic volume was found to be positively correlated with pedestrian crossing safety. Pedestrians under 30 years (about 64%) had a higher tendency to use mobile phones while crossing at unsignalized intersections. In this research, distracted pedestrians (approximately 3 out of 4) were engaged in pedestrian-vehicle collisions.

Shaaban et al. (2018) investigated pedestrian illegal crossing behavior on a major high-speed six-lane divided arterial road. Using four video cameras, road crossing characteristics such as the crossing point, waiting time, the number of unsuccessful attempts, walking style, pedestrian speed change, crossing type, pedestrian path, crossing time, crossing distance as well as traffic-related characteristics (vehicle yielding, gap type, vehicle lane, safety distance, critical distance, and the number of rejected gaps) were collected during the three stages (before crossing, during the crossing, and after crossing) of illegal road crossing activity. Descriptive statistics revealed that middle-aged male pedestrians were more involved in illegal crossings. The authors employed a multiple linear regression (MLR) model to identify the significant influencing factors associated with waiting and crossing time. Crowd size and crossing points were found to be significantly correlated with waiting time. On the other hand, crossing time was affected by gender, age, crossing in a group, and the use of a mobile phone.

2.3 Mobile Phone Induced Distraction:

Chen et al. (2018) examined the impacts of mobile phone-induced distracting activities on road crossing behaviors at a signalized intersection in the context of Taipei City in Taiwan. A total of 1995 distracted pedestrians' demographic characteristics as well as various distracting activities such as talking, texting, gaming, internet browsing, and listening to music were collected using an interactive environment. The outcome of the study pointed towards the fact that smartphone gaming while road crossing potentially increased the likelihood of pedestrian-vehicle collisions.

Hatfield and Murphy (2007) performed a comparative study between male and female pedestrians to visualize the impacts of mobile phone use on pedestrian safety. A total of 270 females and 276 males crossing performances were recorded from an observational field survey, conducted in three Sydney suburbs intersections. In this research, both males and females exhibited slower walking speed while being engaged in phone conversations. The findings also suggested that mobile phone-induced distracting activities adversely affected pedestrian safety.

Alsaleh et al. (2018) applied automated video analysis to explore the influences of mobile phones on pedestrians' walking strategies. The analysis of a total of 357 pedestrians (36 distracted and 221 non-distracted) walking behavior revealed that visual and cognitive distracting activities significantly reduced the walking speed which tends to undermine safety.

Lennon et al. (2017) presented mobile phone-induced distracted pedestrian characteristics in Queensland, Australia. Pedestrians between 18–30 year old exhibited higher frequency to cross roads in distracted conditions.

Nasar and Troyer (2013) studied the relationship between pedestrian casualties and mobile phone use. The authors used US Consumer Product Safety Commission data from 2004 through 2010. Analysis of the data revealed that mobile phone disturbance considerably increased pedestrian casualties and fatalities. Pedestrians under 31 years old were more involved in mobile phone related casualties.

Sobhani and Farooq (2018) illustrated pedestrians' road crossing performances using a virtual environment. A Multinomial Logit (MNL) model revealed that distracted female pedestrians exhibited more risky crossing behavior.

Schwebel et al. (2012) applied a semi-immersive environment to examine the impacts of visual and cognitive distracting on pedestrian safety. Logistic regression analysis revealed that texting and listening to music while street crossing undermines pedestrian safety.

Nasar et al. (2008) compared the performance of crossing behavior between mobile phone-induced distracted and non-distracted pedestrians in a real environment. In this study, distracted pedestrians exposed more risky behavior than non-distracted pedestrians.

Jiang et al. (2018) examined the influences of mobile phone induced distracting activities such as texting, listening to music, and talking on pedestrian safety. Text-messaging was found to be the most detrimental factor associated with safety. Talking, and listening to music also adversely encouraged pedestrians' risky behavior.

2.4 Summary:

Relevant studies have highlighted pedestrians' behavioral patterns of road crossing which is accountable for pedestrian-vehicle crashes at mid-block as well as at intersections. Additionally, several recent studies have particularly emphasized on the influential factors of illegal road crossings at intersections. Furthermore, in recent years, the extensive use of electronic devices, in particular the use of cell phones resulting in distracted crossing behavior which exhibits a potential threat to pedestrian safety. Nevertheless, in developing countries like Bangladesh, pedestrians are often found to jaywalk along the roads because of congested, unsafe, unhygienic, and encroached pedestrian walkway. Hence, this study concentrated on pedestrians' jaywalking behavior. Furthermore, within this study, an attempt has been executed to study pedestrians' awareness towards the threat of jaywalking and mobile phone induced distracted jaywalking behavior together.

Chapter 3
METHODOLOGY

3.1 Introduction:

The thesis aims to present the Bayesian Belief Network (BBN) as a medium to foretell the way pedestrian jaywalk. It includes an advanced approach aimed specifically at identifying factors that influence pedestrians to do walking. A comprehensive knowledge of BBN is needed in order to build such a model. There are also many modeling methods and algorithms used in this study to achieve the basic goals. This chapter demonstrates both methods of data collection and the experimental setup adopted. After that, a thorough BBN discussion will take place. The algorithms used for structural learning and batch learning algorithms will then be discussed. And then, before the analysis is carried out, the variables used for the thesis and the model development methods are defined. Figure 3.1 represents a step-by-step development procedure of the proposed model:

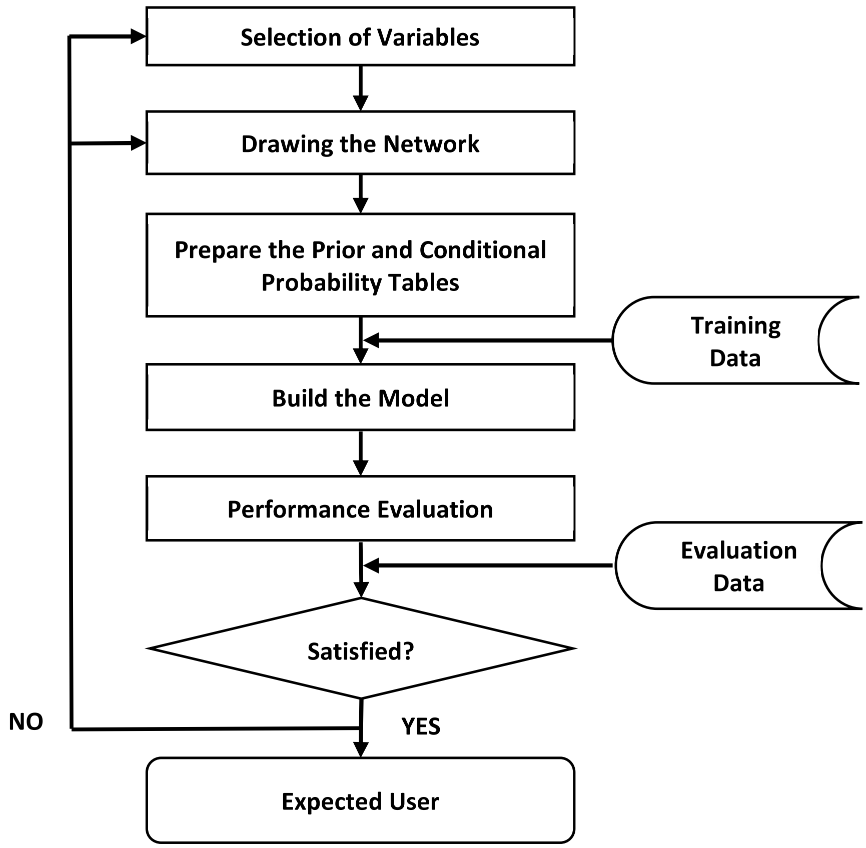

Figure 3.1: Flow Chart of Model Development.

This chapter will select variables and set conditions to draw the network in accordance with the flow chart. We will discuss how we deal with BBNs and how data is collected. In the last chapters, the other steps of the flow chart will be covered.

3.2 Study Area and Data Collection:

3.2.1 Study Area:

Pedestrians' walking and street crossing behaviors were observed in 32 intersections at both national and regional highways using video cameras. The cameras were set up in suitable positions in the intersection locations to record the pedestrians' movements. All pedestrians who were found to perform jaywalk were considered as participants of this survey. After each jaywalking event and the distracted street crossing was completed, the pedestrian was approached and interviewed by expert surveyors about demographic information, handset specifications, etc. Before the main survey, a pilot test was conducted to check whether the respondents were having difficulty in understanding the questions, and then necessary corrections were made accordingly.

Figure 3.2: Survey location map of 32 intersections.

3.2.2 Data:

A questionnaire for the jaywalking survey was formulated based on literature review and transportation expert judgments. The questionnaire included socio-demographic questions (age, gender, education level, occupation) as well as their perception of not following the rules and their deliverance about how the consultant can encourage more people to follow the rules. A total of 20 variables were included under the topics of demographic information, perception of jaywalkers, distraction, jaywalking characteristics, and potential risk. The intersections were selected after careful consideration so that the data collected from these intersections represent most of the population all over the country. Data screening was conducted to expurgate the incomplete and unengaged responses. After the screening process, the final data containing 2016 pedestrian responses were obtained. Jaywalkers were asked in questionnaire surveys about their understanding of the risk of jaywalking.

3.3 Descriptive Statistics:

Descriptive statistics of this study is presented in Table 3.1.

Table 3.1: Descriptive Statistics for Jaywalking.

Attribute	Attribute Category	Frequency	Percentage (%)
Age	a. 10-25 years	322	23.8
	b. 26-40 years	850	62.9
	c. 41 years and Older	180	13.3
Occupation	a. JOB	325	24

	b. BUSINESS	418	30.9
	c. STUDENT	249	18.4
	d. Laborer	287	21.2
	e. OTHERS	73	5.4
Gender	a. Male	1292	95.6
	b. Female	60	4.4
Education Level	a. No education	127	9.4
	b. Primary	270	20.0
	c. Secondary	659	48.7
	d. Graduate	296	21.9
Purpose of journey	a. Education	172	12.7
	b. Office	280	20.7
	c. Business	354	26.2
	d. Home	173	12.8
	e. Medical and Religious Purpose	58	4.3
	f. Others	315	23.3
Driving Experience	a. Yes	876	64.8
	b. No	476	35.2
Jaywalking speed	a. Slow	201	14.9
	b. Normal	960	71.0
	c. Fast	191	14.1
Waiting Time	a. 0-1 minute	1076	79.6
	b. 1-2 minute	229	16.9
	c. more than 2 minute	47	3.5
Types of Jaywalking	a. Grouped or Herd Jaywalking	151	11.2
	b. Jaywalking alone	1201	88.8

Merged with other jaywalkers or jaywalk independently?	a. Merged with other jaywalkers	97	7.2
	b. Jaywalked independently	1255	92.8
Jaywalker Activities	a. Talking	290	21.4
	b. Texting or Reading or Listening to something with headphones	153	11.3
	c. Not using phone	909	67.2
Frequency of phone use while jaywalking	a. Very rarely	742	54.9
	b. Sometimes	610	45.1
Cell phone Screen size	a. Less than 4 inch	513	37.9
	b. 4 to 5.5 inch	645	47.7
	c. Larger than 5.5 inch	194	14.3
Frequency of visit that area	a. Daily	947	70.0
	b. Weekly	312	23.1
	c. New or Couple of times a month	93	6.9
Availability of mobile data?	a. Yes	1084	80.2
	b. No	268	19.8
Risk Perception	a. Positive	987	73
	b. Negative	365	27

3.4 Bayesian Network:

This research employed Bayesian network to understand the pedestrian safety challenges considering jaywalking and cell phone induced distraction. Bayesian networks, also popularly recognized as Bayesian belief networks, belief networks, or causal networks are composed of a directed acyclic graph (DAG) with parent nodes and child nodes as well as a conditional probability table to symbolize and analyze uncertainty using Bayesian inference (Pearl, September 1988). Before proceed to the Bayesian network, it is important to know how conditional probability is used for modelling concepts.

Bayesian probabilities are the assumption of probabilities that talk about partial beliefs. And the Bayesian estimate calculates the validity of the preposition. The validity of the preposition is calculated on the basis of two things. They are:

1. Prior Estimates,
2. New relevant evidence.

Focusing on the posterior Bayesian estimation, the best approach to this is an essential theorem called Bayes Theorem.

Bayes' theorem, named after Thomas Bayes, an 18th-century British mathematician, is a mathematical criterion for assessing conditional probability. Conditional probability is the likelihood that an outcome will occur on the basis of a prior outcome. Bayes' theorem provides a mechanism to revise existing predictions or theories based on new or alternative evidence. Suppose we provide different possible competing hypothesis, and we want to find out the probabilities of individual hypothesis, based on the data, so that we can find out which one is the most probable or most likely hypothesis.

Bayes rule: $p(A|B) = \frac{p(B|A)p(A)}{\mathrm{p(B)}}$

Here,

$p(A)$ = prior probability of hypothesis A

$p(\mathrm{B})$ = prior probability of training data B

$p(A|B)$ = probability of A given B (posterior density)

$p(B|A)$ = probability of B given A (likelihood of B given A)

According to the theorem, probability of hypothesis given data by P(B/A) times prior probability of the hypothesis A divided by P(B). here, P(A) is the prior probability and P(B/A) is the probability of the data. For example, Covid-19 tests are common nowadays, but some results of tests are not true. Let's assume; a diagnostic test has 99% accuracy and 60% of all people have Covid-19. If a patient tests positive, what is the probability that they actually have the disease?

$$p(covid19|positive) = \frac{p(positive|covid19)p(covid19)}{p(positive)}$$

Here,

P(positive|covid19) = 0.99

P(covid19) = 0.6

P(positive) = 0.6*0.99+0.4*0.01=0.598

$p(covid19|positive) = \frac{0.6*0.99}{0.598}$ = 99.33%

Now the question arises that, if the hypothesis is true, then what is the likelihood that the data will be generated? If A is true what is the probability of B being generated and is P (B) the likelihood of the data? The most likely hypothesis can be identified, which is called the maximum posterior hypothesis. So if the conditional probability is known, we can use the Bayes rule to establish the reverse probabilities.

Now come to the Bayesian belief network that is a representation of a probabilistic model where several variables are conditionally independent. Thus, "Bayesian belief networks provide an intermediate approach that is less constraining than the global assumption of conditional independence made by the naive Bayes classifier, but more tractable than avoiding conditional independence assumptions altogether." (Machine Learning Mastery, 1997, p. 184).

It is a type of probabilistic graphical model. "A graph comprises nodes (also called vertices) connected by links (also known as edges or arcs). In a probabilistic graphical model, each node represents a random variable (or group of random variables), and the links express probabilistic relationships between these variables." ((Machine Learning Mastery, 2006, p. 360).

The Bayesian Network is a restrictive model, where the edges of the graph are directed, indicating that they can only be managed to navigate in one direction. It means that cycles are not possible and that the structure can be more simply referred to as a directed acyclic graph (DAG).

Components of Bayesian Networks: "A Bayesian belief network describes the joint probability distribution for a set of variables." (Machine Learning Mastery, 1997, p. 185)

- Structure of the graph $\longleftrightarrow$ Conditional independence relations.

In general,

P (X1, X2,…..,Xn) = ∏ P(Xi | parents(Xi))

The full joint distribution   The graph-structured approximation

- Requires that graph is acyclic.

- Two components:

  - The graph structure (conditional independence assumptions),

  - The numerical probabilities (for each variable given its parents).

Representation of a Bayesian model:

Nodes: Random variables of interest in the domain (X1, X2, X3, ……). There are mainly two types of nodes. They are:

1. Parent Node,

2. Child Node.

Arcs: Represent relations between the variables. The lack of arc denotes the lack of relation between the variables in certain ways that represents conditional independence.

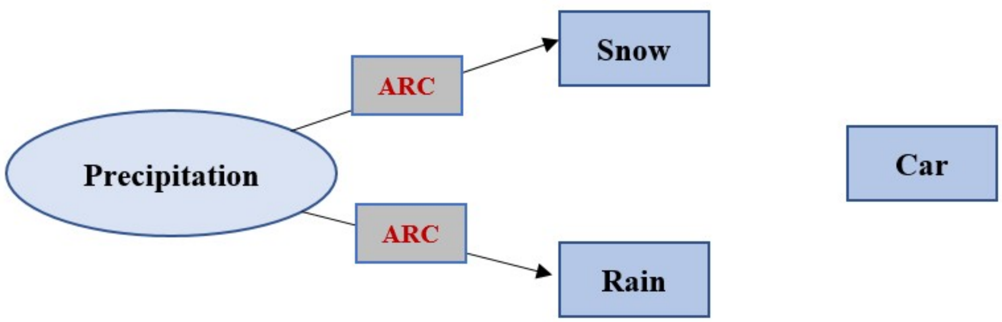

Figure 3.3: Relation between nodes and arcs.

Here, precipitation is the causes of snow and rain, so precipitation is the parent node of snow and rain. One the other hand, snow and rain are the child node of precipitation.

There is no relation between precipitation and car, so it represents there is no conditional dependency.

Causality: Some of the variables can be causes of the other, some may be effects of other. Bayesian network can be represented without causality but the representation of causality makes the structure of the Bayesian network more efficient. Bayesian networks are particularly strong in their ability to capture causality and intuitively attractive interfaces (Murphy, 1998) that helps to ensure effective communication between statisticians and non-statisticians. (Airoldi, 2007)

Inference:

Illustration of a Bayesian Network:

- The acyclic graph below shows two possible independent causes of computer failure. The two causes of this banal example are assumed to be independent. Electricity failure and computer malfunction both are ancestors and parent nodes of computer failure, On the other hand, computer failure is a descendant and a child node of both electricity failure and computer malfunction.

