Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

USE OF MACHINE LEARNING FOR DRIVER´S BEHAVIOR AND CONCENTRATION

Wilson Arias-Rojas, PhD
Universidad Nacional de Colombia Sede Medellín: wariasr@gmail.com wiariasro@unal.edu.co

Jorge Eliecer Córdoba Maquilón, PhD
Professor Universidad Nacional de Colombia: sede Medellín jecordob@unal.edu.co

Abstract:
This research is the result of the analysis of drivers´ behavior in a controlled scenario, using a driving simulator, in which by measuring brain waves, the degree of concentration was measured when driving and through the use of neural networks. Driver behavior model was proposed to be subjected to a distracting effect while driving, which allows analyzing the most relevant factors that are reflected in errors and bad practices at the time of driving.In this research it was determined a population sample of men and women whose ages oscillate between sixteen to ninety years. A driving simulator was built, and it was using a software for the simulation that allows different driving scenarios. Finally, risk behaviors were classified to be a factor of distraction.

Keywords: Driver´s behavior, Simulation, Concentration, Neurosky, human behavior, risky behaviors

An Introduction to the generation of traffic accidents

The study of the frequency and causality of traffic collisions in public roads is a prominent topic due to its impact and cost to general society. Each year, according to statistics from the now extinct Colombia's Highway Prevention Fund, on average, every 2.5 minutes there is a traffic accident, every ten minutes there is a traffic accident with wounded victims, and every hour there is a fatal victim; mainly due to the poor highway safety features in public roads in Colombia. These statistics point out that the segments of the population at a higher risk are: youth between fifteen to thirty years of age, pedestrians, cyclists, and particularly in the last five years, motorcyclists.

According to Ferrer et al., (2003), the statistics for traffic accidents and fatalities that are gathered by Colombia's National Police, National Administrative Department of Statistics and the National Institute for Legal Medicine and Forensic Sciences do not add up.

In the above referenced statistics, speeding is identified as the main cause for traffic accidents. 895 of the traffic accidents, are related to human error, while the remaining 11% are related to mechanical failure, weather, and public road defects among others (Arias, W. and Colucci, B., 2006). The majority of traffic accidents happen five minutes from destination, in average, as the driver relaxes once close to destination and driver attention diminishes as a result. (Arias, W. and Colucci, B., 2006). Nonetheless, the

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

authors didn't take into account that the road conditions contribute to about 28 to 34% while the vehicle's safety devices contribute to about 8 to 12% (Rumar, 1985 and Johnston, 2006).

Traffic Accident severity can be related to sudden changes in the speed limit in segments of the road, which could be involuntarily disregarded by the driver, if the geometry or the general conditions of the road do not vary considerably. This can be pronounced in a segment of the road where speed limits are not congruent to the driver's expectations, when considering the characteristics of said segment. Also, to analyze the causes of traffic accidents, factors such as education, engineering and incident management must be accounted for.

It can be considered that a driver in an urban area, where the speed limit is sixty kilometers per hour (37.3 mph), and then changes to a rural area where the speed increases to 100 kilometers per hour (62.1 mph), and encounters a construction zone with a sudden, substantial reduction in speed limit to thirty kilometers per hour (18.6 mph), there is a possibility the driver wouldn't react accordingly, and create a hazardous condition leading to a traffic accident with wounded or fatal victims.

Lack of traffic safety devices or its maintenance, such as inadequate signage and pavement marking, are a possible cause for the increase in traffic accidents with fatalities; which have a very high cost for society and the economy of a country.

An increase in the price of gasoline can have a direct reduction in trips during certain periods of time, increasing or reducing the probability of traffic accidents. Intense police patrolling can also result in a transient change in driver behavior, given the perceived risk of fines. Shinar, Stiebel (1986) and Benekohal *et al*. (1992) demonstrated the transient reduction in driver speeds given an increase in police presence in the roads. Vaa (1997) demonstrated that intensive police patrolling in a particular segment of road will have speed reducing effect in the drivers for a maximum duration of eight months.

