Association Rule Mining for Driving Behaviors and Road Traffic Accidents in Kuwait

Atheer AlMutairi, Dana AlKandari, Loulwah Shummais, Reem AlAjmi, Takeaki Toma
Department of Industrial Engineering
American University of the Middle East
P.O.Box: 220 Dasman, 15453 Kuwait
30606@aum.edu.kw, 30109@aum.edu.kw, 20161@aum.edu.kw, 23238@aum.edu.kw,
Takeaki.Toma@aum.edu.kw

Abstract

One of the serious problems in Kuwait society is the Road Traffic Accident (RTCs). Our research objective is to find unknown association rules between RTAs and/or driving behaviors from our questionnaire data. An online questionnaire collected RTAs and driving behavior data (n=296). The Apriori algorithm found 1,733 rules, and each association rule was scrutinized. If pedestrians-involved traffic accidents occur, the probability of no experience in driving school was 2.98 times higher. If drivers had impaired vision problems (e.g., myopia), the probability of a pedestrians-involved accident was 2.04 times higher. These findings may indicate the importance of driving school experience and visual acuity test at driving license issuance in Kuwait.

Keywords
Road traffic accidents, driving patterns, association rules.

1. Background

On November 16, 2019, a Kuwaiti citizen was killed, and nine others were injured due to an accident on Wafra road. The wounded victims were between the ages of twelve and seventeen. On the same day, a bicycle collision on Sheikh Jaber Causeway caused five injuries (Kuwait Times, 2019). Ideally, we do not have any accidents. But Kuwait is known for the high rate of RTAs. This project is about road traffic accidents in Kuwait. Many reports about speeding, mobile phone usage, and ignoring the fundamental traffic rules as the root causes in Kuwait (e.g., Alsaleh, 2006). Even police and government agencies are working hard to reduce traffic accidents. Why do we still have so many RTAs? Our objective is to find unknown or hidden association rules between road traffic accidents and/or driving behaviors from questionnaire data. We used the Apriori algorithm and applied the knowledge and skills of Probability and Statistics (IE330), Work Design and Analysis II (IE486), and Safety Engineering (IE558).

2. Literature Review

2.1 Primary Factors

Shope and Bingham (2008) reported that there are many reasons teens are prone to having car crashes. First, a driver's attention could be diverted by other passengers' conversations. Further, driving at night time or on weekends increases car accident rates. Drinking or eating while driving and the non-use of the seatbelts can increase the risk of injury during a crash. However, we do not know which factor is the leading cause behind Kuwaiti youth's high rate of accidents. The lack of safety among young drivers is a well-documented and studied problem. Several studies help us estimate the size and nature of many different aspects of the problem.

Deery and Fildes (1999) used cluster analysis in car accident categories. They categorized five distinct subtypes of drivers based on different levels of driving-related attitudes and behaviors. They are general personality traits and hostility and aggression. Their second study showed that performance on the attentional control task, which is the time to reach the desired speed and the proportion of correct responses, provided the most significant practical factor in differentiating the young novice driver's driving skill subtypes.

2.2 Mental Factors

Alsaleh (2018) reported that road traffic accidents could occur due to young people's reckless driving, significantly associated with anger, anxiety, road rage, and drug usage. Young people may have a wrong concept towards driving and may not recognize how dangerous it could be (Čabarkapa, Čubranić-Dobrodolac, &Antić (2018). They may want to show off their cars, and it can lead to car accidents. In addition, Hortwood and Fergusson (2000) reported that more than 90% of drivers engaged in frequent risky driving behaviors were males. They also reported that male drivers suffered more likely from alcoholism or cannabis abuse. Male drivers were involved in violent crime and who are affiliated with deviant peers. Alsaleh (2006) reported that driving habits, aggressive driving, and road rage were consequences of high rates of anxiety, sensation seeking, and anger. Such factors may cause unsafe and traffic accidents. The most common forms of aggressive driving in Kuwait are tailgating and road rage. Tailgating is when the driver is behind a car and leaving barely any space between them. The rate at which young Kuwaiti drivers were engaging in tailgating was recorded at 95%. Furthermore, the top types of road rage that lead to RTA's in Kuwait include forcing a car of the road while driving (94%), passing stop signs (96%), purposeful obstructions on the road (93%), and passing red lights (88%).