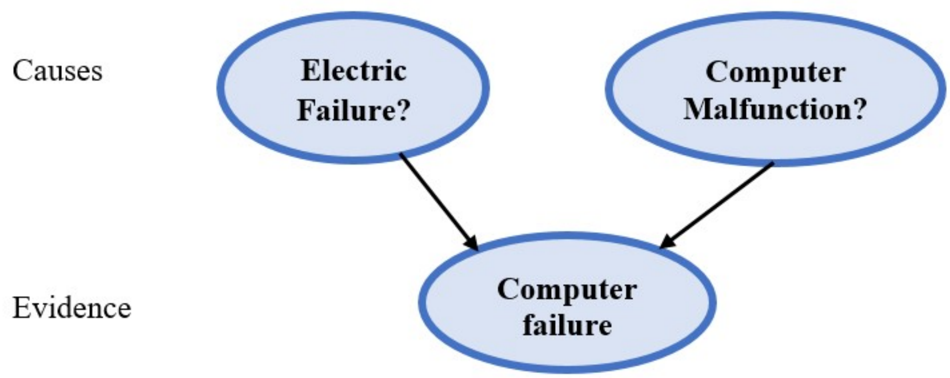

Figure 3.4: Illustration of a Bayesian Network.

Suppose that electricity failure, denoted by E, happens with a probability of 0.1, P [E = yes] = 0.1, as well as a computer malfunction, denoted by M, occurs with a probability of 0.2, P [M = yes] = 0:2. It is logical to assume that electricity failure and computer malfunction have always been independent. Furthermore, it is predicted that if there is no problem with electricity and the computer does not have a malfunction, the computer operates smoothly. In other words, if C indicates a computer failure, then P [C = yes |E = no; M = no] = 0. Since there is no electrical problem but the computer has a malfunction, the probability of computer failure is 0.5, P [C = yes | E = no; M = yes] = 0.5. Eventually, if the electricity is switched off, the computer will not operate regardless of its potential malfunction, P [C = yes | E = yes; M = no] = 1 and P [C = yes |E = yes; M = yes] = 1. The probability of computer failure P [C = yes] can be calculated as follows:

$$P\,[C = \text{yes}] = \sum_{E,M} p[C = yes, E, M]$$

$$= \sum_{E,M} (p[C = yes|E, M].\,p[E].\,p[M])$$

$$=0.19$$

Before we mention any evidence, the probability P [C = yes] = 0.19 as a prior probability of computer failure.

- Two independent possible causes of computer failure are expressed by a directed graphical model with a prior probability distribution, i.e. before any evidence is observed.

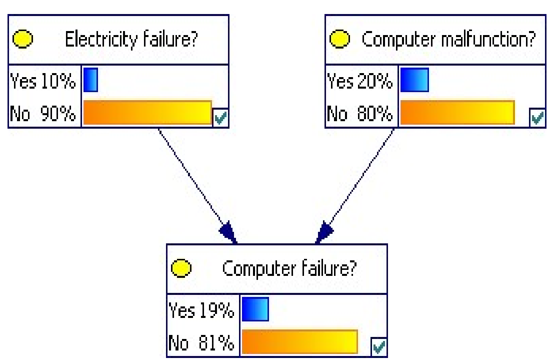

Figure 3.5: Prior probability of computer failure.

Suppose we tried to turn the computer on now, but it didn't start. In other words, we observe C = no with probability 1 and wonder how the probability distribution of electricity failure E and computer malfunction M has changed due to the evidence observed. Using the Bayes formula, we're going to find

P [E = yes | C = yes] $= \sum_M P[E = yes,\ M|C = yes]$ = 0.53

P [M = yes | C = yes] $= \sum_M P[E = yes,\ M{=}yes|C = yes]$ = 0.58

- Directed graphical model, representing two independent possible causes of computer failure with posterior distribution of probability, after observation of evidence.

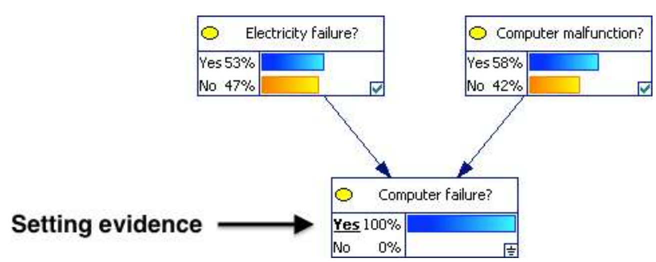

Figure 3.6: Setting evidence.

Assume that the example is extended by incorporating another piece of evidence into the model, specifically a light failure L. Light failure is assumed to be independent of computer malfunction. Again as, if the electricity is switched off, the light will not flourish under any circumstances, P [L = yes | E = yes] = 1. If there is no problem with electricity, we still claim a 0.2 chance that the light will go off, P [L = yes |E = no] = 0.2. Using the same algorithm as before, we obtain a prior probability of P [L = yes] = 0.28.

- Directed graphical model, representing two independent possible causes of computer failure, is a single possible cause of light failure with prior probability distribution, i.e., before any evidence is observed.

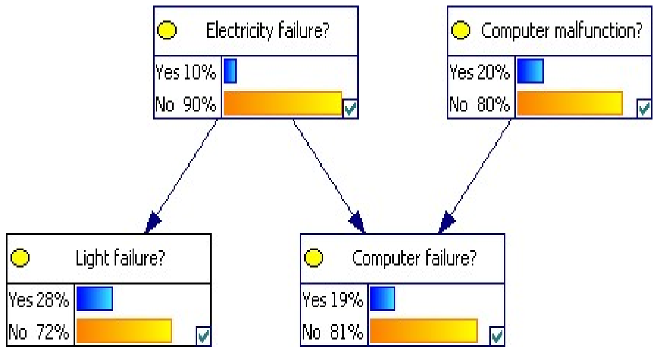

Figure 3.7: Two independent causes of computer failure without setting evidence.

Now, after observing evidence for all four combinations of light failure and computer failure outcomes, the changes in posterior probability will be observed.

- Directed graphical model, representing two independent possible causes of computer failure, is a single possible cause of light failure with prior probability distribution, i.e., after any evidence is observed.

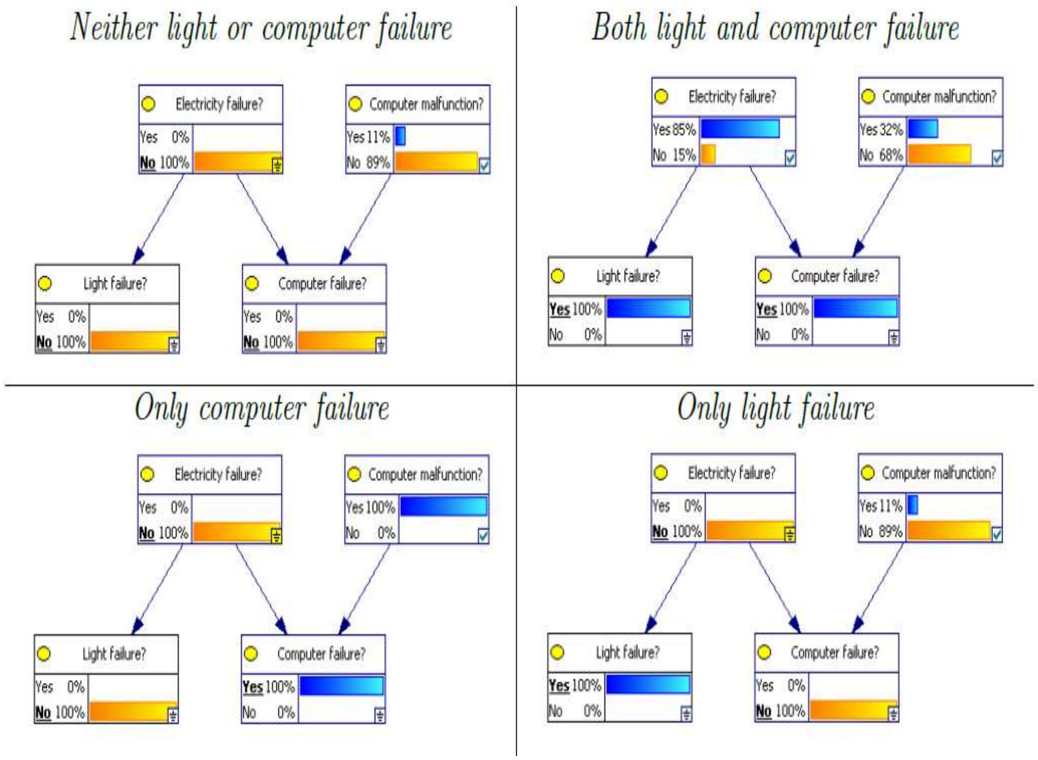

Figure 3.8: Different possible scenario

In practice, the Bayesian networks are considerably more complex than our example. It is therefore important to note that each node in the graph should be connected with at least one edge to another node. Otherwise, the isolated node is independent of all the remaining nodes and thus there is no need to take this node into consideration.

Construction of Bayesian network: There are two ways to construct a Bayesian network. They are:

1. Manual construction,
2. Automatic learning.

Manual construction: Manual construction of the Bayesian network relies on the assumption prior expert knowledge of the underlying domain. The first step is to construct a directed acyclic graph, followed by a second step to evaluate the distribution of the conditional probability in each node.

Directed acyclic graph: The construction of the directed acyclic graph begins with the identification of the relevant nodes (random variables) and the structural dependence between them. (Cowell RG, 1999) (Lucas PJ, 2004) (Airoldi, 2007). It is not necessary to observe all variables; in fact, some random variables can define unobserved quantities that are thought to affect the measurable outcomes. Data, latent variables, and parameters are all treated as nodes in the graph in the same way The underlying conditional probability distribution must be understood, or at least assumed. Since the Bayesian approach assumes that all unknown quantities are random variables, it's only natural to include parameters, as well as all latent variables and theoretically measurable quantities, as nodes in a graph. The next step is to sketch the network, (Airoldi, 2007) taking relationships among the random variable into account, (Lucas PJ, 2004) .The graph structure is usually based on substantive knowledge, although model criticism and revision are often essential, (Spiegelhalter, 1988).

Conditional probability distribution: The constructed directed acyclic graph has to include conditional probability distributions for every node in the graph, (Lucas PJ, 2004). If the variables are discrete, this can be represented as a table (multinomial distribution), which lists the probability that the child node takes on each of its different values for each combination of values of its parents. If the conditional probability distribution is not available, other statistical methods may be applied to derive this conditional distribution from data (e.g., empirical conditional probability distribution/frequencies estimation). Possible computational methods are outlined e.g., in (Spiegelhalter, 1988), or (Lucas PJ, 2004). At this point, the Bayesian network is fully specified. However, it is necessary to perform a sensitivity analysis before the network can be used in real-life application, (Lucas PJ, 2004). The sensitivity analysis may be performed either as one-way deterministic sensitivity analysis (i.e., varying one parameter at a time over a specified range), or as a probabilistic sensitivity analysis (i.e., varying all parameters of the network at once over a specified probability distribution).

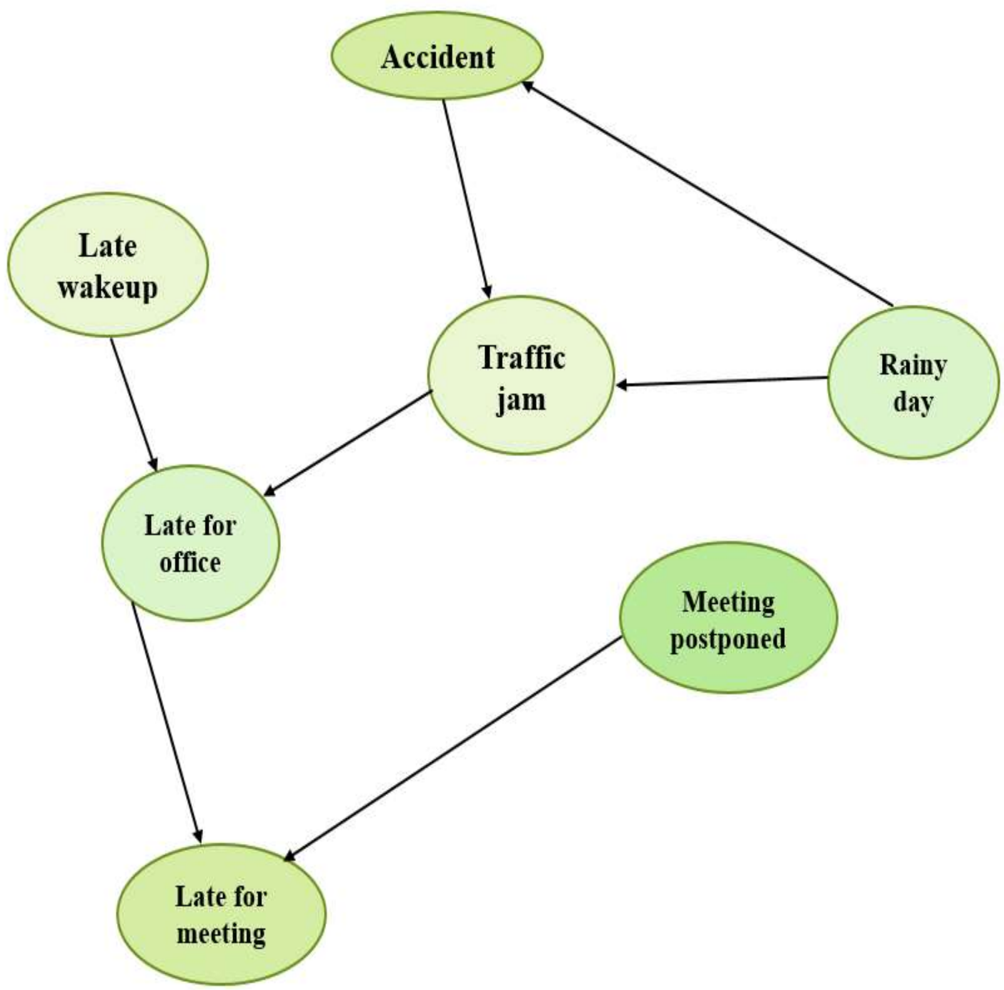

Figure 3.9: Representation of Bayesian Network.

In this example, let us say, walking up late can influence late for office. And if there is accident, it will be more likely there will be a traffic jam. Again, if the day is rainy, it will also likely to be a traffic jam. We might say that if it is rainy day, there is higher likelihood for an accident. If one passenger is caught in a traffic jam, it affects whether he is late for office. Late for office affects whether he is late for meeting and if the meeting is postponed that also influence late for meeting.

So, from this network, a directed acyclic graph is established that represents different conditional independence relationship. For example, late for office is influenced as when one has woken up late and it directly influence late for meeting. Therefore, it seems that if one wakes up late that affects whether he is late for meeting. However, if it is known that the person is late for work or not. Suppose, he is not late for office in spite of waking up late then if he is woken up late or not does not influence late for meeting. So, waking up late and being late for meeting are not normally independent but they are conditionally independent given that it is known to us that he is late for work. Again, accident and wake up late are two independent variables. But if it is known to us that he is late for office, then these two variables do not remain independent. So, this particular graphical representation encodes certain conditional independence relationships.

Automatic learning: In contrast to manual construction, automatic learning does not require expert knowledge of the underlying domain. Bayesian networks can be learned automatically from databases using experience-based algorithms that are often integrated into appropriate software. However, the disadvantage is that more data requirements are placed on automatic construction. Most automatic learning algorithms do not require missing data in the data set, which is often a very strong assumption in practice. If data is missing from the dataset, it must be imported, imputed or estimated from other sources, (Lucas PJ, 2004). In addition, sufficient data must be provided to meet the algorithm's requirements for reliable estimates of conditional probability distributions. Conditional probability distributions are assumed to be a priori known for manual construction. Automatic learning involves the creation of a network structure and the estimation of conditional probability distributions. Several network learning algorithms are discussed in the literature, for example in (Lucas PJ, 2004).

Software Used: There are a number of options for useful software to deal with graphical models. Genie, Hugin, BUGS and R are the most common packages. For our research purpose, Genie 4.2 is used.

GeNIe: GeNIe Modeler is an environment for the development of graphical decision-making models. It was established and developed at the Decision Systems Laboratory at the University of Pittsburgh between 1995 and 2015.

Using GeNIe for Bayesian Networks: For automatic learning, the database must be produced into the program by File-> Open Data File... or File-> Import ODBC data. The preferred algorithm can be selected under the Network->Algorithm option. Additional functionality of the Genie package include, for example, sensitivity analysis, strength of influence, or the calculation of the probability of total evidence.

Structural Analysis: One of the key elements of maximum likelihood estimation is the ability of directed probabilistic graphs to describe the causal structure of the observed domain. The structure itself is very meaningful and an important source of knowledge. Models built using the GeNIe can be examined conceptually. The structure itself, viewing the strengths of the influences and pathways through the graph, is an important element of this analysis. In this section we have to import the dataset and select the algorithm.

There are many learning algorithms. They are:

- Bayesian Search
- PC
- Greedy Thick Thinning
- Naïve Bayes
- Augmented Naïve Bayes
- Tree Augmented Naïve Bayes.

PC algorithm is used for our study purpose.

Parameter Learning: The method of studying the distributions of a Bayesian network or Dynamic Bayesian network using data is known as parameter learning. The Expectation Maximization (EM) algorithm is used by Bayes Server to perform maximum likelihood estimation, and it supports all of the following:

- Learning both Bayesian networks and Dynamic Bayesian networks. (e.g. Learning from Time Series data).
- Learning both discrete and continuous distributions.
- Learning with missing data (discrete or continuous).
- Learning on multiple processors.
- Learning a subset of nodes/distributions.
- Learning with noisy nodes.
- Advanced initialization algorithm.