In relation to the deaths associated to speed, 54% of the accidents happened in segments of the road with a speed limit of 56 kilometers per hour of les (NCSA, 2006). AASHTO (2004) recommends the speed limits to be consistent with what users expect, based on the geometric features of the road. Nonetheless, it begs the question as to what relationship there is between design speeds, geometric features of the road, and marked speed limits in these segments, that may be a contributing factor to these fatalities. It is necessary to evaluate and analyze the relation between the geometric design of a road, and the perceived risk of traffic accidents by the drivers in these segments.

Research conducted by the University of Puerto Rico, Mayagüez Campus, demonstrated that 70% of the accidents with pedestrians take place in urban areas, while 98.5% take place outside of intersections and 73% happen in flat, straight

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

segments of the road (Alicea, 2004). This research also had short, mid and long-term recommendations to mitigate pedestrian safety issues in Puerto Rico. Some of these recommendations are to establish public outreach and education strategies, the installation of pedestrian barriers, the installation of pedestrian traffic signage in select segments of the roads, amendments to local traffic laws and traffic calming techniques to reduce the speed of the vehicles.

According to Regan et al., (2008) the known advantages and disadvantages is that driver simulators:

- **Advantages**

1. Have the capacity of creating likely scenarios of crashes without damage or risk, such as for fatigue, driving under the influence, extreme climate conditions, or early adoption of new technology, among others.
2. Allow for the control of many variables that may cause driver confusion, such as weather, traffic, illumination, wind, potholes, erratic or spontaneous behavior from other users in the road, among others.
3. since driver-perceived data is known and can be reliably reproduced using simulators (Gibson, 1986), not all real-world sensory data is perceived by the driver in an usable way, or in a way that will impact driver's decision making process or behavior.
4. Can be reproduced identically, for all subjects participating in the experiment.
5. Provide inexpensive flexibility and configurability that allow for the modelling of diverse research scenarios (Jamson, 2001).
6. Help expose a wide array of relevant research topics, even with low-fidelity, low cost simulators.
7. Are convincing and stimulate adequate reactions in drivers that are similar to those in real driving conditions.
8. Are an excellent tool to evaluate the performance of drivers, or measure what a driver can do (Evans, 2004).
9. Can be configured as a tool for training of new drivers, to help in the development of certain skills, to then bring these to real driving conditions (Pollatsek et al., 2006).

- **Disadvantages**

1. Since simulated accidents do not have the exact same consequences as a real accident, subsequent behavior can be affected. Accidents in a simulator can produce an unknown psychological impact in the participants.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

2. Confusion variables or interactions that take place in real world dynamics must be understood, and since these cannot be fully recreated in simulator models with current technology, are not fully susceptible to being tested.
3. Real driving scenarios cannot be perfectly reproduced.
4. Each event or decision in the driving simulation will impact later driving decisions or events.
5. High end simulators will require cutting edge hardware, and development of adequate software, will in turn only approach limited variables.
6. More affordable, lower end simulators can be imprecise, or pose issues of application flexibility, hence not all research queries would be addressed with these.
7. Drivers often do not believe in the authenticity of simulators.
8. Simulators cannot address questions related to driver behavior, since the driver doesn't drive as them would in their own vehicles (Evans, 2004).
9. Since the driver's level of experience and skills that are transferred into the simulator are unknown, the software's cost-effectiveness is affected (Jamson, 2001).

- **Materials and Methods**

Materials used in this experiment are described as follows:

- Neurosky MindWave Mobile2 Headset used to capture brain waves.
- Matlab software, to register and keep record of the brain waves captured by the Neurosky headset.
- Foerst Simulation Sowftware, to perform the simulations.
- Simulation Station, comprised of the following:
  - PC unit with Intel Core I7 7700ATX - Board MSI-H110M - 16GB DDR4 - 1 TERA HDD-unit DVD-RW, Keyboard and Mouse, Solid state drive SSD ADATA SPP550 - 240GB, Graphics card: PNY GTX 1070 TI 8GB DDR5, Bluetooth.
  - Logitech Steering wheel and pedals, model G29, with adapter to fit a regular steering wheel.
  - Chassis with structural square steel tubing, with support to fit a driver seat, steering wheel and pedals mounts, screen mount, paint finished.
  - TV LED SAMSUNG 32" – Model UN32J4000DKXZL
  - Logitech Z506 Speakers

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

- This simulator was built by the author, as part of this PhD research project.
- Alcatel U5 cellular phone

Figures 1, 2 and 3 show the Neurosky headset employed, a screen capture of the simulator in action, and the simulator built for this research.