Dahlen and Martin, Ragan, and Kuhlman (2005) examined whether risky and aggressive driving should be studied along with the personality traits, sensation seeking, and driving anger, as they are usually studied separately. They asked three hundred and fifteen college students to participate in surveys based on the above traits, plus the average number of miles they drove in a week. It was useful to use these predictors in one study to explain different aspects of driving behaviors.

On the other hand, Palinko et al. (2010) used a remote eye-tracking device to assess cognitive load in a simulated driving environment. Their results showed that the driver's cognitive load was the lowest during the first set of twenty questions due to the uneasiness of the drivers, the driver was given during the driving process, the cognitive measurement during the second set of the question presented to the drivers was significantly different in results due to the change of mental state. The study result indicated that the driving performance of the test subjects was too rough and does not meet fast mental adjustments properly (where drivers can react to sudden circumstances on the road) where subjects could focus on two operations at the same time. It took approximately 4.6 seconds on average to complete both driving and answering questions turns. This research indicates that driving with multitasking can decrease the response time.

2.2 Multitasking Factor

Strayer, Drews, and Johnston (2003) revealed the effect of multitasking; even driving with a hands-free phone affected drivers' reaction time for braking, and it increases the chance to crash with the front car. According to Hyman, Boss, Wise, McKenzie, and Caggiano (2010), an individual's driving performance became weaker when their attention was divided. When a subject was walking or driving, cellular phones for either texting or talking increased the risk of accidents due to difficulties staying focused. Individuals also become less likely to notice new and distinctive stimuli. This phenomenon is known as inattentional blindness. People who walked, listened to music, or talked with other passengers had even less inattentional blindness compared to cell phone users. When people drive while talking on the phone, their driving performance became poorer, as they were slow to respond to a car braking in front of them, are less likely to notice stimuli in their environment, and fail to notice signal changes. Thus, conversing on the phone while driving was regarded to lead to visual attention impairment.

2.3 Pedestrian Factors

According to Wood Tyrrell, and Carberry (2005), the primary safety road issue is pedestrian fatalities. In 2003, four thousand seven and forty-seven pedestrians were killed, Whereas seven thousand pedestrians were injured due to road safety issues and problems. Additionally, the visibility of pedestrians at night can be nonexistent, and it can be enhanced by aids of visibility such as head lightening systems and street lights. Furthermore, several studies were conducted by Wood, Tyrrell, and Carberry (2005) to address how to use conspicuity aids to enhance the pedestrian's safety.

Divers and policymakers may not pay enough attention to the safety of pedestrians. The non-industrialized nations and road users were the most remaining vulnerable pedestrians. Koushki and Ali (2003) investigated four thousand

and forty-three injuries. The leading causes of these accidents were the failure of drivers to spot pedestrians from a safe distance.

3. Methodology

3.1 Overall approach

The overall approach for finding association rules between road traffic accidents and/or driving behaviors is shown in Figure 1. The first step was the questionnaire design. We adopted questions from a study conducted by Al-Hemond et al. (2010). We added demography questions, including gender and age group, in our questions. The second step was to conduct a questionnaire survey. The third step was data cleaning. Raw data obtained from our questionnaire survey was cleaned by identifying, correcting, or removing data entry errors. The fourth step was data mining. Finally, their analysis result was integrated and interpreted.