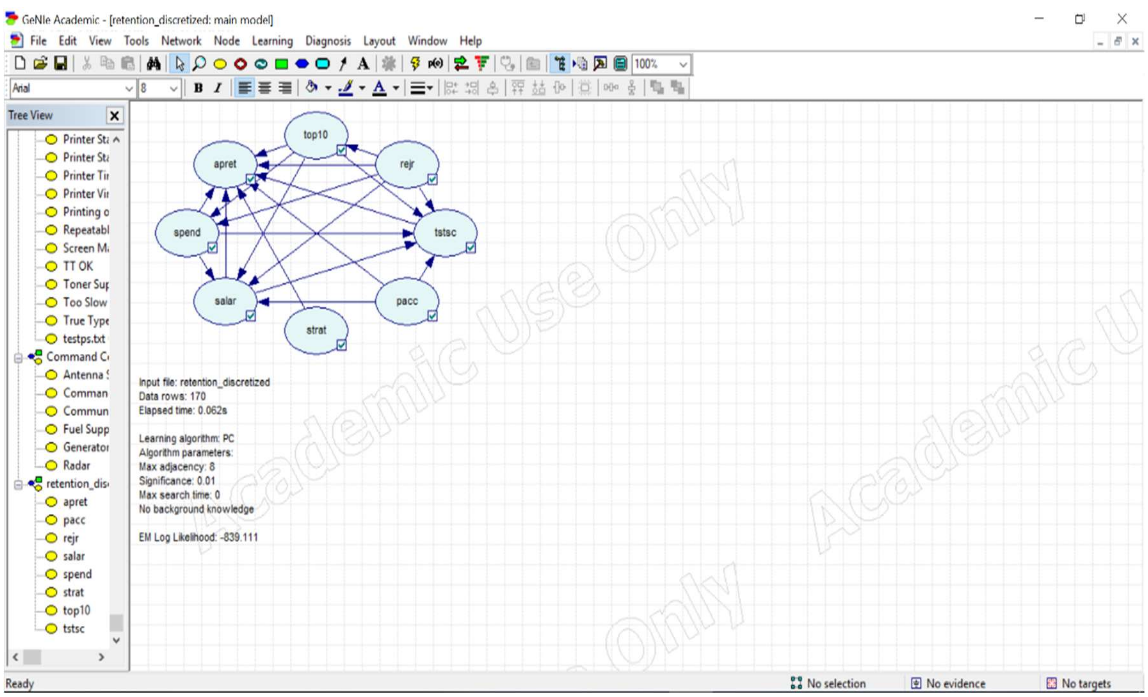

Figure 3.10: Parameter learning

Prior Probability: Prior probability is referred to the target node that is the prior one and the proportion of occurrence of a given target class relative to the other target state is the prior likelihood of that target class. For example, a probability distribution representing the relative proportion of jaywalkers who will do jaywalk always, in future they will do the same in all situation. Using the prior probability Bayes' Theorem calculates the probability function to calculate the posterior probability, the unpredictable estimation given the data. Prior probability can be calculated by previous experiment, historical data. When no data is available an uninformative prior is used to adjust the outcomes.

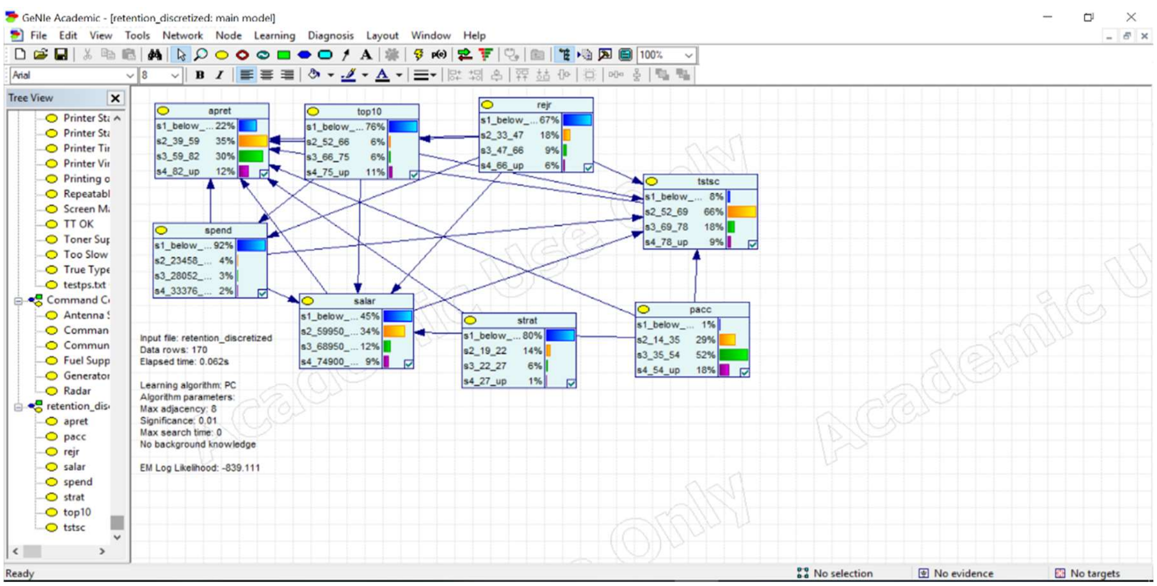

Figure 3.11: Prior Probability.

Posterior Probability: The posterior probability of a random occurrence or an unknown proposition is the conditional probability assigned after the appropriate data or context is taken into consideration in Bayesian statistics. In this context, "posterior" means "after taking into consideration all relevant facts related to the particular case under investigation."

The posterior probability distribution is the probability distribution of an unknown quantity that is viewed as a random variable that is based on data from an experiment or survey.

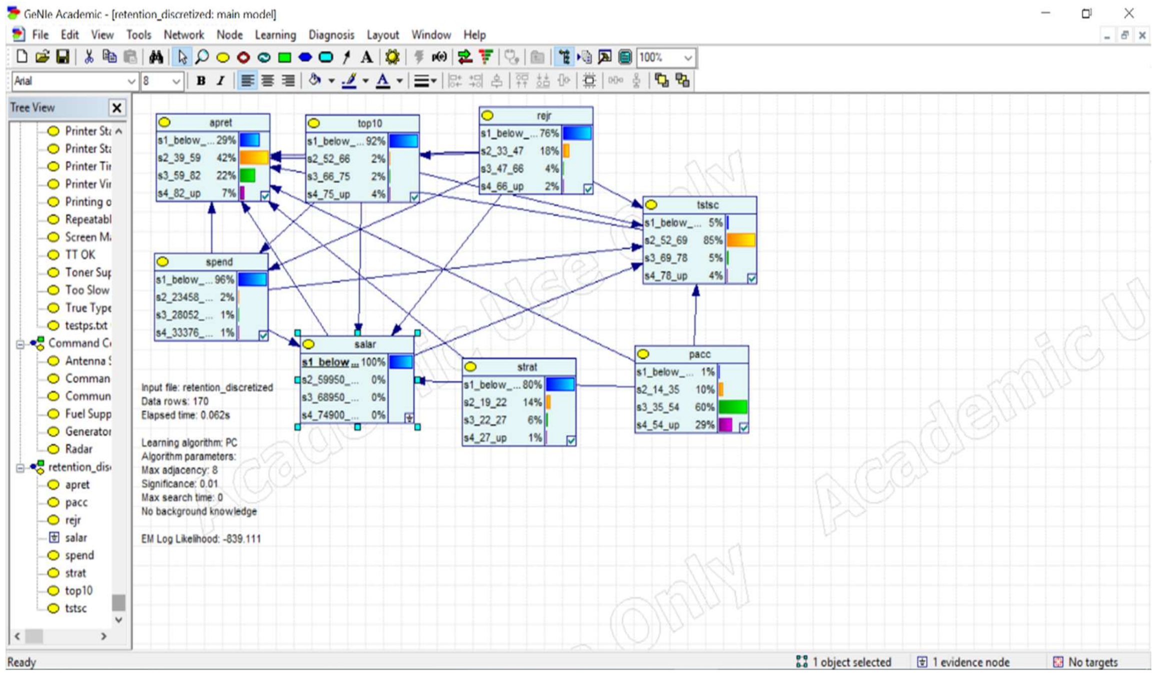

Figure 3.12: Posterior Probability.

Probability of evidence: One of the useful estimates in a probabilistic model is the probability of evidence. In this section a question is asked that "How likely is this set of observations within this model?". Then the likelihood of evidence is being chosen to run the probability of evidence estimate.

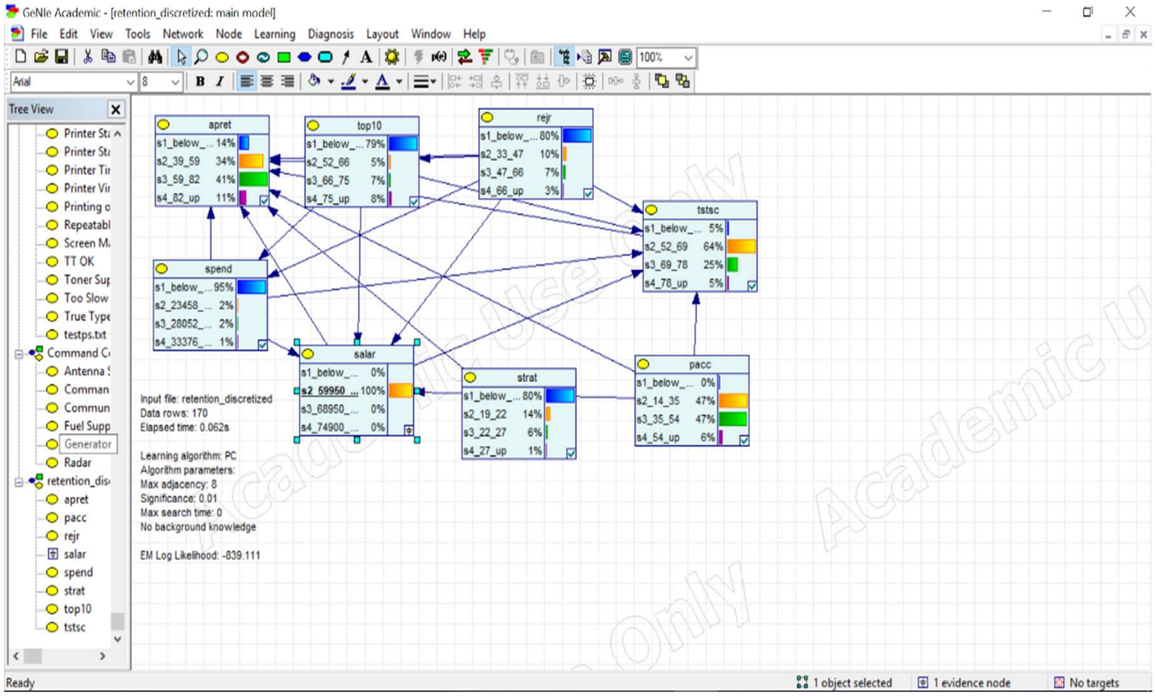

Figure 3.13: Probability of Evidence.

Sensitivity analysis in Bayesian networks: A sensitivity analysis is a systematic technique for visualizing and quantifying the effects of small changes in parameters on target node parameters. The Tornado diagram shows the most important parameter for a targeted node state, in which the color of the bar indicates the degree of change in the target state, with red representing negative changes and green representing positive changes.

Chapter 4

ANALYSIS & RESULTS

4.1 Introduction

This chapter focuses on the in-depth analysis of the model structure and the theoretical aspect of the study. The previous chapter, entitled "Methodology" provides an elaborate illustration of the various techniques and algorithms employed for the model construction along with the different cases. The contents of this chapter are intended to further expand the previous understanding and direct the review process. In various sub-sections, the required steps involved in the analytical section will be illustrated. GeNIe academic BBN model building software tool is used in this analysis. This chapter outlines the working steps of the GeNIe. The research results will be addressed. Priority will also be given to evaluating the model results and the various types of models. This chapter will include a review of the accuracy of the models. Finally, remarks are addressed on the model's success and the extent of achievement.

4.2 Model Preparation:

At first, data after screening process, was imported in GeNIe. The imported dataset is illustrated in Figure 4.1. Then network is learnt as automatic learning is used in this research. Structural learning is done from a complete data set using the PC algorithm with a significance threshold of 0.01. The network obtained is adjusted according to the correlation between variables, literature review, and engineering judgment and Figure 4.3 represent that. The built-in EM algorithm is used to produce marginal probabilities of all nodes. Figure 4.4 represents marginal probability of the network structure.

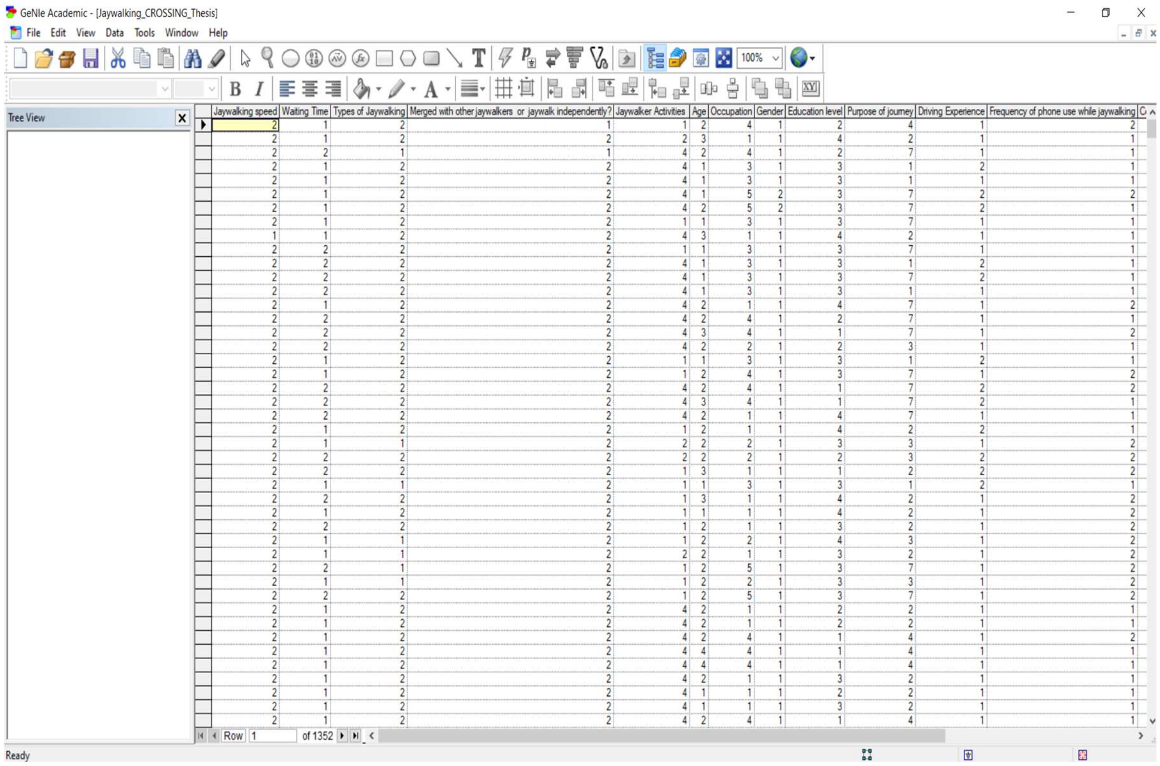

Figure 4.1: Dataset.

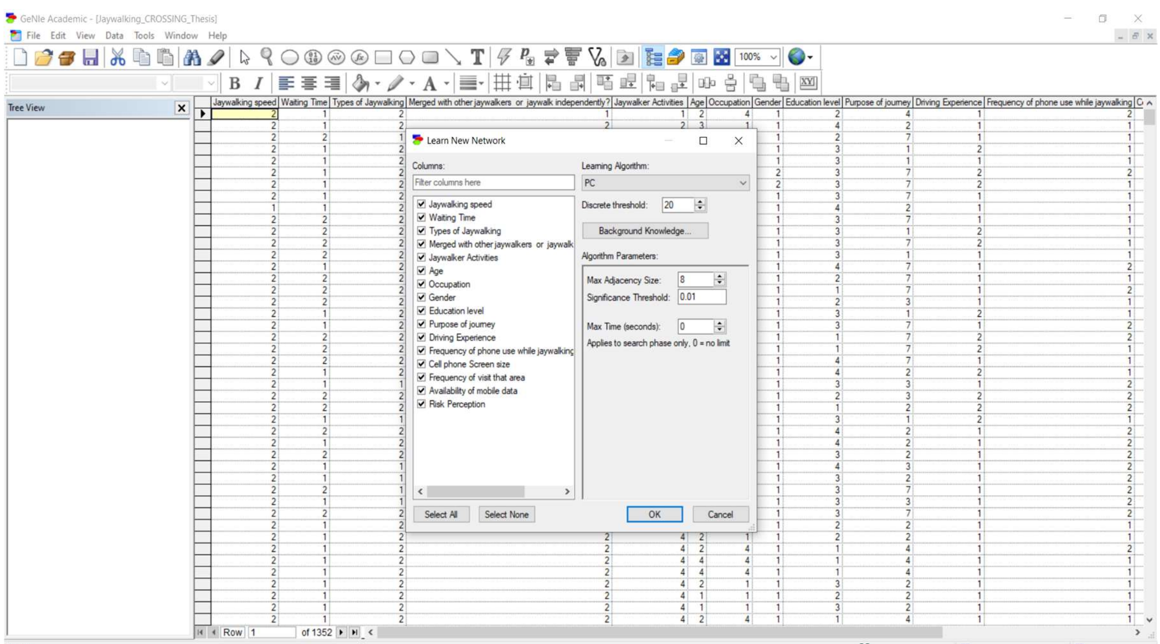

Figure 4.2: Learning New Network.