.

Figure 1 - Neurosky Headset

Source: www.neurosky.com

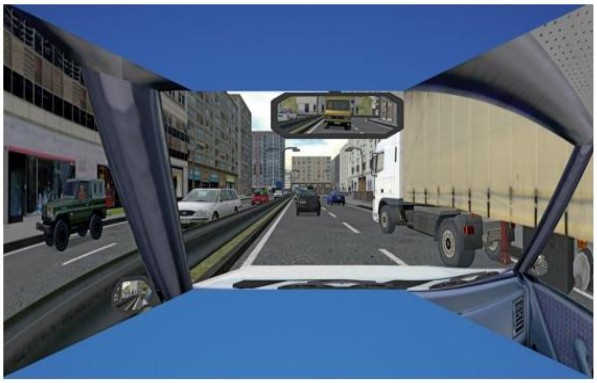

Figure 2 - Driver simulator in urban setting

Source: http://www.fahrsimulatoren.eu/

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Figure 3 - Actual driver simulator built for this research

Source: the authors

The population sample that participated in this research project met certain basic criteria, which are presented below:

• Have a valid driver's license

• Have driven for at least 1 year

• Have driven for at least 1 week in the last year of the test

• Do not have any physical or cognitive disorders

• Do not have central nervous system diseases

• Do not be pregnant

• Not having consumed alcoholic beverages or hallucinogenic substances before the test

Simulator disease is a phenomenon that is affected by the simulator's specifications and characteristics of the participants. It produces similar symptoms, but typically, participants who drive in the driving simulator experience dizziness, nausea, eye discomfort, and disorientation (Kennedy et al., 1993).

Driving area conditions

The scenario for running the test drive was the same for all 167 participants in the experiment. It was conducted in an urban area, with rush hour traffic, between 6:00 PM and 7:00 PM, in an environment inside and outside urban areas, with pedestrians and

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

vehicles circulating on stage. The urban area has platforms, shopping areas, roadside parking areas, gas stations, traffic signals and roundabouts throughout the development of the route.

Transit scenario

The effect of traffic flow and its distracting effect while driving is important in this research. The simulation of the environment (for example, the behavior of other vehicles in the simulation of the road network of the experiment) can be very complex in some cases. It is important to observe the simulation of the behavior of no more than 1 or 2 vehicles in relation to the simulated vehicle in this experiment. The simulation of the traffic environment is much more demanding than the classic traffic micro simulation model, for this reason, the software has implemented a similar movement framework as when driving naturally, in this way, the environment is not static. The distribution of the participants in the experiment, by age, gender, socioeconomic status and level of education. 39 participants were women and 128 were men, for a total of 167 individuals.

Distribution of participants by group (gender and age)

Table 1 shows the distribution of the participants by range and age, with the largest female population concentrated in the range of 31 to 40 years and the male population in the range of 21 to 30 years, followed by the range of 31 to 60 years.

(Table 1. Distribution of participants by gender and age)

Age	Female	Male	Grand Total
0 to 20	2	15	17
21 to 30	8	28	36
31 to 40	15	23	38
41 to 50	7	23	30
51 to 60	3	23	26
61 to 70	2	7	9
71 to 80	1	6	7
81 to 90	1	3	4
Grand Total	**39**	**128**	**167**

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Figure 4 show distribution for gender and age of all participants in this research (23.25% female and 76.35% male)