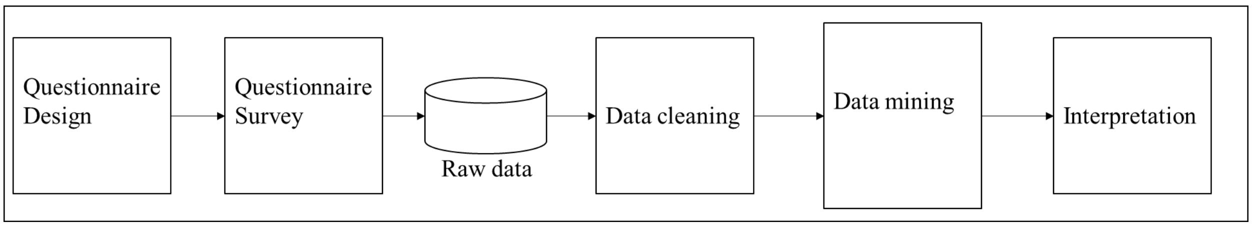

Figure 1 Overall approach

3.2 Questionnaire Design

The following questions were used for our online survey (Table 1).

Table 1 Questionnaire Content

No	Questions
1	What is your age?
2	What is your gender?
3	How do you drive when the weather condition changes?
4	With whom did you learn to drive?
5	Do you have any vision problems?
6	Do you wear glasses or contact lenses when you drive?
7	Do you wear a seat belt?
8	How many times did you pass through the red light in the last one year?
9	Do you drive over the speed limit when you experience time pressure?
10	Do you use a mobile phone while driving?
11	What is your multitasking in driving?
12	Do you have daydreaming while driving?
13	How many speeding violations did you receive during the past 24 months?
14	What is your education level?
15	How do you usually react when you are overtaken by another car?
16	Does the over-speeding traffic cause you to speed as well?
17	What kind of accidents have you ever get involved in?
18	How many hours do you sleep per day?
19	What are you afraid of while driving?
20	Are you confident about driving?

3.3 Survey Method

The online survey was distributed to faculty members, students, their family members, and friends, and three hundred and six records were collected. After data cleaning, we could use n=296 datasets for our analysis.

3.4 Apriori Algorithm

The Apriori algorithm was used for identifying any association rules of driving behavior and accidents among Kuwaiti drivers. Apriori algorithm was initially developed for finding the association rules in market basket analysis. (Agarwal, R., & Srikant, R. ,1994). In this study, an association is defined as the co-occurrence of two or more question responses. For example, a reckless driving experience may be associated with a male, young, sports car, and night driving. It is not an easy task to find an association manually. First, the combination of 20 questions could exponentially increase, and it could be a difficult task for an investigator to check each association. Given 20 questions answered "Yes" or "No" in a set, there are $2^{20} = 1048576$ possible item sets. We need to make 190 pivot tables for finding associations between just two questions and 1140 pivot tables for finding associations between three questions. Second, an association may not always be strong. In the Kuwait traffic accidents questionnaire, each question can be associated with each question else. However, most of those associations are weak. The Apriori algorithm was used to find association patterns from large, sparse data set to extract insight into the road traffic accidents in Kuwait. In the Apriori algorithm, a set of items that co-occurred in observation is called a transaction. The apriori algorithm is used for data pertaining RTAs and road traffic behavior.

3.4.1 Support

The support criteria are about the probability of an item set in all the transactions.

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Total\ number\ of\ transactions} \tag{1}$$

3.4.2 Confidence

The confidence criterion is the conditional probability when two items are present in the shopping cart.

$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Transactions\ containing\ X} \tag{2}$$

3.4.3 Lift

The Lift criterion is the conditional probability divided by the fraction of transactions containing the item.

$$Lift(\{X\} \rightarrow \{Y\}) = \frac{(Transactions\ containing\ both\ X\ and\ Y)/(Transactions\ containing\ X)}{Fraction\ of\ transactions\ containing\ Y} \tag{3}$$

In this project, `arules` package in R was used for the apriori analysis.