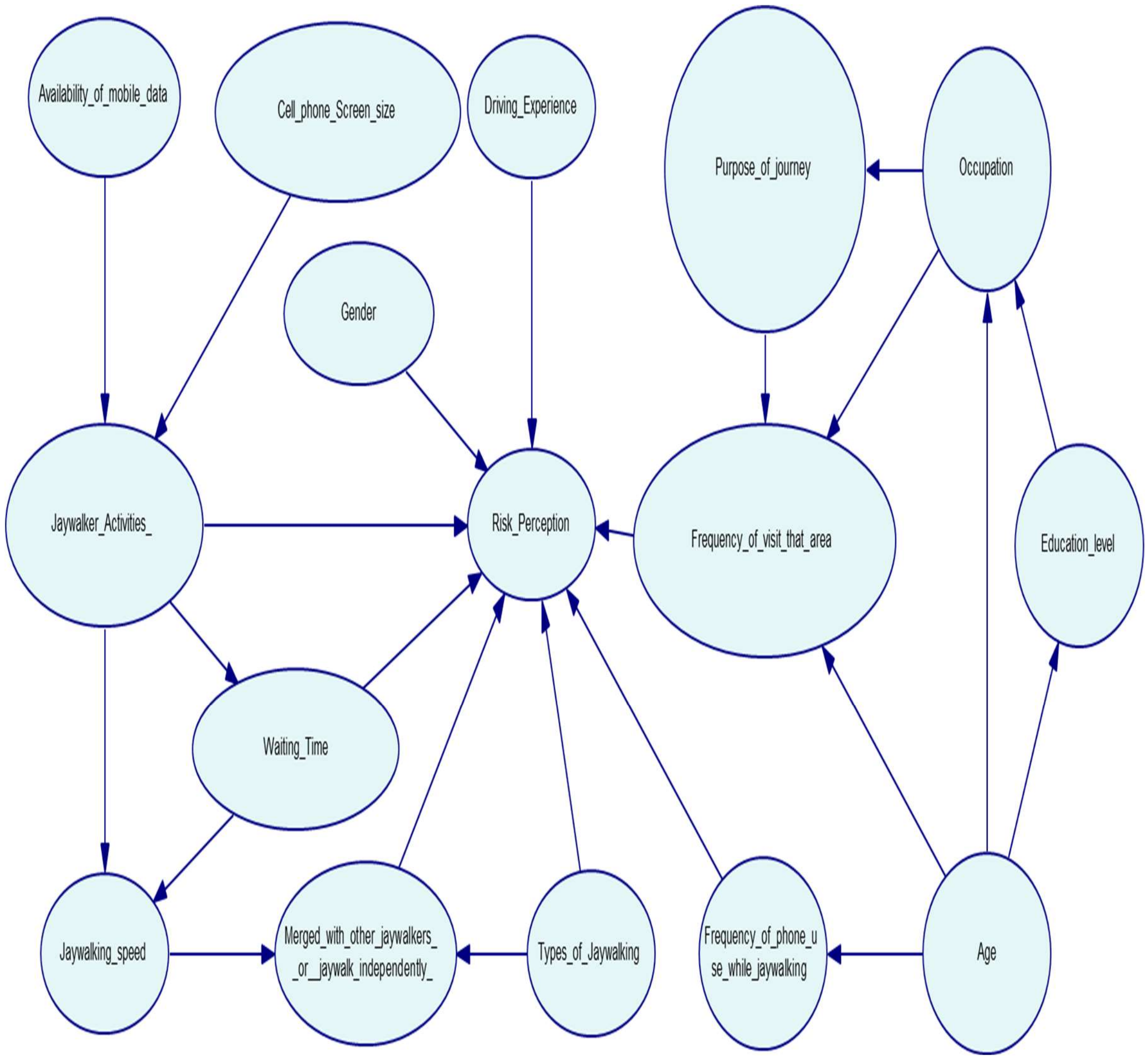

Figure 4.3: Developed Network.

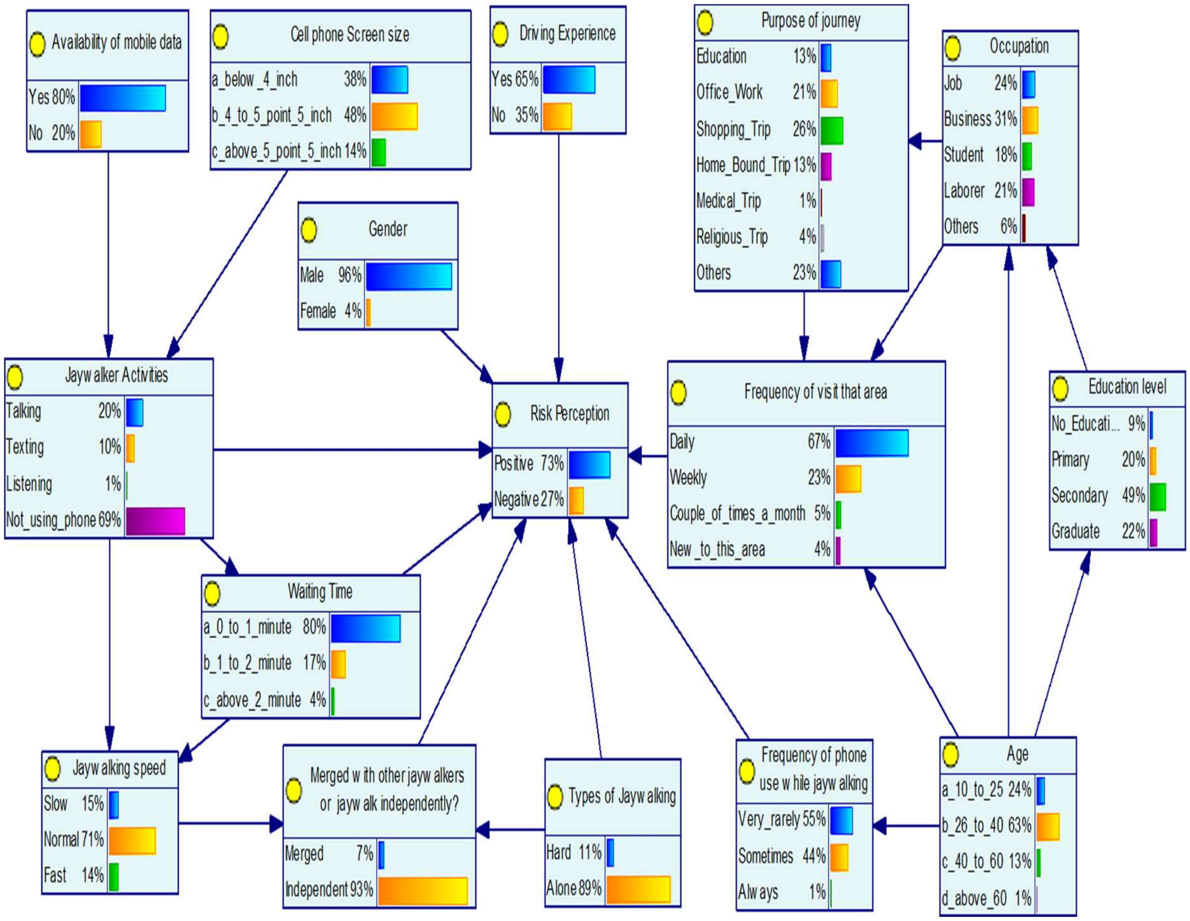

Figure 4.4: Prior probability distribution of the developed BBN.

4.3 Model Validation:

A crucial element of learning is validation of the results. We will show it on an example data set and a Bayesian network model learned from this data set. Model validation is a process to determine whether the model accurately reflects the functionality of the application. The risk perception is calculated by the model from the dataset consisting 1352 records of jaywalkers.

Accuracy:

Risk Perception = 0.911982 (1233/1352)

Positive = 0.964895 (962/997)

Negative = 0.76338 (271/355)

The validation result for a single class node is represented by estimating the accuracy. In this case, our model achieved 91.2% accuracy in predicting the risk perception that means it guessed correctly 1233 out of the total of 1352 records. Sensitivity of the model in detecting the Positive Risk Perception is roughly 96.5% (962 records out of all 997 records for which the Risk Perception was Positive), with specificity of roughly 76.3% (271 records out of all 355 records for which the Risk Perception was Negative).

ROC Curve: The *ROC Curve* tab shows the Receiver Operating Characteristic (ROC) curves for each of the states of each of the class variables. ROC curves originate from Information Theory and are an excellent way of expressing the quality of a model independent of the classification decision. The ROC curve is capable of showing the possible accuracy ranges, and the decision criterion applied by GeNIe is just one point on the curve. Choosing a different point will result in a different sensitivity and specificity (and, hence, the overall accuracy). The ROC curve gives insight into how much we have to sacrifice one in order to improve the other and, effectively, helps with choosing a criterion that is suitable for the application at hand. It shows the theoretical limits of accuracy of the model on one plot.

The Receiver Operating Characteristic (ROC) curve is also established for our model. It provides a graphic representation of the number of various cut points with their associated sensitivity vs. 1-specificity (i.e., false positives rate). This illustrates the merits of a specific predictor/predictive model, which makes it possible to identify different cut-points for specific applications – depending on the 'cost' of misclassification. The area under the curve (AUC) estimates provides an indication of the predictor's effectiveness and a means of comparing (testing) two or more predictive models.

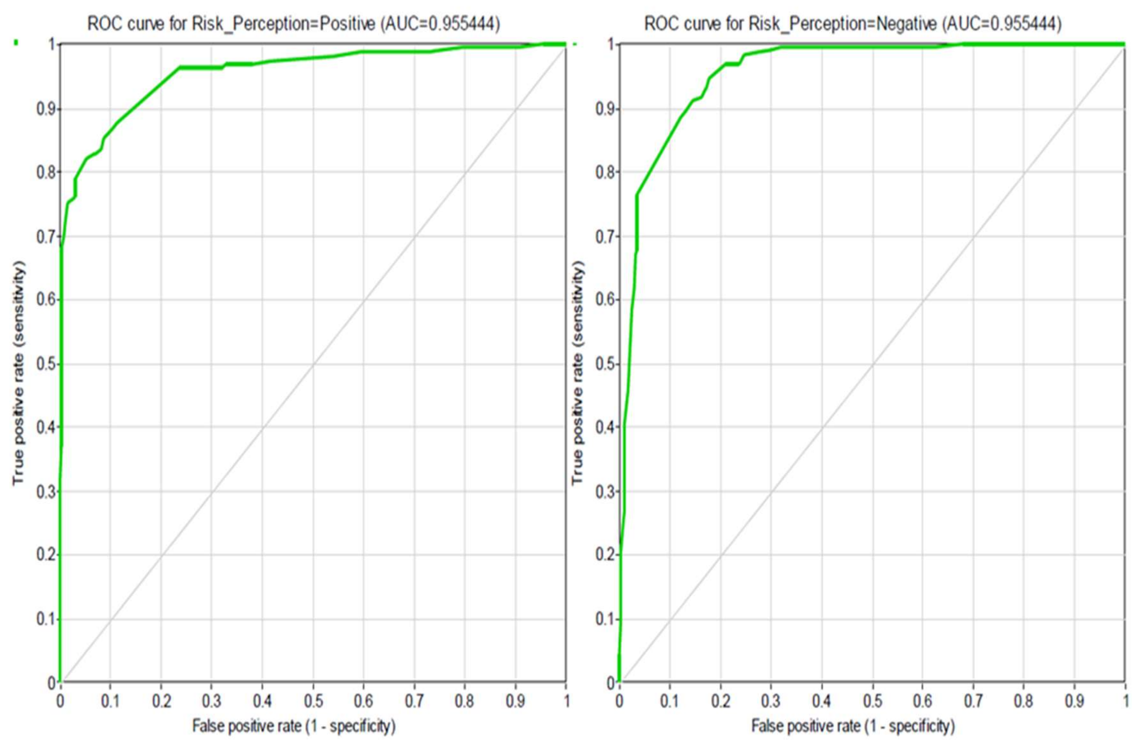

Figure 4.5: ROC Curve.

The AUC values for both positive and negative risk perception we obtained from the Receiver Operating Characteristic (ROC) curves are 0.96 that is quite close to 1.

<u>Calibration curve:</u> Another way of measuring the accuracy of a model is comparing the output probability to the actually observed frequencies in the data. The calibration curve shows how the two of them compare. One way of measuring the accuracy of a model is comparing the output probability to the actually observed frequencies in the data. The calibration curve shows how these two compares. For each probability p produced by the model (the horizontal axis), the plot shows the actual frequencies in the data (vertical axis) observed for all cases for which the model produced probability p. The dim diagonal line shows the ideal calibration curve, i.e., one in which every probability produced by the classifier is precisely equal to the frequency observed in the data. Because p is a continuous variable, the plot groups the values of probability so that sufficiently many data records are found to estimate the actual frequency in the data for the vertical axis.

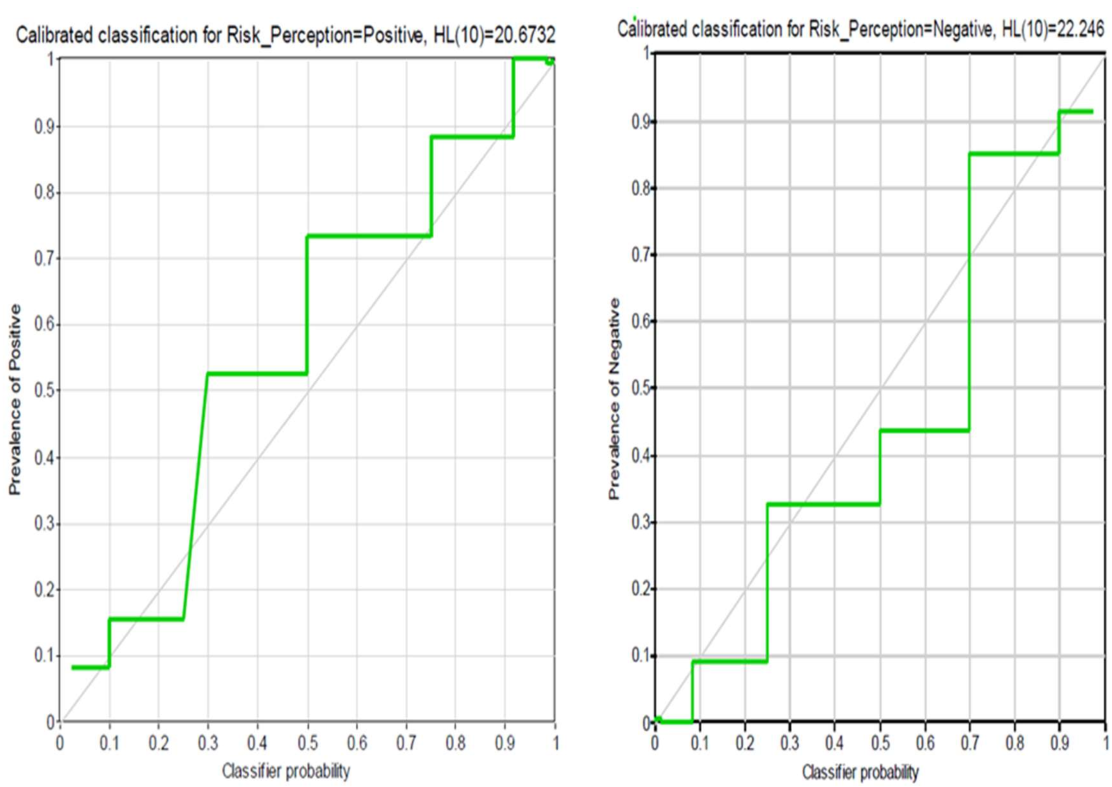

Figure 4.6: Calibration Curve.

4.4 Analysis and Result of Developed BBN:

Initially, 'Positive' state was set as the evidence in the 'Risk Perception' node to classify various attributes of pedestrians who have a positive perception of the risk of jaywalking. This is illustrated in Figure 4.7. Here, it shows that, pedestrians who are male, age group between 26-40 years, having secondary education and occupation as business, have the most positive risk perception regarding jaywalking. While 25% of pedestrians used their mobile phone while jaywalking, 18% of them talked. Pedestrians who jaywalked alone and went to that area daily could be more vulnerable when jaywalking. Further findings are that driving experience allows one to consider the possible danger of jaywalking more often than without driving experience.