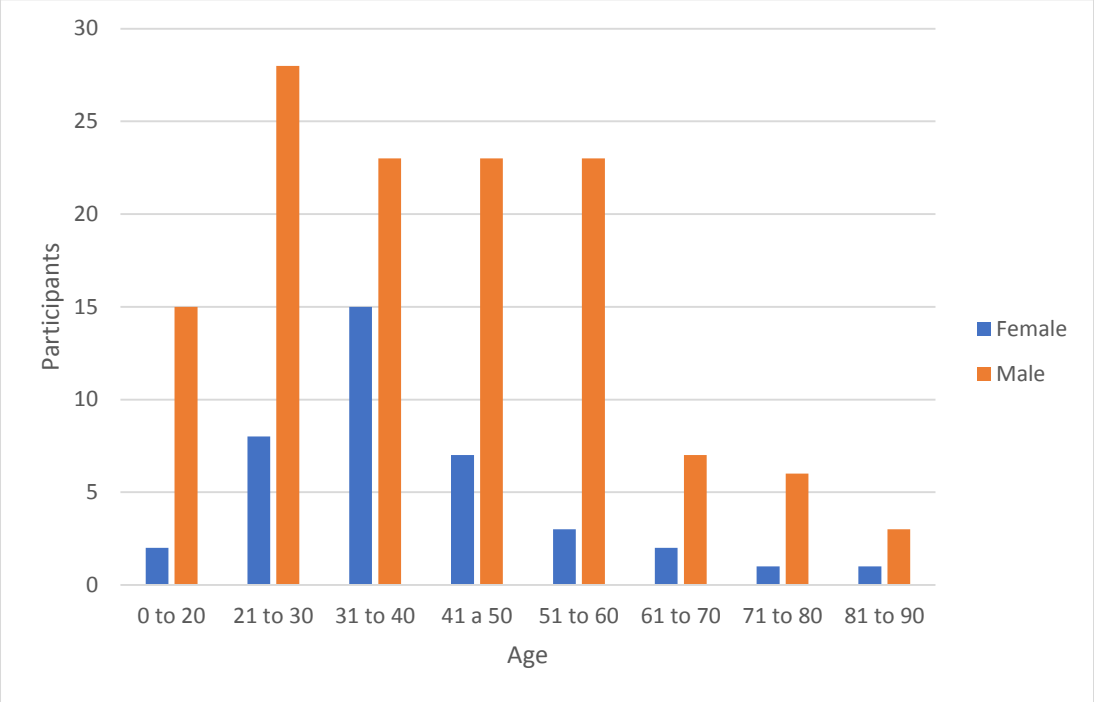

Figure 4. Distribution of participants by gender and age range)

Source: the authors

The socioeconomic stratum is a classification of residential properties in Colombia to charge home public services differentially by strata, subsidizing the lower strata (1 and 2) and taxing the higher strata (5 and 6). (In the table 2). Therefore, those who receive a higher income must pay more for home public services in order to subsidize the lower strata.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Table 2: Distribution of participants by socioeconomic stratum.

Socioeconómic stratum	Female	Male	Grand Total
1		5	5
2	7	30	37
3	28	71	99
4	3	17	20
5		5	5
6	1		1
Grand Total	**39**	**128**	**167**

Figure 5 show distribution for socioeconomic stratum of all participants (stratum 1, 3%, stratum 2, 22%, stratum 3, 59.4%, stratum 4, 12%, stratum 5, 3% and stratum 6, 0.6%)