4. Result

4.1 Association Rules related to Road Traffic Accidents

The Apriori algorithm found 1,733 rules. We used `inspect(subset(rules, lift>x)` R command to extract the association rules. The obtained rules were saved in CSV data format, and we scrutinized the rules with Excel. Two useful association rules were found that related to pedestrians-involved traffic accidents.

The first useful rule was about the driving school experience and pedestrians-involved accidents. R command `arulesViz` generated a graph of the top fifteen association rules mined from our dataset related to pedestrians-involved accidents (Figure 2). Each pink-colored circle shows a rule between two green-colored items. The pink circle size indicates the magnitude of each rule's support value, while the color darkness indicates the level of lift value. A dark red circle between "Training School=No" and "Pedestrians Accidents=Yes" indicates a strong link. The pedestrians-involved road traffic accidents were usually rare events (support=0.023). However, if it occurs, the driver is probably learned to drive without a driving training school (Confidence = 0.875 and Lift= 2.98). The second useful association rule was that people with compromised vision had a higher rate of pedestrian-involved accidents. Compromised vision problems (e.g., myopia, hyperopia, and astigmatism) and pedestrian accidents had a lift value of 2.04. Figure 3 indicates the association rule between self-learning of driving and other factors. The lift value was 3.41 for no driving training school experience when they learn driving by themselves.

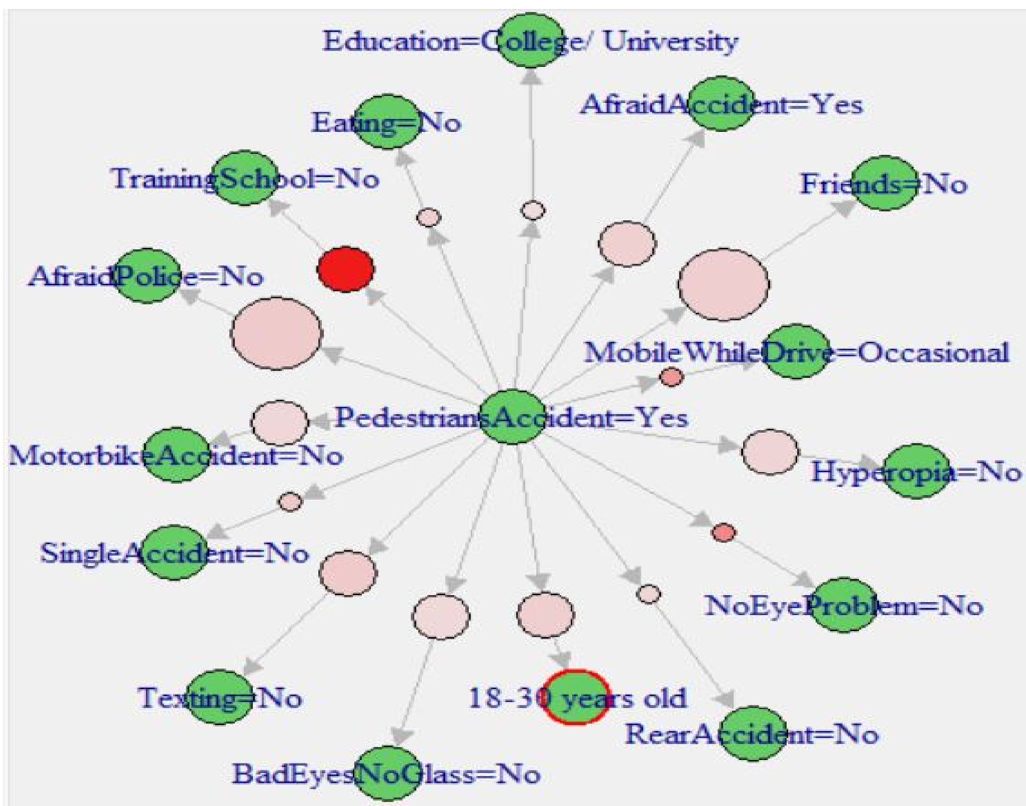

Figure 2 Association Rule Between Pedestrians-Involved Accidents = Yes and Other Factors