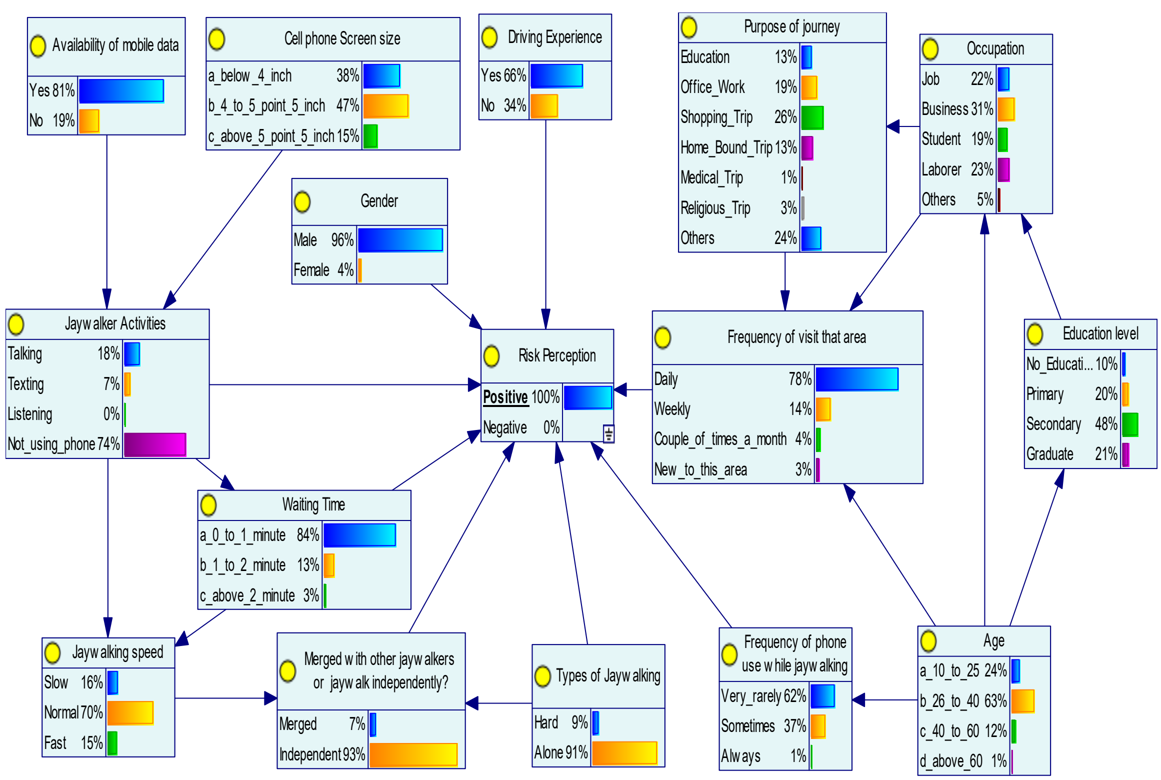

Figure 4.7: Posterior marginal probability distribution over all of the nodes when 'Positive' is set as the evidence on node 'Risk Perception'.

Again, 'Negative' state was set as the evidence in the 'Risk Perception' node to classify various attributes of pedestrians who have a negative perception of the risk of jaywalking. This is illustrated in Figure 4.8. Marginal probabilities of all nodes while evidences are set for 'Risk Perception' node, is displayed in Table 4.1.

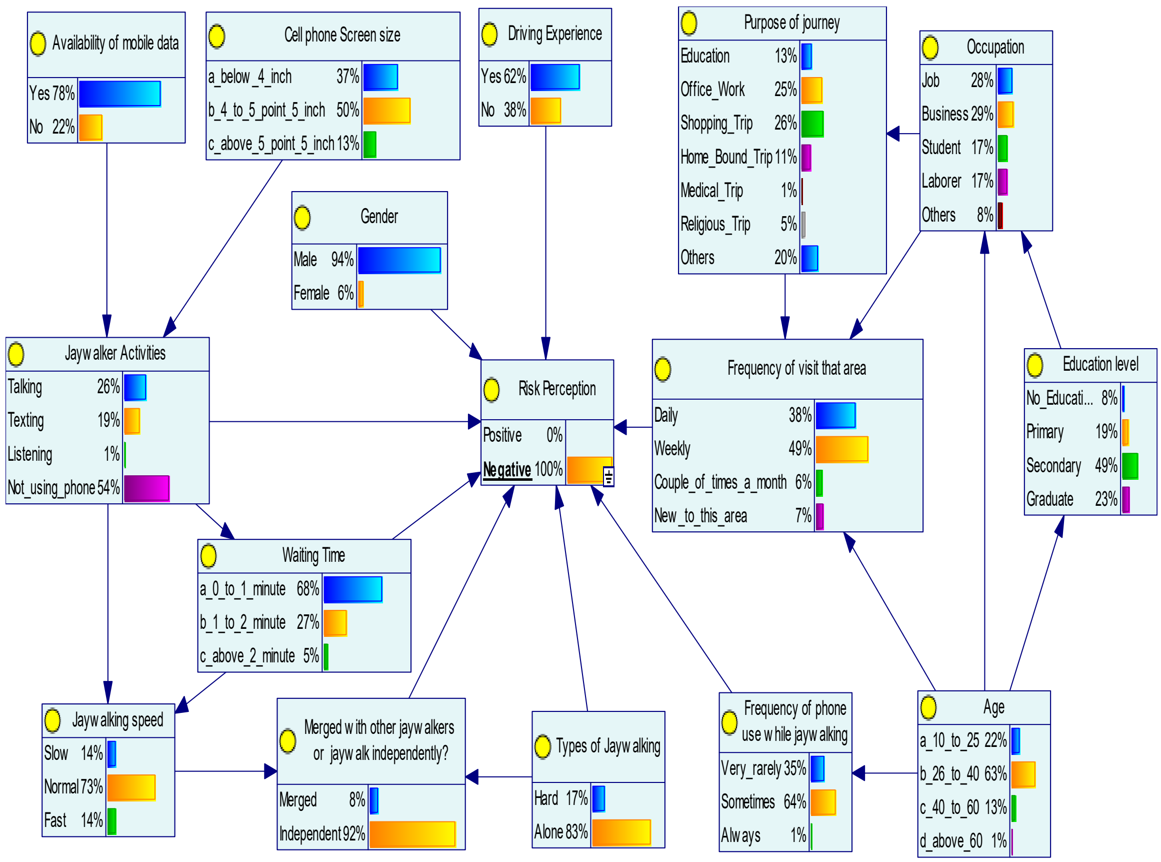

Figure 4.8: Posterior marginal probability distribution over all of the nodes when 'Negative' is set as the evidence on node 'Risk Perception'.

Table 4.1: Marginal probabilities of all nodes while evidences are set for "Risk Perception".

Attribute	**Attribute Category**	**Evidence**	
		Risk Perception	
		Positive (%)	**Negative (%)**
Age	a. 10-25 years	24	22
	b. 26-40 years	63	63
	c. 41-60 years	12	13
	d. Above 60 years	1	1
Occupation	a. Job	22	28
	b. Business	31	29
	c. Student	19	17
	d. Laborer	23	17
	e. Others	5	8
Gender	a. Male	96	94
	b. Female	4	6
Education Level	a. No education	10	8
	b. Primary	20	19
	c. Secondary	48	49
	d. Graduate	21	23
Purpose of journey	a. Education	13	13
	b. Office work	19	25
	c. Shopping Trip	26	26
	d. Home Bound Trip	13	11
	e. Medical Trip	1	1
	f. Religious Trip	3	5
	g. Others	24	20
Driving Experience	a. Yes	66	62
	b. No	34	38

Jaywalking speed	a. Slow	16	14
	b. Normal	70	73
	c. Fast	15	14
Merged with other jaywalkers or jaywalk independently?	a. Merged	7	8
	b. Independently	93	92
Types of Jaywalking	a. Herd	9	17
	b. Alone	91	83
Jaywalker Activities	a. Talking	18	26
	b. Texting	7	19
	c. Listening	0	1
	d. Not using phone	74	54
Waiting Time	a. 0-1 minute	84	68
	b. 1-2 minute	13	27
	c. more than 2 minute	3	5
Frequency of phone use while jaywalking	a. Very Rarely	62	35
	b. Sometimes	37	64
	c. Always	1	1
Cell phone Screen size	a. Below 4 inches	38	37
	b. 4 to 5.5 inch	47	50
	c. Larger than 5.5 inch	15	13
Frequency of visit that area	a. Daily	78	38
	b. Weekly	14	49
	c. Couple of times a month	4	6
	d. New to this area	3	7
Availability of mobile data	a. Yes	81	78
	b. No	19	22

For better understanding of the impacts of all variables in the network, ‘Risk Perception’ is specified as target variable in sensitivity analysis. Gender, waiting time, types of jaywalking, frequency of phone use while jaywalking and jaywalker activities are found to be most impactful variables for risk perception. They are illustrated in deep red color in Figure 4.9. Frequency of visit that area is also significant. Socio-demographic characteristics have a likely influence on the target variable. While, jaywalking speed is found to be most insignificant variable for risk perception.

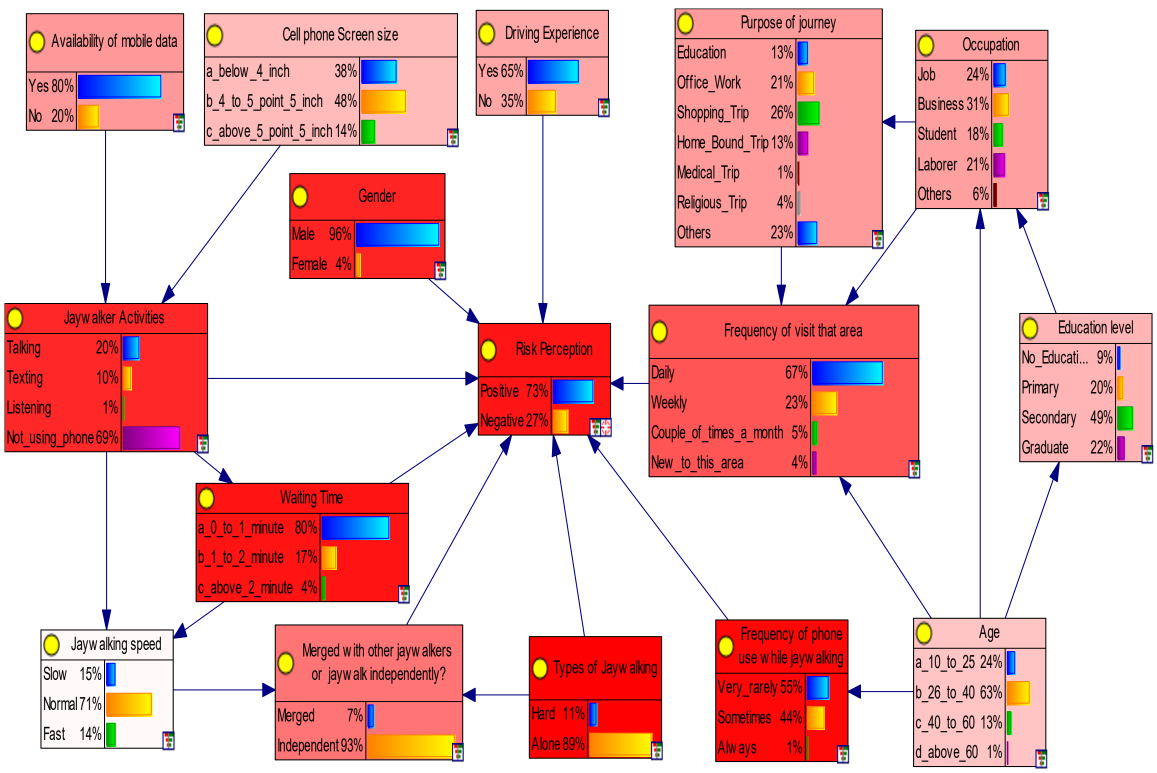

Figure 4.9: The BBN and significant variables for jaywalking.

Based on the evidence left behind it, the approximate BN can be used to assess various risk perceptions of jaywalkers. To identify different attributes of pedestrians during jaywalking, evidence was set on various nodes to find out their awareness level by accessing “Risk Perception” node in Figure 4.4.

Setting evidence for age group of pedestrians in between 10-25 years and having driving experience, from the BN it is seen that, probability of having positive risk perception is 76%. Which is the highest. If the age group is changed to between 26-40 years, probability of positive risk perception decreases to 74%. This declination can also be seen for remaining age groups. When the age groups have no driving experience, the probability of positive risk perception is 72% and 71% respectively which less than the previous case. It concludes that, driving experience can help to increase awareness of pedestrians of any age group.

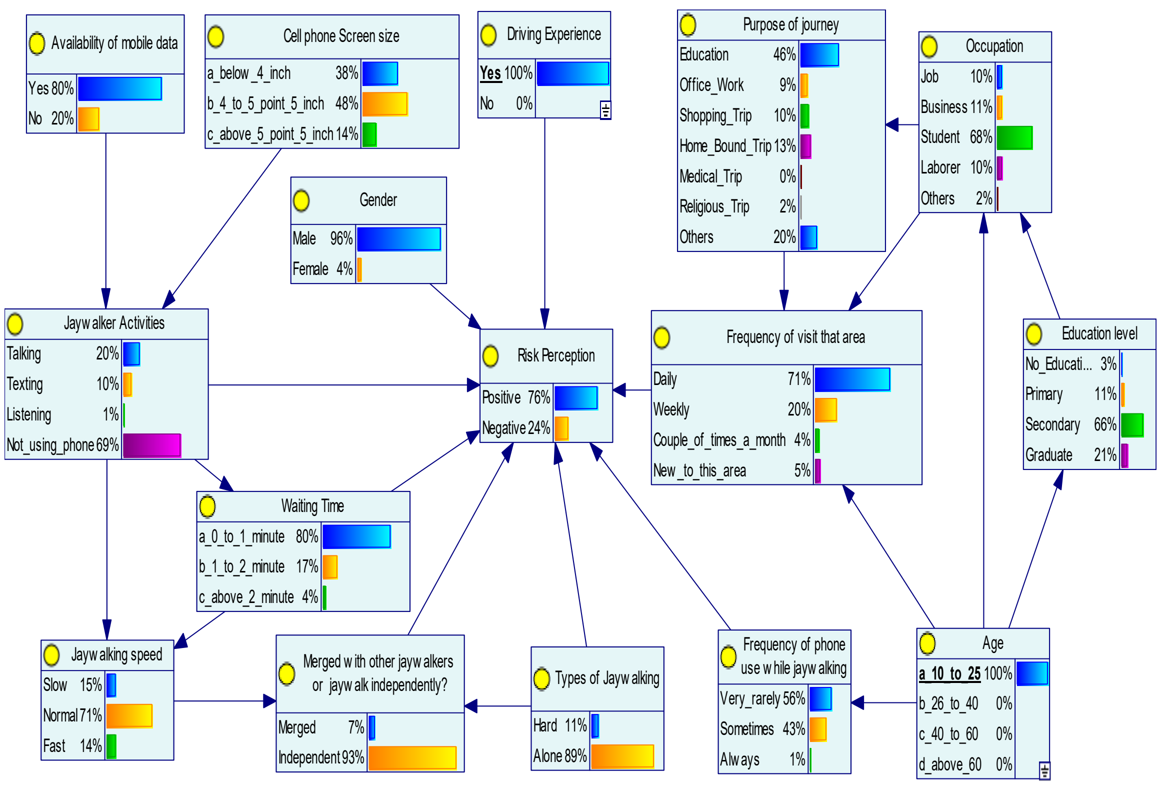

Figure 4.10: State of Bayesian Network for Risk Perception when pedestrian have driving experience and age between 10-25 years.

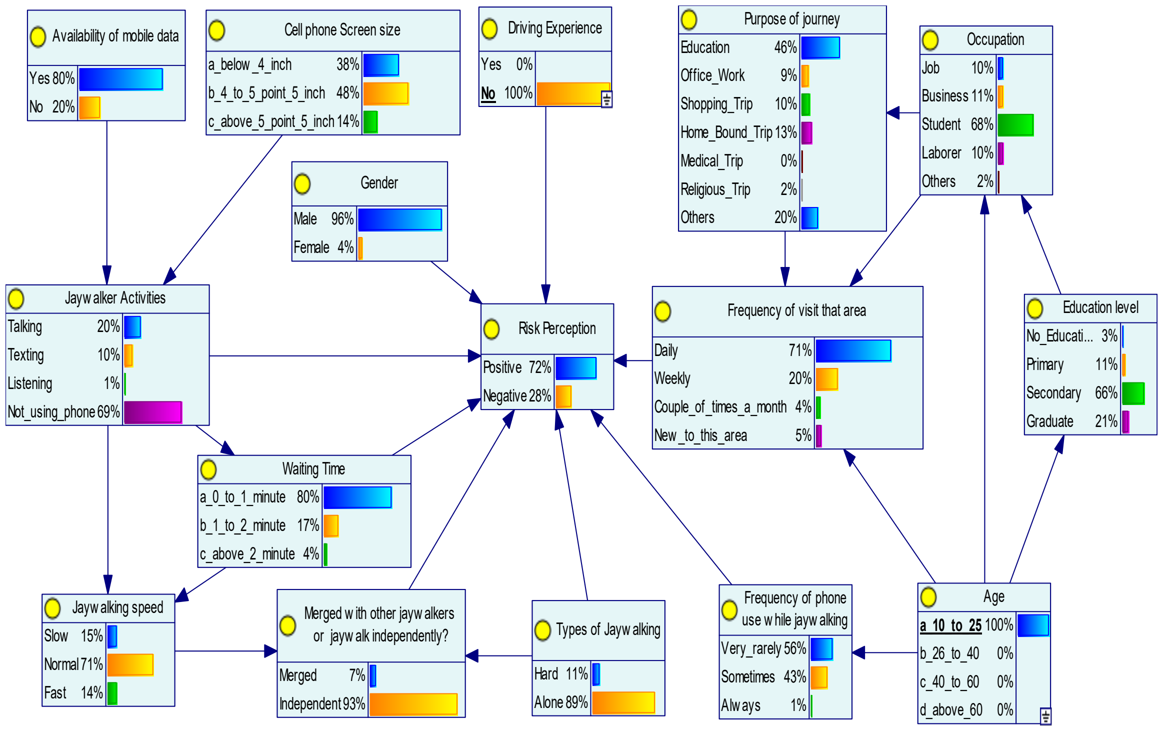

Figure 4.11: State of Bayesian Network for Risk Perception when pedestrian have no driving experience and age between 10-25 years.