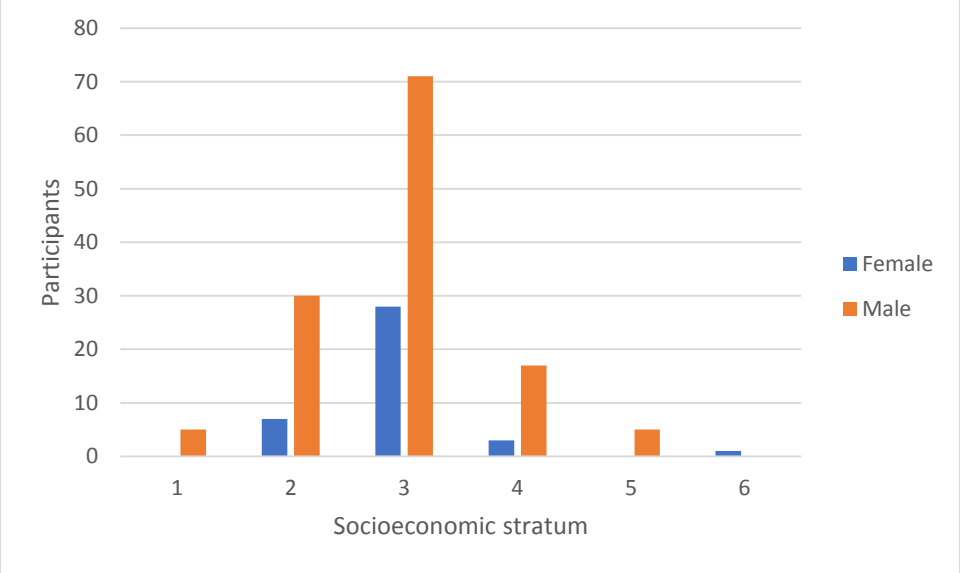

Figure 5: Distribution of participants by socioeconomic stratum.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Source: the authors

Table 3: show the scholarity level for gender of all participants in this research. From elementary school to doctoral degree for gender.

Table 3. Distribution of participants by scholarly level.

Scholarly level	**Female**	**Male**	**Grand Total**
Elementary School	1	9	10
High school	10	33	43
Asocciate degree	12	26	38
BSc	12	52	64
MSc	4	5	9
PhD		3	3
Grand Total	**39**	**128**	**167**

Figure 6 :show distribution by scholarity level of all participants (elementary 6%, High school 26%, associate degree 23%, BSc 38%, MSc 5% and PhD 2%)

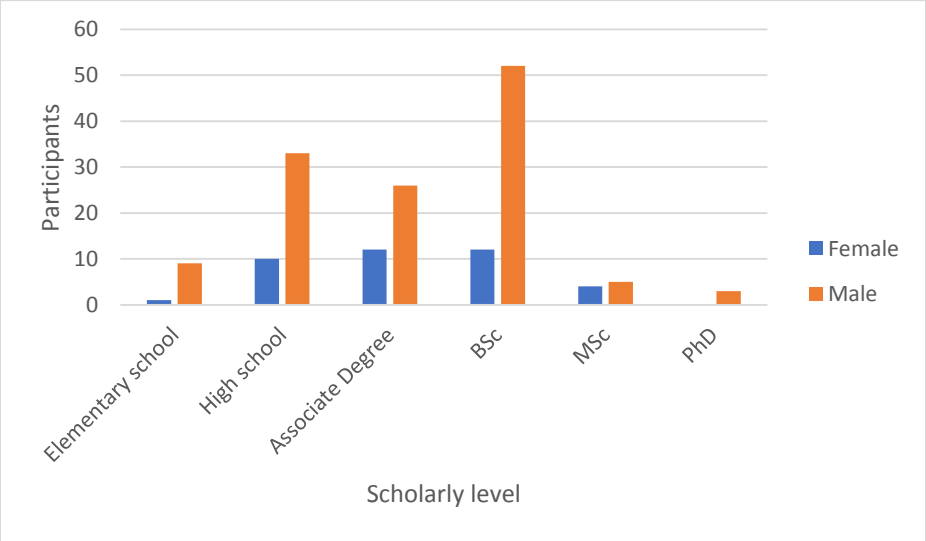

Figure 6. Distribution of participants by scholarly level.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Source: the authors

Twilight (from 6:00 PM to 7:00 PM) in an urban environment was chosen as the setting, with a driving time of 10 minutes under dry weather conditions.

The driver was asked if he wanted to drive the vehicle with a mechanical transmission or an automatic transmission. This information is necessary to configure the simulator software. The driver was fitted with the Neurosky headband on his head and the information obtained from it was recorded via Bluetooth to the computer, in which, with the Neurosky SDK program and the development of the program in Matlab, the information was recorded in real time of the degree of concentration of the driver during the test.

In addition, the simulator program recorded the errors made by the driver during the test.

At the end of the scenario, the data of the subject were recorded including those obtained with the headband, as well as those processed by the simulator, among others the beginning and end time of the test and the time, in seconds, of each of the errors made in the driving. Figure 7 show the driver simulator used for obtain all data in this research.