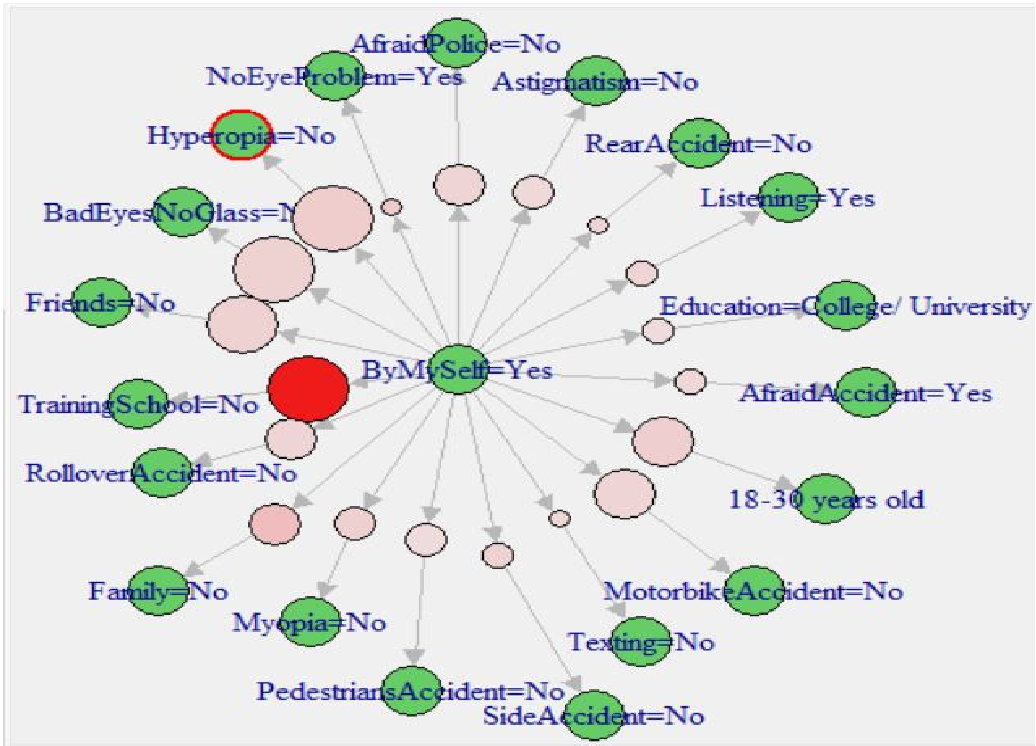

Figure 3 Association Rule Between Learning Driving ByMySelf = Yes and Other Factors

4.2 Association rule related to driving behavior

Some people with myopia drove their cars without wearing glasses, and mostly they were drivers with myopia vision. Figure 4 shows the association rule related to “BadEyesNoGlass = Yes” item. The “BadEyesNoGlass” is the data analysis code representing driving without wearing glasses even if they have compromised vision (such as myopia). The lift value of the association rule between “BadEyesNoGlass = Yes” and “Myopia = Yes” was 4.71. Figure 5 shows that everybody with hyperopia or astigmatism visions wear glasses, while eleven percent of myopia people did not wear glasses when they drive their cars.