Jaywalker activity is one of the impactful variables for accessing risk perception. Incorporating evidence of age group with jaywalker activity, it is found that, jaywalker aged between 10-25 years and not engaged in any mobile phone induced jaywalker activity, have the highest probability of positive risk perception of 80%. Age group 26-40 years holds the second place with 79% probability. But when jaywalker engaged in any mobile phone induced jaywalker activity, the probability of positive risk perception decreases drastically for any age group. Here, it can be seen that, texting while jaywalking have the lowest probability for positive risk perception of 50%. This trend is also seen in other jaywalker activities (talking-68% and listening through headphones-56%). So, mobile phone induced jaywalker activity for any age group can decrease pedestrians' risk perception.

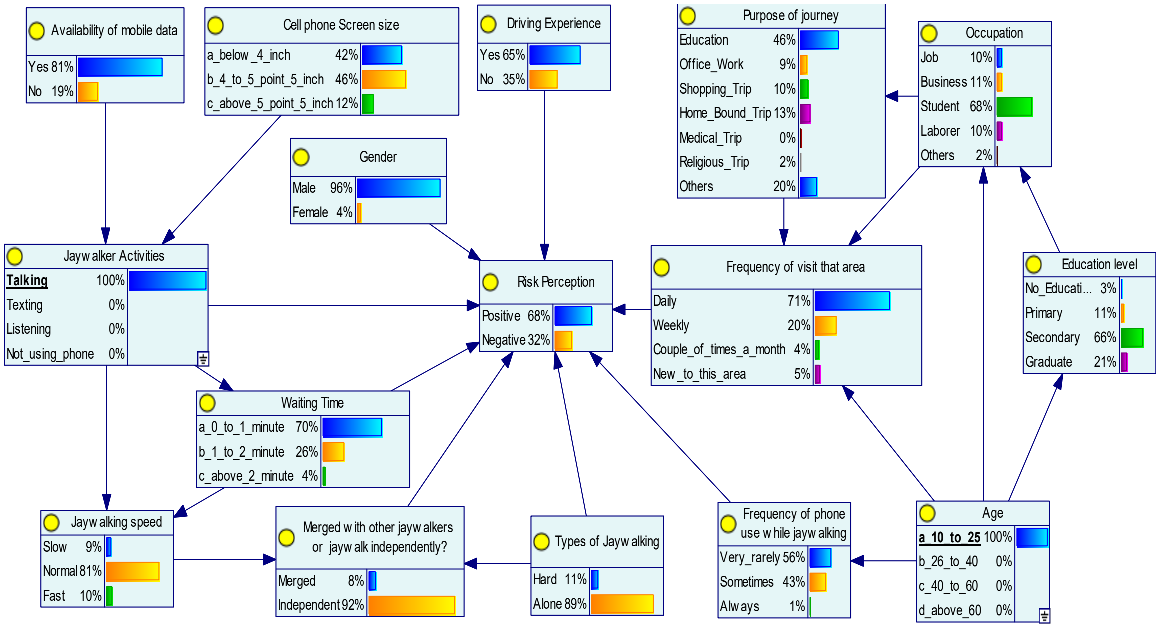

Figure 4.12: State of Bayesian Network for Risk Perception when pedestrian talk via phone while jaywalking and age between 10-25 years.

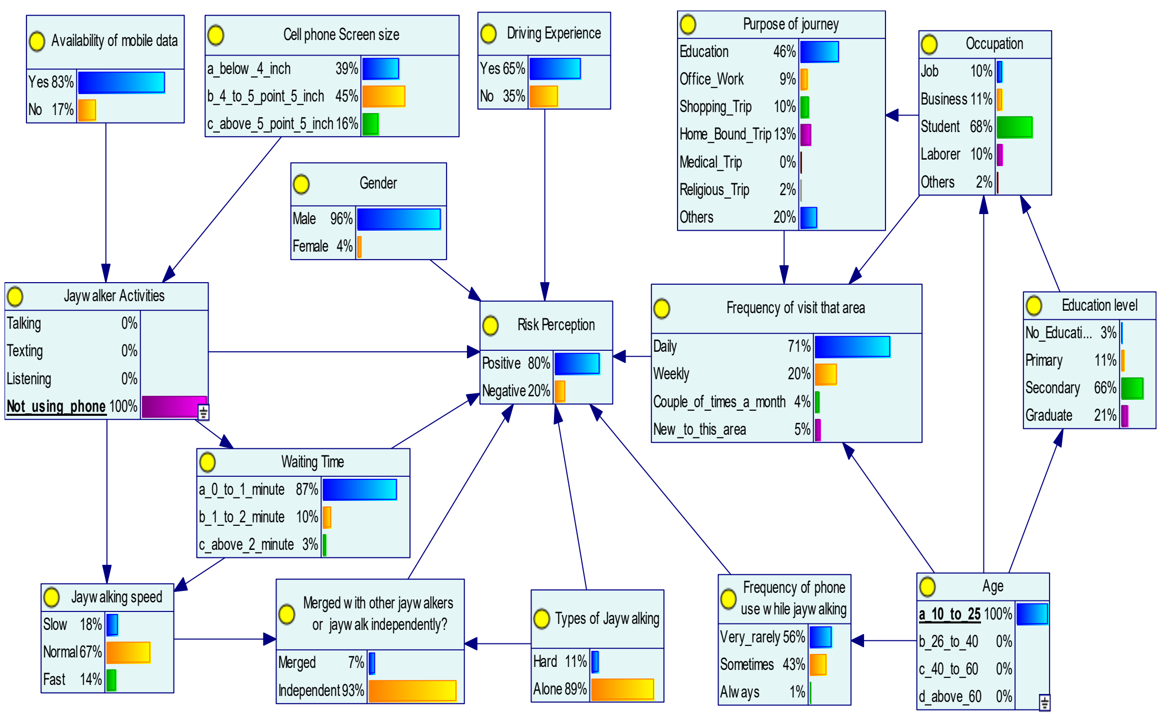

Figure 4.13: State of Bayesian Network for Risk Perception when pedestrian not using phone while jaywalking and age between 10-25 years.

Differences in risk perception can be seen with different age groups and gender. 10-25 years male pedestrian have 75% probability of positive risk perception. On the other hand, in the same age group, the female pedestrian has only 64% probability of positive risk perception. Declination of risk perception can be seen with increasing age for both male (75%-72%) and female (64%-61%) pedestrians. It concludes that female pedestrians have less awareness regarding the risk of jaywalking.

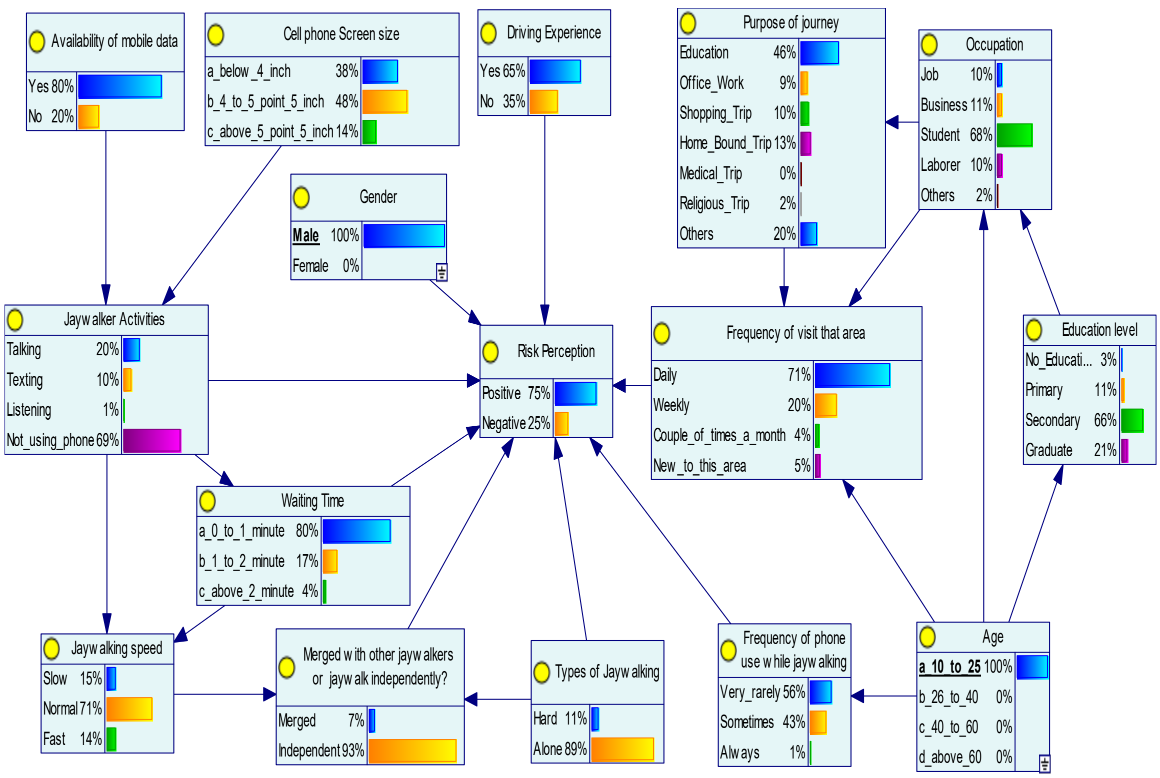

Figure 4.14: State of Bayesian Network for Risk Perception when pedestrian is male and age between 10-25 years.

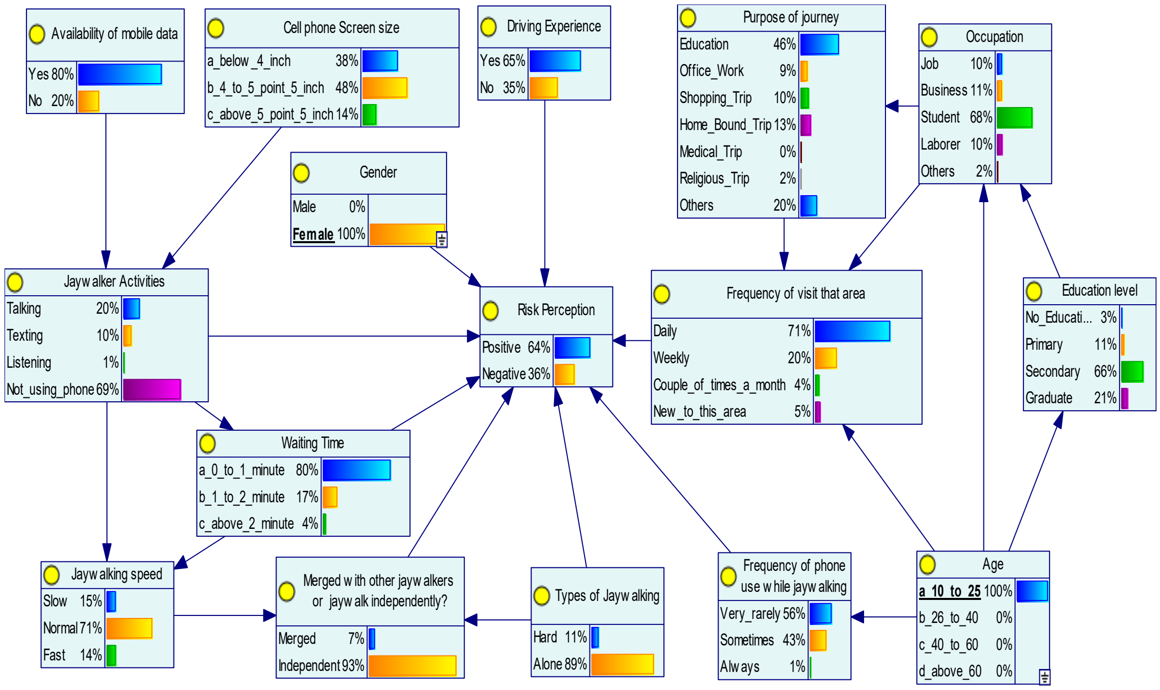

Figure 4.15: State of Bayesian Network for Risk Perception when pedestrian is female and age between 10-25 years.

From sensitivity analysis, it is seen that, types of jaywalking can be a deciding variable for risk perception. When incorporating it with occupation, laborer pedestrian who were walking alone found to be most aware of risk perception (80%). But, when they were walking in herd, they found to be less aware (61%). Same trend can be seen in pedestrian of other occupation. Job holder, businessmen, student all seen to be less conscious of risk regarding jaywalking when they walk in herd. Therefore, it is safe to assume that if they jaywalk alone, pedestrians of every occupation remain fairly alert.

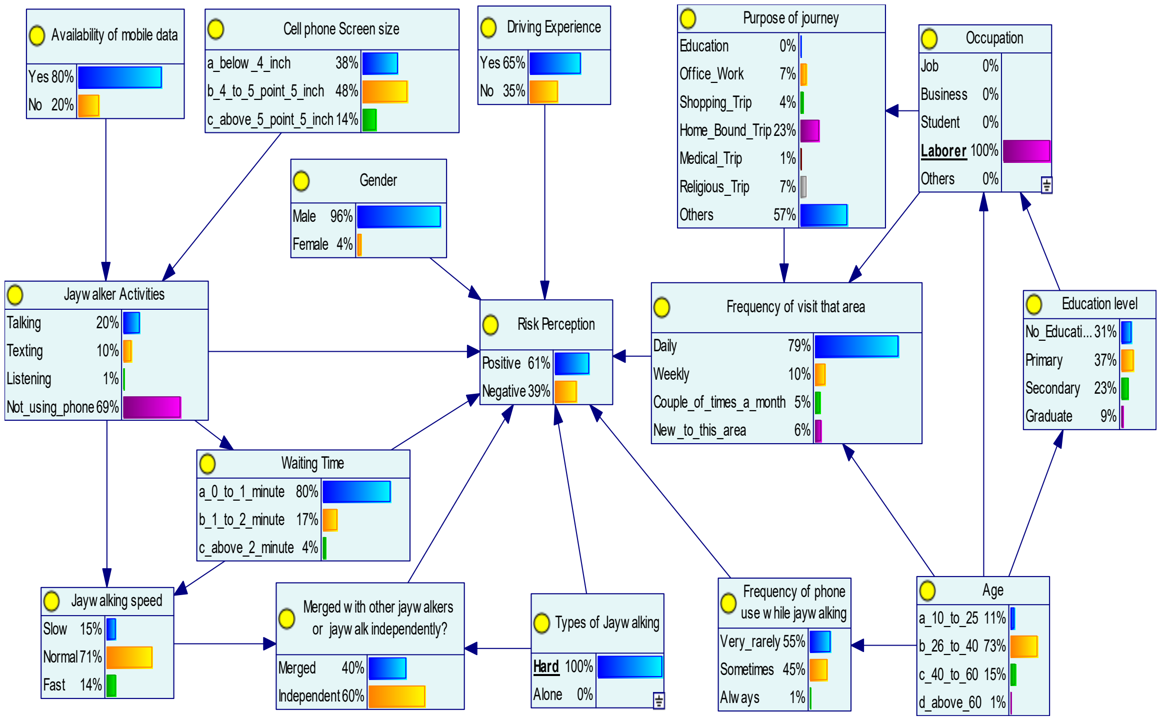

Figure 4.16: State of Bayesian Network for Risk Perception when pedestrian is a laborer and types of jaywalking is herd.

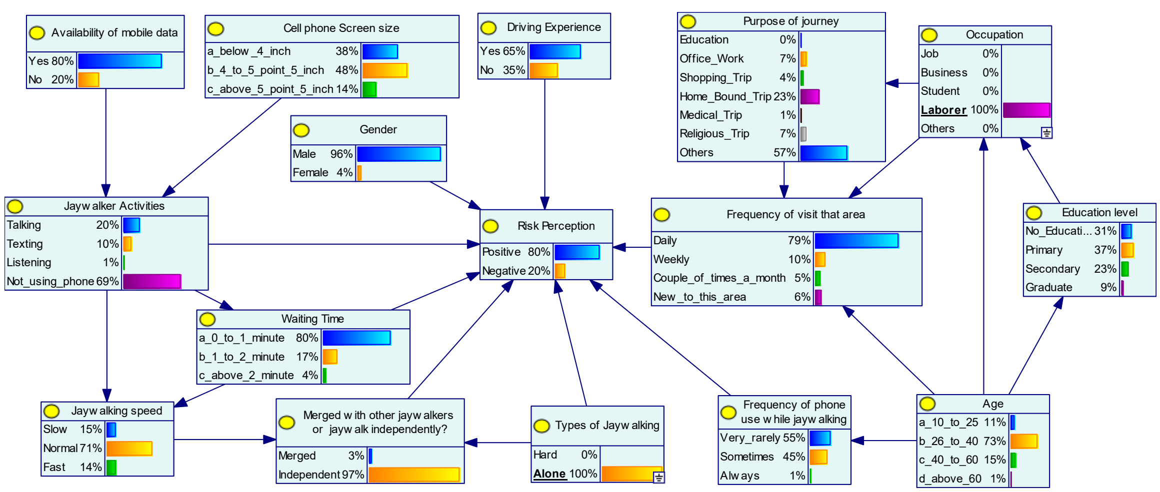

Figure 4.17: State of Bayesian Network for Risk Perception when pedestrian is a laborer and types of jaywalking is alone.