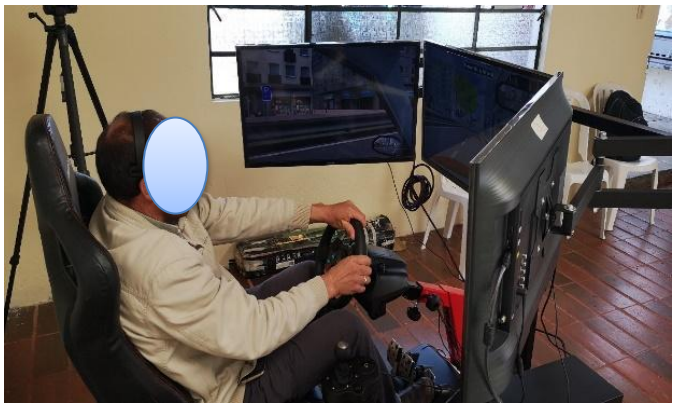

Figure 7. Participant using the driving simulator.

Source: the authors

Results

Information from transportation studies is generally modeled using two types of approximations: classical statistics and artificial intelligence. In statistics, numerical information is collected, organized and interpreted with mathematical tools, particularly when this information is related to population characteristics such as the inference of a sample (Glymour *et al*., 1997). Statistical models have very robust, widely accepted,

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

mathematical foundations and and provide insights into the mechanisms of data creation. However, they often fail when it comes to complex and highly non-linear data. Artificial intelligence combines concepts of learning, adaptation, evolution, and fuzzy logic to create models that are “intelligent in the sense that the structure arises from an unstructured beginning (Engelbrecht, 2007; Sadek *et al*., 2003). Neural networks are an extremely popular class of artificial intelligence. and they have been widely used in different transport problems, particularly because they are very generic, precise and convenient mathematical models, capable of easily simulating numerical components of a model. Neural networks have an inherent propensity to store empirical knowledge and can be used in any of three basic ways (Haykin, 1999): i. As models of biological nervous systems. ii) As real-time adaptive signal processors / controllers. iii) As data analytical methods. In transport investigations, neural networks have been used mainly as data analysis methods due to their ability to work with large volumes of multidimensional data, their modeling flexibility, their learning, their adaptability and prediction that is good in general. (Karlaftis and Vlahogianni, 2010). Quite often, researchers who have mastered one of the two approaches argue fervently in support of "their" chosen method. The selection of the analytical approach is one of the most debated topics in research meetings and publications; and, although these arguments provide interesting scientific debates, they manage to completely confuse younger researchers and, in particular, to professionals who are more interested in the model they should use rather than concentrating on the philosophical or mathematical foundations of the approaches (Karlaftis and Vlahogianni, 2010). In the experimental stage of this project, data from 167 subjects distributed in 39 women and 128 men, were recorded, meeting the age ranges of the population sample and with different types of scholarly level and socioeconomic status.

The total information volume of the entire experiment is summarized with the following mathematical operation:

167 subjects x 30 driving errors x 28,000 records = 140,280,000 average data

After the test, the simulator delivered 30 parameters which are coded according to a number and that can be seen in the file of driving errors or types supplied by the software developer. These parameters are errors or situations that arise while the subject is driving in the time span of the test.

Exploratory analysis of the experimental results

The information on accidents committed by each of the 167 participants in the experiment were analyzed to evaluate the trend in the demographic and combinatorial variables of these variables as presented below. Figure 8 shows the percentage of concentration and deconcentration for the two genders, where it is observed that 79.65% of the participants were concentrated in the experiment and 20.35% of the

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

women were concentrated. 70.37% of the deconcentrated participants were men and 29.63% of the women were deconcentrated in the execution of the experiment.