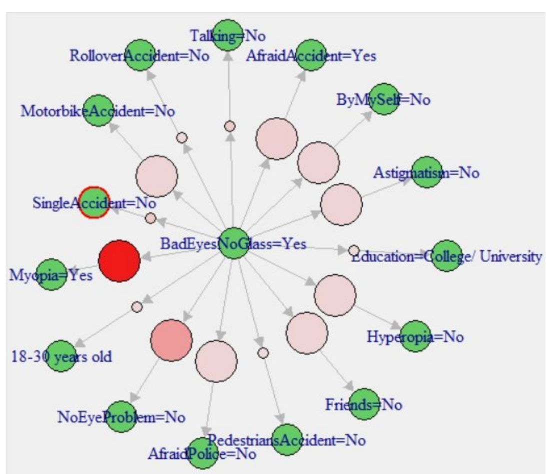

Figure 4 Association Rules Between BadEyesNoGlass = Yes and Other Factors

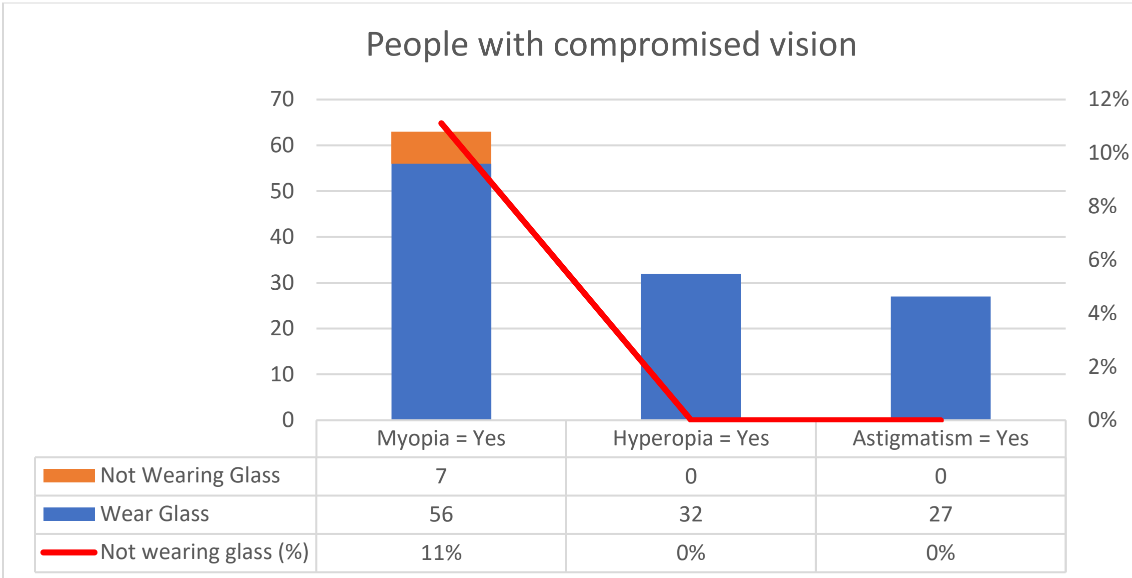

Figure 5 People with compromised vision

5. DISCUSSION

5.1 Importance of driving school

Our association rule study indicated that driving school experience is related to pedestrian-related road traffic accidents. A lack of driving school experience was also related to destructive driving manner. For example, people with no training school reported wearing seatbelts 22% less than people with training school experience.

5.2 Myopia and Accidents

We also found an association rule between myopia and pedestrians-involved accidents. If we know that drivers have vision problems, then the probability of pedestrians-involved accidents was 2.04 times higher. Vision problems may be one of the important factors that lead to having an accident. However, the Kuwait government requires glasses to new drivers on issuing a new driving license. The Kuwait government may need to check the driver's vision problems more rigidly when people renew their driving licenses.

6. CONCLUSION

We used the Apriori algorithm to find the association rules between a road traffic accident and Kuwait's driving behaviors. We discovered that pedestrian-related accidents were related to a lack of driving school experience. Pedestrian-related accidents were also associated with the myopia problem in Kuwait. Eleven percent of myopia people did not wear glasses while driving, while all hyperopia and astigmatism people reported wearing glasses. This information may be helpful for the law-enforcement agency and the government reduces road traffic accidents in Kuwait. Since this is a retrospective study and our sample size is small, further study is needed.