People have to frequently visit some area for different purposes. From the BN, it indicates that, whatever the purpose of the journey is, pedestrian remain very alert when they daily visit that area (85%). Interestingly, their risk perception remains low when they visit the area weekly, couple of times or newly arrive (43%-56%). It concludes that, with visiting any area daily, pedestrian seems to understand the risk of jaywalking in that particular area and tries to remain aware every time they visit that area.

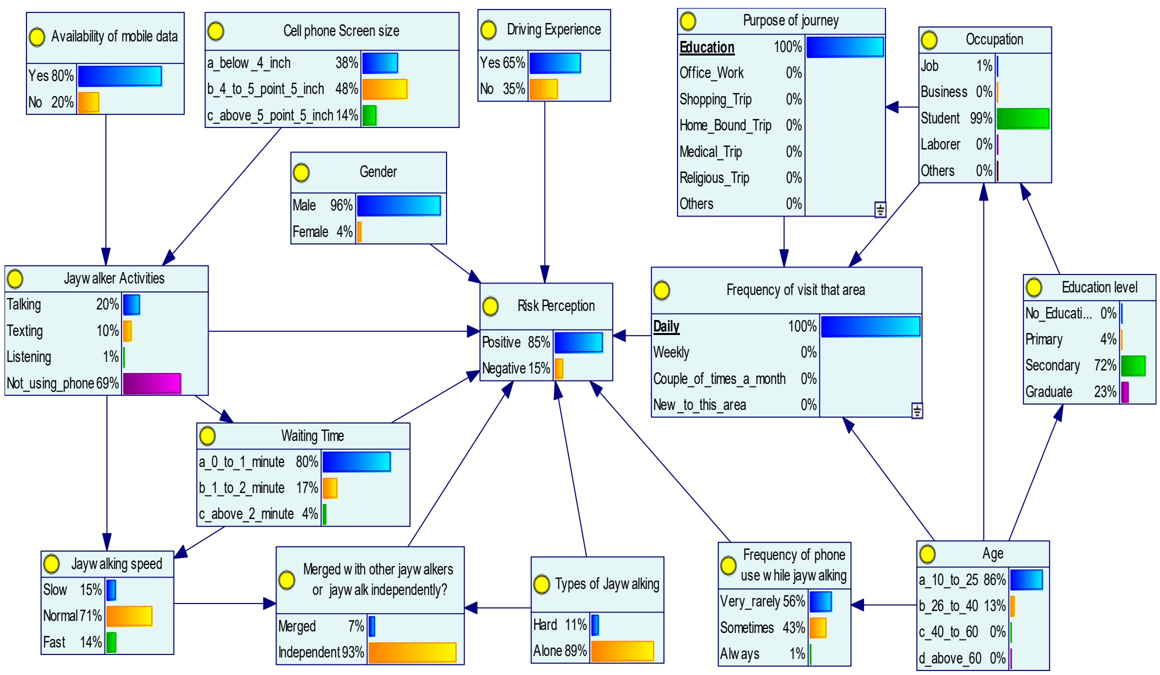

Figure 4.18: State of Bayesian Network for Risk Perception when pedestrian daily visit to educational institution.

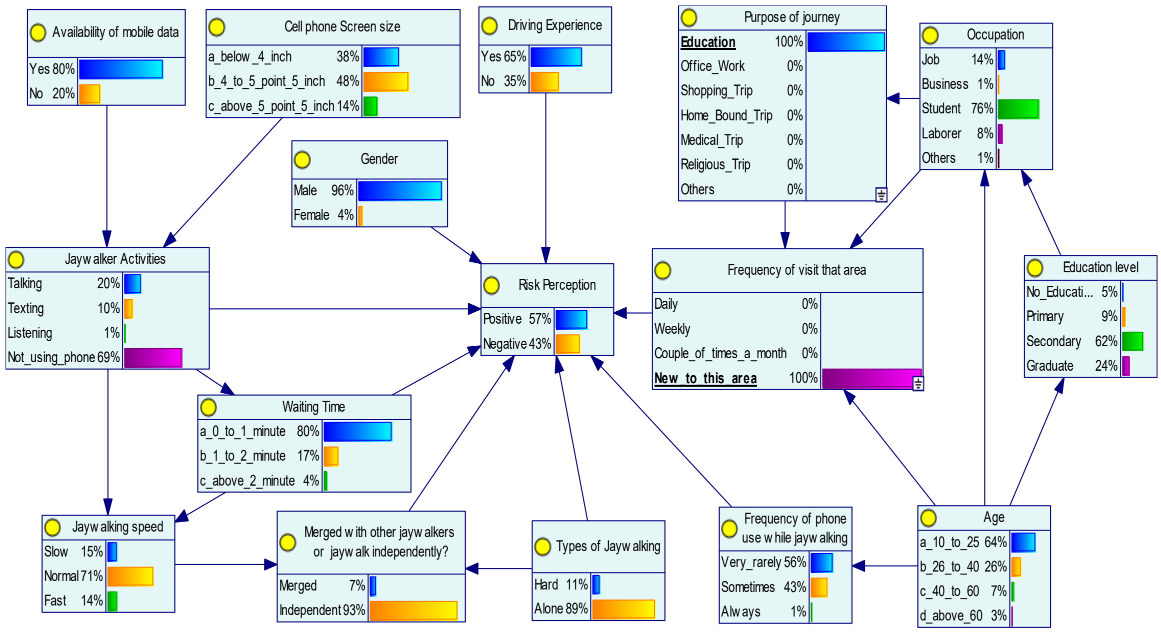

Figure 4.19: State of Bayesian Network for Risk Perception when pedestrian visiting educational institution for the first time.

As gender has been incorporated with types of jaywalking, male pedestrians remain conscious of the fact that they walk alone (75%) than jaywalk in the herd (59%). For female pedestrians, the same pattern is seen. However, in both situations, their awareness level is surprisingly low than male pedestrians.

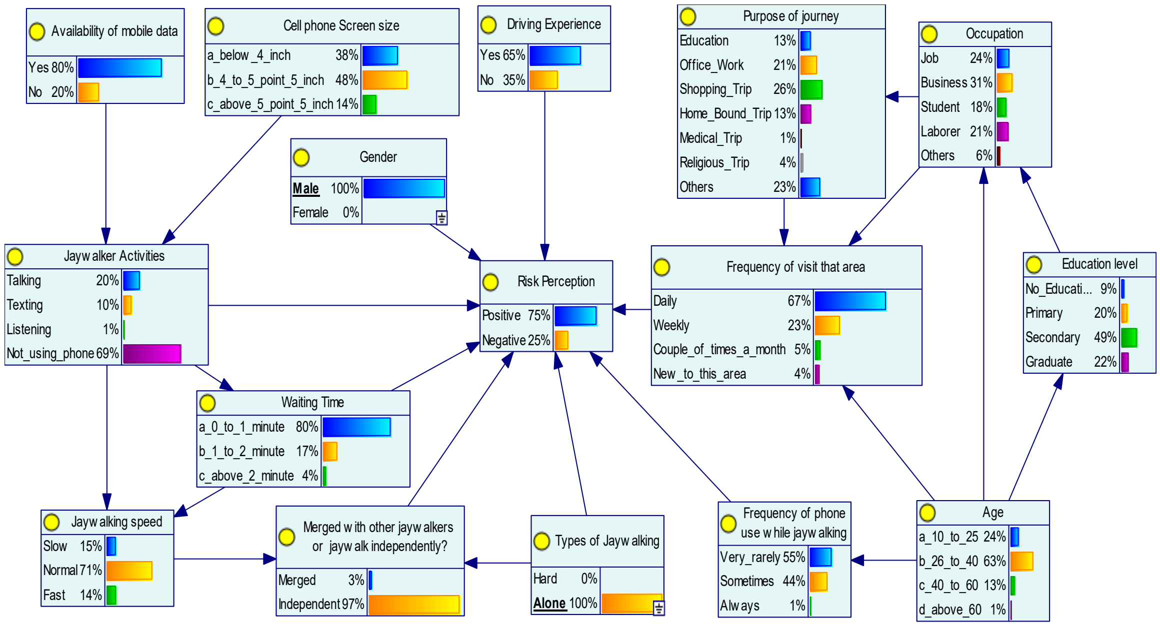

Figure 4.20: State of Bayesian Network for Risk Perception when male pedestrians jaywalk alone.

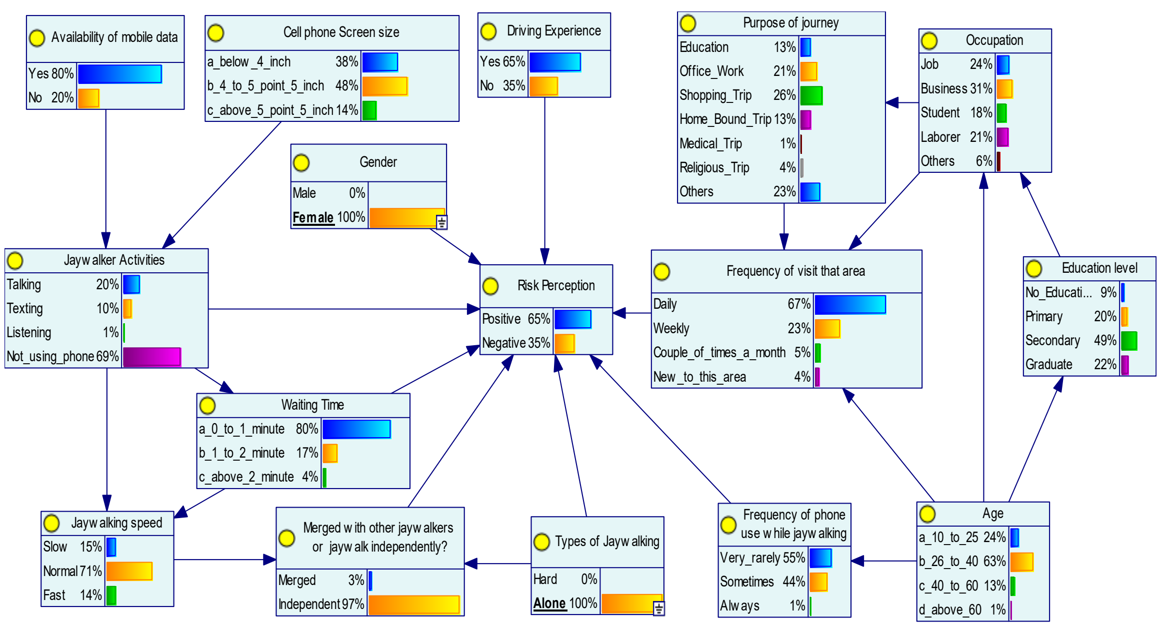

Figure 4.21: State of Bayesian Network for Risk Perception when female pedestrians jaywalk alone.

For different waiting times, a disparity in risk perception can be observed. If the waiting time is low (0-1 minutes), the level of consciousness for both males and females is high (78% and 67%). But when the time of waiting is high (over two minutes), their patience continues to fall (64% and 49%). Another interesting discovery is that, whatever the waiting time is, awareness is lacking for female pedestrians than male pedestrians.

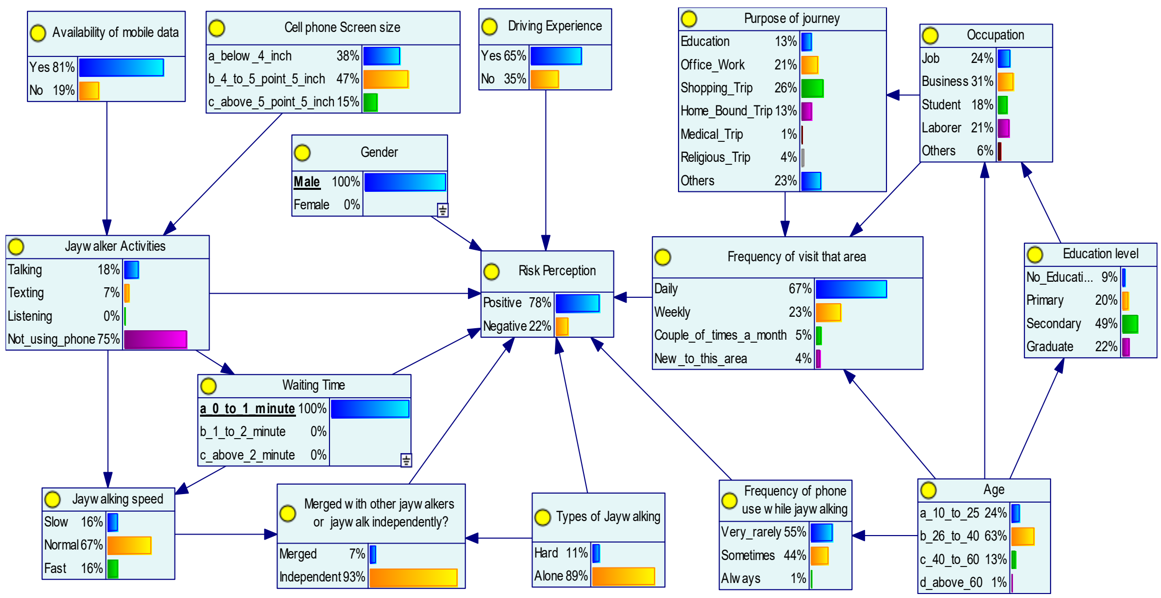

Figure 4.22: State of Bayesian Network for Risk Perception when waiting time is low for male pedestrians.

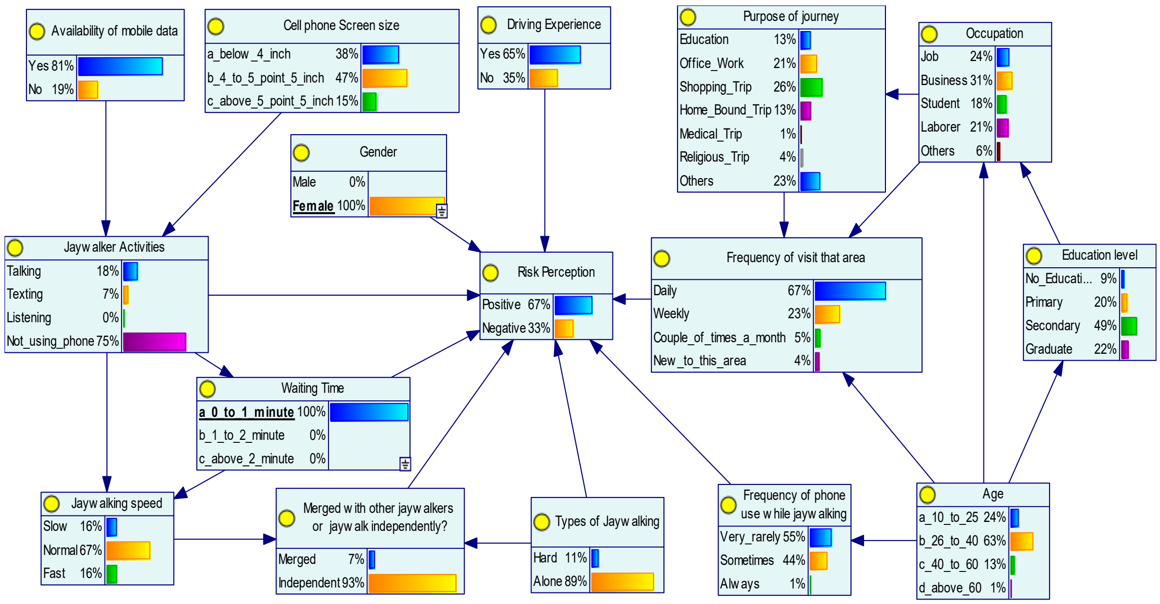

Figure 4.23: State of Bayesian Network for Risk Perception when waiting time is low for female pedestrians.

Jaywalking types can affect phone use while jaywalking. Pedestrians seems to remain alert when they jaywalk alone and as a result, they use their mobile phone very (85%). On the other hand, they seem quite careless when they jaywalk in herd and uses the mobile phone always (52%). So, jaywalk in a group can boost confidence of pedestrians and sometimes they become less conscious about the risks regarding jaywalking.

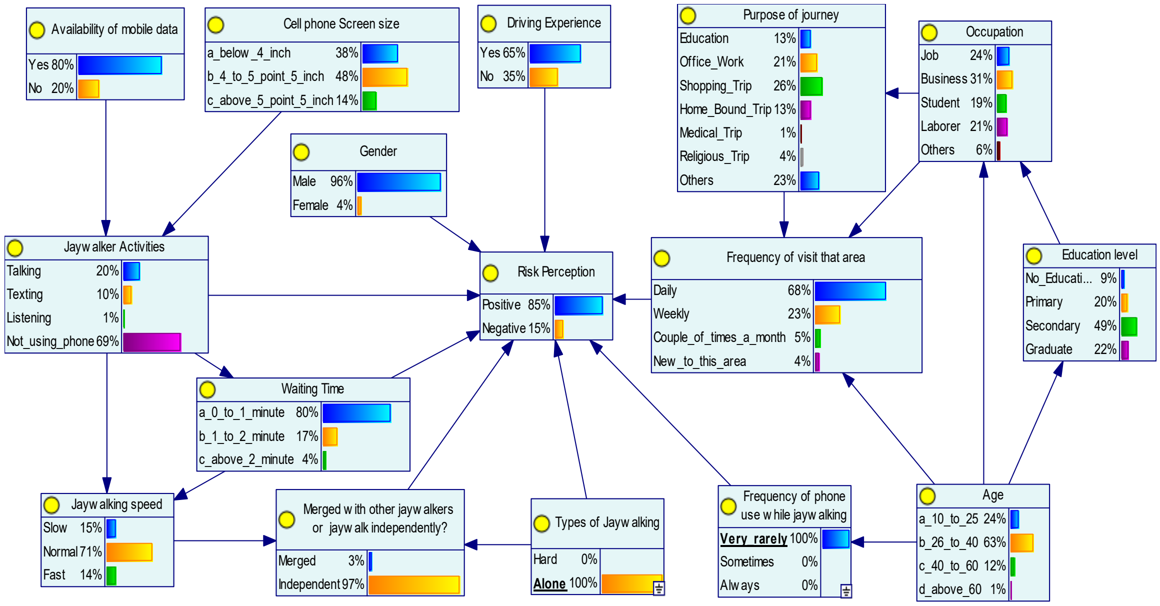

Figure 4.24: State of Bayesian Network for Risk Perception when pedestrians jaywalk alone and uses phone very rarely.