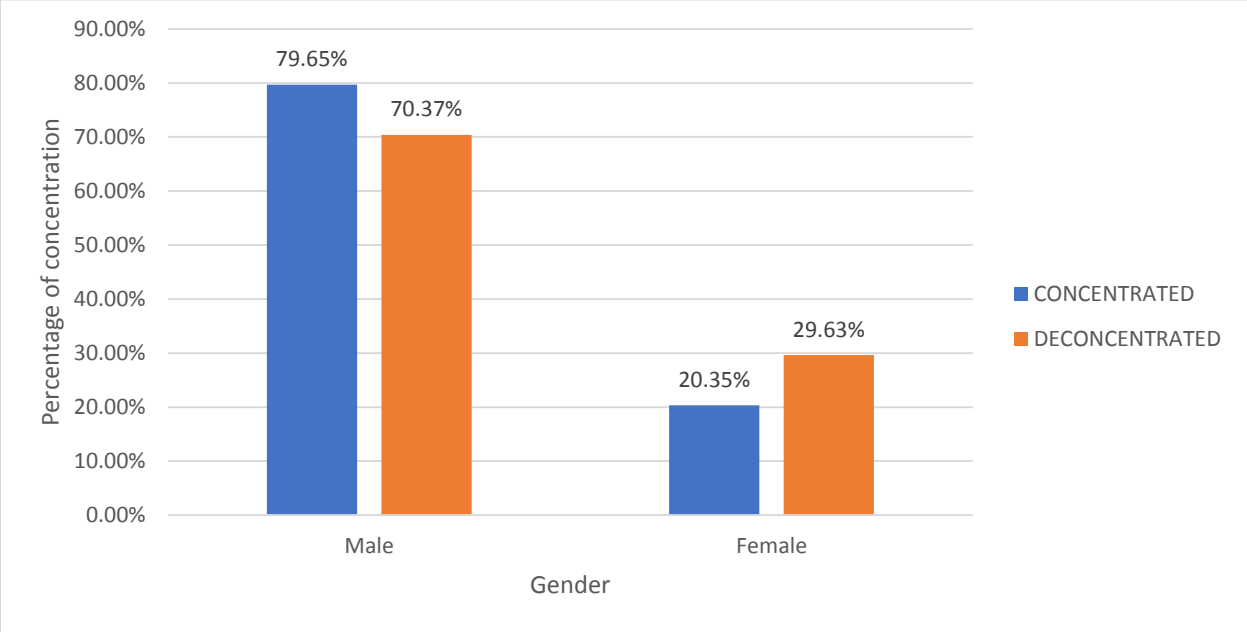

Figure 8: Percentage of concentration by gender

Figure 9 presents the percentage of concentration and deconcentration by socioeconomic stratum of each of the participants in the experiment.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