References

Agarwal, R., and Srikant, R., Fast algorithms for mining association rules. In Proc. of the 20th VLDB Conference (pp. 487-499), 1994

Alsaleh, A., The impact of social and psychological factors on car accidents in Kuwait. Digest of Middle East Studies,15 (1), 1-17., 2006

Alsaleh, A., Road Traffic Accidents in Kuwait City: A Triangulation Approach Journal of Social Sciences and Humanity Studies, 4(5)7-18, 2018

Al-Hemoud, A. M., Simmons, R. J., & Al-Asfoor, M. M., Behavior and lifestyle characteristics of male Kuwaiti drivers. Journal of safety research, 41(4), 307-313., 2010

Čabarkapa, M., Čubranić-Dobrodolac, M., Čičević, S., &Antić, B., The Influence of Aggressive Driving Behavior and Impulsiveness on Traffic Accidents., 2018

Dahlen, E. R., Martin, R. C., Ragan, K., & Kuhlman, M. M., Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving. Accident analysis & prevention, 37(2), 341-348., 2005

Deery, H. A., & Fildes, B. N., Young novice driver subtypes: Relationship to high-risk behavior, traffic accident record, and simulator driving performance. Human factors, 41(4), 628-643., 1999

Horwood, J. L., & Fergusson, D. M., Drink driving and traffic accidents in young people. Accident Analysis & Prevention, 323(6), 805–814., 2000

Hyman, I. E., Boss, M. M., Wise, B. E., McKenzie, K. M., &Caggiano, J. undefined., Did You See the Unicycling Clown? Inattentional Blindness while Walking and Talking on a Cell Phone. Applied Cognitive Psychology, 24, 597–607. DOI: 10.1002/ACP.1638, 2010

Koushki, P. A., & Ali, M. A., An analysis of pedestrian accidents in Kuwait. Kuwait Journal of Science and Engineering, 30(1), 139-152., 2003

Kuwait Times., Kuwait Times. Retrieved from https://news.kuwaittimes.net/website/man- killed-in-wafra-road-crash-bikers-injured-in-causeway-accident/., 2019

Palinko, Oskar, Kun, Andrew &Shyrokov, Alexander and Heeman, Peter., Estimating cognitive load using remote eye tracking in a driving simulator. Eye Tracking Research and Applications Symposium (ETRA). 141-144. 10.1145/1743666.1743701., 2010

Shope, J. T., & Bingham, C. R., Teen driving: motor-vehicle crashes and factors that contribute. American journal of preventive medicine, 35(3), S261-S271., 2008

Strayer, D. L., Drews, F. A., & Johnston, W. A., Cell phone-induced failures of visual attention during simulated driving. Journal of experimental psychology: Applied, 9(1), 23., 2003

Wood, J. M., Tyrrell, R. A., & Carberry, T. P. Limitations in drivers' ability to recognize pedestrians at night. Human factors, 47(3), 644-653., 2005

Biographies

Atheer AlMutairi, Dana AlKandari, Loulwah Shummais, Reem AlAjmi are recently graduated students from the American University of the Middle (AUM) in Kuwait, majoring in Industrial Engineering. During their four years of Bachelor's degrees, they gained several engineering and computational skills. They worked with computer software such as MS Office, AutoCAD, Minitab, MATLAB, Arena, Jack, and Visual Studio. They participated in many AUM academic activities. In addition to their major graduation project presented in this paper, they worked on several course projects in manufacturing processes, safety, ergonomics, operations research, quality control, simulation, and lean six sigma.

Takeaki Toma is an Assistant Professor in Industrial Engineering at the American University of the Middle East, Kuwait. He earned a BS in Information Engineering from the University of the Ryukyus, Japan, Masters in Industrial and Management Engineering from Montana State University, and a Ph.D. in Industrial Engineering from Oregon State University. He has experience in both industry and academia in both United States and Japan. His research interests include Cognitive Engineering, Safety Engineering, Quality Control, Statistical Data Analysis, and Machine Learning. He is a member of IEOM and IISE.