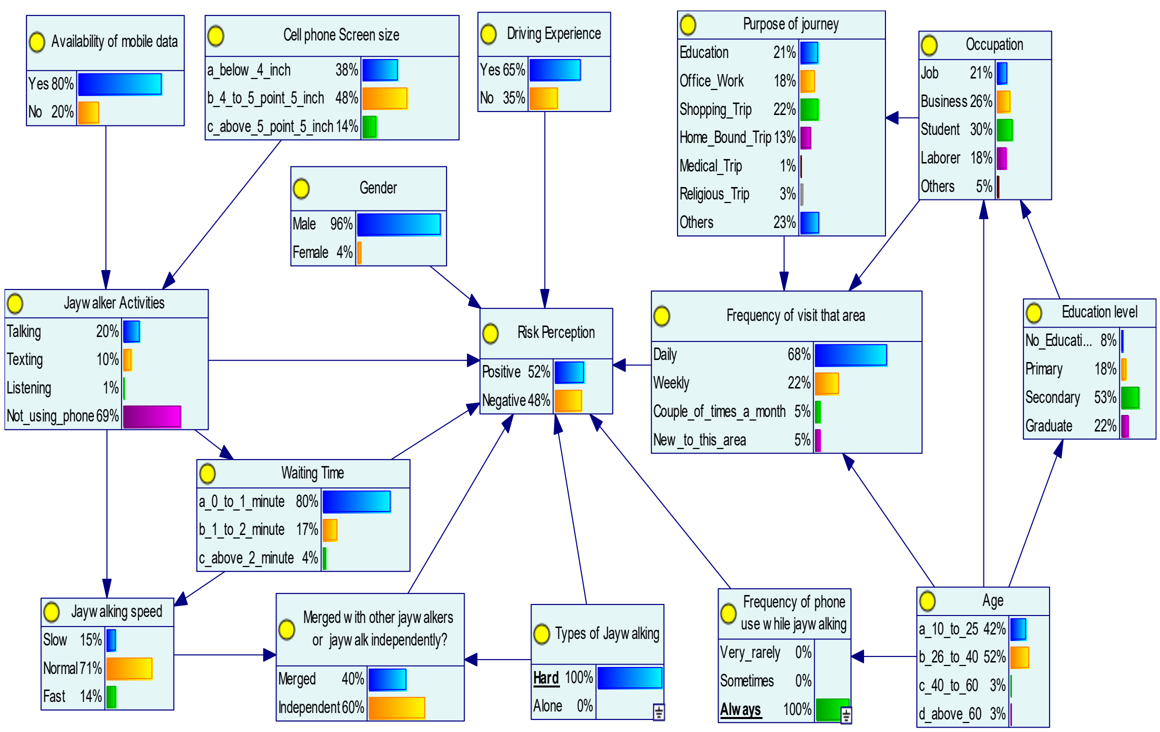

Figure 4.25: State of Bayesian Network for Risk Perception when pedestrians jaywalk in herd and uses phone always.

Marginal probabilities of 'Risk Perception' while evidences are set for other nodes, is displayed in Table 4.2.

Table 4.2: Marginal Probability of 'Risk Perception' by setting evidences on other nodes.

Evidence			**Risk Perception**	
Attribute	**Attribute Category**	**Attribute Category**	**Positive**	**Negative**
Driving Experience-Age	Yes	a) 10-25	76	24
		b) 26-40	74	26
		c) 40-60	73	27
		d) Above 60	73	27
	No	a) 10-25	72	28
		b) 26-40	71	29
		c) 40-60	69	31
		d) Above 60	68	32
Jaywalker Activities-Age	Talking	a) 10-25	68	32
		b) 26-40	66	34
		c) 40-60	64	36
		d) Above 60	63	37
	Texting	a) 10-25	50	50
		b) 26-40	49	51
		c) 40-60	49	51
		d) Above 60	52	48
	Listening	a) 10-25	56	44
		b) 26-40	56	44
		c) 40-60	56	44
		d) Above 60	54	46
	Not Using	a) 10-25	80	20
		b) 26-40	79	21
		c) 40-60	77	23
		d) Above 60	77	23
Gender-Age	Male	a) 10-25	75	25
		b) 26-40	73	27
		c) 40-60	72	28
		d) Above 60	72	28
	Female	a) 10-25	64	36
		b) 26-40	63	37
		c) 40-60	62	38
		d) Above 60	61	39

Types of Jaywalking-Occupation	Herd	a) Job	56	44
		b) Business	59	41
		c) Student	59	41
		d) Laborer	61	39
		e) Others	52	48
	Alone	a) Job	70	30
		b) Business	76	24
		c) Student	77	23
		d) Laborer	80	20
		e) Others	63	37
Purpose of journey-Frequency of visit that area	Education	a) Daily	85	15
		b) Weekly	44	56
		c) Couple of times	66	34
		d) New to area	57	43
	Office work	a) Daily	85	15
		b) Weekly	43	57
		c) Couple of times	66	34
		d) New to area	56	44
	Shopping trip	a) Daily	85	15
		b) Weekly	43	57
		c) Couple of times	66	34
		d) New to area	56	44
	Home bound trip	a) Daily	85	15
		b) Weekly	43	57
		c) Couple of times	66	34
		d) New to area	56	44
	Medical trip	a) Daily	85	15
		b) Weekly	43	57
		c) Couple of times	66	34
		d) New to area	56	44
	Religious trip	a) Daily	85	15
		b) Weekly	43	57
		c) Couple of times	66	34
		d) New to area	56	44
	Others	a) Daily	85	15
		b) Weekly	43	57
		c) Couple of times	66	34

		d) New to area	56	44
Gender-Types of Jaywalking	Male	a) Herd	59	41
		b) Alone	75	25
	Female	a) Herd	51	49
		b) Alone	65	35
Gender-Waiting time	Male	a) 0-1 minute	78	22
		b) 1-2 minutes	56	44
		c) above 2 minutes	64	36
	Female	a) 0-1 minute	67	33
		b) 1-2 minutes	50	50
		c) above 2 minutes	49	51
Types of Jaywalking-Frequency of phone use while jaywalking	Herd	a) very rarely	63	37
		b) sometimes	53	47
		c) always	52	48
	Alone	a) very rarely	85	15
		b) sometimes	62	38
		c) always	57	43

Chapter 5
CONCLUSION

5.1 Introduction

This research aims to identify the most probable causes associated with the understanding of risks in developing countries like Bangladesh of jaywalking and mobile phone induced distractions at highway intersections. BBN and Sensitivity Analysis, an effective measure of artificial intelligence (AI), is used to fulfil this aim by prediction modelling of entirely unpredictable events. In this study, a video graphic survey was used to document pedestrian jaywalking behavior, a questionnaire survey to gather the socio-demographic characteristics of pedestrians' and safety perception, and field observations were used to explore potential explanations for pedestrians' jaywalking and overall conditions of the walkway. The findings of this study are comprehensive observations made from 2016 pedestrians involved in jaywalking and their different perceptions of risk. The most significant variables affecting risk perception are gender, waiting time, types of jaywalking, frequency of phone use while jaywalking and jaywalker activities. In contrast, the most irrelevant risk perception variable is jaywalking speed. In order to classify various attributes, evidence was placed on different nodes to discover pedestrians' level of awareness through "Risk Perception". Result of this research indicates that pedestrians of all age groups can benefit from driving experience to increase awareness. A further finding is that when pedestrians use mobile phones, awareness about danger is lowest. Shockingly, female pedestrians are found to be less aware than male pedestrian and awareness remains high as pedestrian jaywalk alone. And finally, with a regular tour to any area, a pedestrian seems to be aware of the danger of jaywalking in that particular area and always tries to be aware of it. Therefore, it is expected that addressing the established significant variables associated with risky jaywalking, as well as the triggering factors of jaywalking triggered by mobile phones, would result in enhanced overall protection of pedestrians.

5.2 Policy Implementation:

The results obtained from this study possess strong policy implications, since it will improve the overall pedestrian safety. The findings of this study supports similar research findings on pedestrians'' demographic characteristics who are mostly involved in traffic law violation Ma et al. (2020), Marisamynathan and Vedagiri (2018), Gong et al. (2019, Ren et al. (2011). Another significant finding of the research is that having 'driving experience' increases pedestrians' attitudes towards the danger of jaywalking, which is similar to previous research findings Dommes et al. (2015). One who has driving experience knows the laws and risks associated to driving and jaywalking, which enables him/her to refrain from such actions. Hence, driving education could play an integral role in improving overall safety. Congested and unsafe passenger walkway which is reported to be one of the most key factors affecting the risks of jaywalking. Governmental agencies, therefore, need to successfully implement interventions to ensure proper and safe walkways. In addition, on the basis of this current study, mobile phone induced distracting activities potentially undermines pedestrian safety which supports previous findings Jiang et al. (2018), Nasar et al. (2008), Sobhani and Farooq (2018). Therefore, legislators and government agencies must implement effective countermeasures against cell phone usage while walking. There must also be strict enforcement of laws, fines, and sanctions to achieve maximum cognitive power of pedestrians'.

5.3 Limitations and Future Scope:

This subsection concludes the thesis by stating the limitations of the study. Also, further advancement that can be ventured from this study is presented as future scope in this subsection.

- The study is conducted only in 32 national and regional highway intersections of Bangladesh. For further study, the number of intersections can be increased in order to get better inspection of this study.
- In this study, data is collected only from highway intersections of Bangladesh. In future data can be collected from mid-blocks of highways.
- In this study, we have mainly emphasized on the awareness level of pedestrians associated with jaywalking. In future, pedestrian-vehicle collision can also be incorporated.

REFERENCES

Alsaleh, R., Sayed, T., & Zaki, M. H. (2018). Assessing the Effect of Pedestrians' Use of Cell Phones on Their Walking Behavior: A Study Based on Automated Video Analysis. Transportation Research Record, 1–12.

Brosseau, M., Zangenehpour , S., Saunier, N., & Miranda-Moreno, L. (2013). The impact of waiting time and other factors on dangerous pedestrian crossings and violations at signalized intersections: A case study in Montreal. Transportation Research Part F, 159–172.

Chen, P.-L., Saleh, W., & Pai, C.-W. (2018). Pokemon gaming causes pedestrians to run a red light: An observational study of crossing behaviours at a signalised intersection in Taipei City. Transportation Research Part F, 380–388.

Dommes, A., Granié , M.-A., Cloutier, M.-S., Coquelet, C., & Huguenin-Richard, F. (2015). Red light violations by adult pedestrians and other safety-related behaviors at signalized crosswalks. Accident Analysis and Prevention, 67–75.

Gong, Q., Liang, X., & Xu, M. (2019). Pedestrian violations crossing behavior at signal intersections: A case study in Anning District of Lanzhou. In IOP Conference Series: Materials Science and Engineering. Vol. 688, No. 4, p. 044006). IOP Publishing.

Hatfield, J., & Murphy, S. (2007). The effects of mobile phone use on pedestrian crossing behaviour at signalised and unsignalised intersections. Accident Analysis and Prevention, 197–205.

Jiang, K., Ling, F., Feng, Z., Ma, C., Kumfer, W., Shao, C., & Wang, K. (2018). Effects of mobile phone distraction on pedestrians' crossing behavior and visual attention allocation at a signalized intersection: An outdoor experimental study. Accident Analysis and Prevention, 170–177.

King, M. J., Soole, D., & Ghafourian, A. (2009). Illegal pedestrian crossing at signalised intersections: Incidence and relative risk. Accident Analysis & Prevention, 485-490.

Lennon, A., Oviedo-Trespalacios, O., & Matthews, S. (2017). Pedestrian self-reported use of smart phones: Positive attitudes and high exposure influence intentions to cross the road while distracted. Accident Analysis and Prevention, 338–347.

Li, N., Feng, X., & Jimenez, R. (2017). Predicting rock burst hazard with incomplete data using Bayesian networks. Tunnelling and Underground Space Technology, 61. 61-70.

Lord, S., Cloutie, M.-S., Garnie, B., & Christoforou, Z. (2018). Crossing road intersections in old age—With or without risks? Perceptions of risk and crossing behaviours among the elderly. Transportation Research Part F, 282–296.

Ma, Y., Lu, S., & Zhang, Y. (2020). Analysis on Illegal Crossing Behavior of Pedestrians at Signalized Intersections Based on Bayesian Network. Journal of Advanced Transportation.

Mamun, S., Caraballo, F. J., Ivan, J. N., Ravishanker, N., Townsend, R. M., & Zhang, Y. (2018). Identifying association between pedestrian Identifying association between pedestrian considering demographics and traffic context. Journal of Transportation Safety & Security. doi: https://doi.org/10.1080/19439962.2018.1490369

Marisamynathan, S., & Vedagiri, P. (2018). Modeling Pedestrian Crossing Behavior and Safety at Signalized Intersections. Transportation Research Record: JOURNAL OF THE TRANSPORTATION RESEARCH BOARD.

Mukherjee, D., & Mitra, S. (2019). A comparative study of safe and unsafe signalized intersections from the view point of pedestrian behavior and perception. Accident Analysis and Prevention.

Nasar, J. L., & Troyer, D. (2013). Pedestrian injuries due to mobile phone use in public places. Accident Analysis and Prevention, 91–95.

Nasar, J., Hecht, P., & Wener, R. (2008). Mobile telephones, distracted attention, and pedestrian safety. Accident Analysis and Prevention, 69–75.

Priyadarshini, P., & Mitra, S. (2018). Investigating Pedestrian Risk Factors Leading to Pedestrian Fatalities in Kolkata City Roads. Transportation in Developing Economies.

Ren, G., Zhou, Z., Wang, W., Zhang, Y., & Wang, W. (2011). Crossing Behaviors of Pedestrians at Signalized Intersections Observational Study and Survey in China. Transportation Research Record Journal of the Transportation Research Board, 65-73.

Rosenbloom, T. (2009). Crossing at a red light: Behaviour of individuals and groups. Transportation Research Part F, 389–394.

Schwebel, D. C., Stavrinos, D., Byington, K., & Davis, T. (2012). Distraction and pedestrian safety: How talking on the phone, texting, and listening to music impact crossing the street. Accident Analysis and Prevention, 266–271.

Shaaban, K., Muley, D., & Mohammed, A. (2018). Analysis of illegal pedestrian crossing behavior on a major divided arterial road. Transportation Research Part F, 124–137.

Shiwakoti, N., Tay, R., & Stasinopoulos, P. (2017). Exploring Jaywalking at Intersections. Australasian Transport Research Forum. Auckland, New Zealand.

Sobhani, A., & Farooq, B. (2018). Impact of smartphone distraction on pedestrians' crossing behaviour: An application of head-mounted immersive virtual reality. Transportation Research Part F, 228–241.

WHO. (2015). Global status report on road safety 2015. Geneva: World Health Organization.

WHO. (2018). Global status report on road safety 2018. Geneva: World Health Organization.

Xu, Y., Li, Y., & Zhang, F. (2013). Pedestrians' intention to jaywalk: Automatic or planned? A study based on a dual-process model in China. Accident Analysis and Prevention, 811–819.

Zhang, H., Zhang, C., Wei, Y., & Chen, F. (2017). The effects of mobile phone use on pedestrian crossing behavior and safety at unsignalized intersections. International Conference on Transportation Information and Safety(ICTIS). Banff,Canada.

Zhang, W., Wang, K., Wang, L., Feng, Z., & Du, Y. (2016). Exploring factors affecting pedestrians' red-light running behaviors at intersections in China. Accident Analysis and Prevention, 71–78.

Hoque, M. M., Anowar, S., & Raihan, M. A. (2008). *Towards sustainable road safety in Bangladesh.* Paper presented at the International Conference on Sustainable Transport for Developing Countries: Concerns, Issues and Options, Dhaka, Bangladesh, 63–69.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Mateo, CA.

Castillo, E., Gutiérrez, J. M., & Hadi, A. S. (1997). *Sensitivity analysis in discrete Bayesian networks*. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *27*(4), 412-423.

Kjærulff, U., & van der Gaag, L. C. (2013). *Making Sensitivity Analysis Computationally Efficient*. Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI 2000), pages 317-325.

Lucas PJ, V. d.-H. (2004). Bayesian networks in biomedicine and health-care. Artificial intelligence in medicine., (pp. 201-214).

Machine Learning Mastery. (1997). Machine Learning. In *Machine Learning* (p. 184).

Machine Learning Mastery. (2006). In *Pattern Recognition and Machine Learning* (p. 360).

Spiegelhalter, D. J. (1988). Bayesian graphical modelling: a case-study in monitoring health outcomes. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 115-133.

Cowell RG, D. A. (1999). Spiegel halter, DJ (1999). Probabilistic Networks and Ex pert Systems. New York: Springer-Verlag, New York.

Airoldi, E. (2007). Getting started in Probabilistic Graphical Models. PLoS Computational Biology., (pp. 2421-2425).