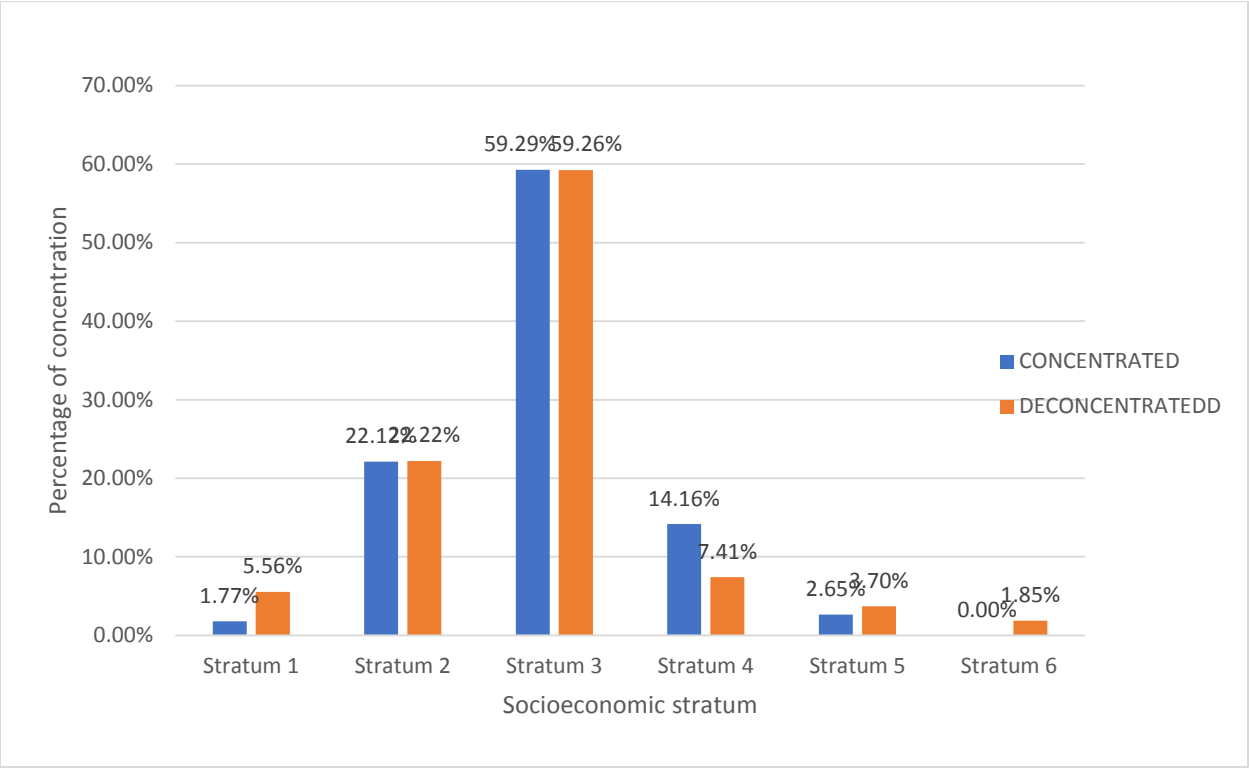

Figure 9: Concentration in percentage by stratum

The figure 10: shows the percentage of concentrated and deconcentrated participants by educational level

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

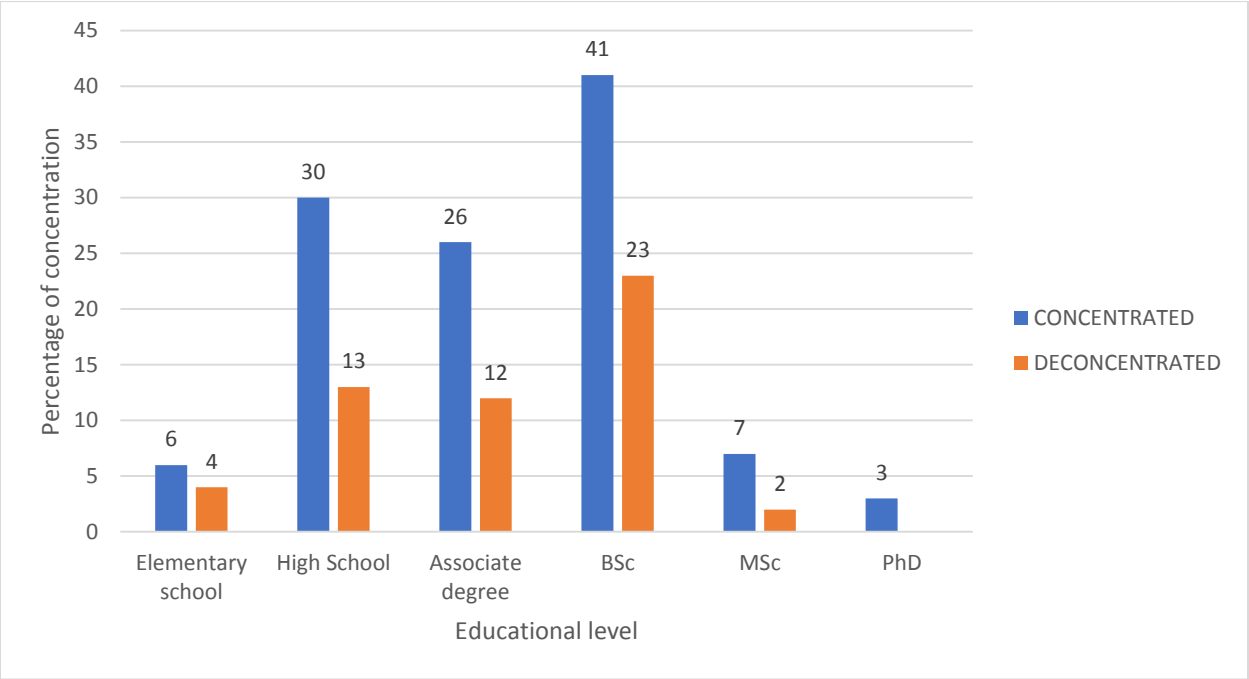

Figure 10: Concentration by percentage by educational level

The figure11 shows the percentage of concentrated and deconcentrated participants by age range

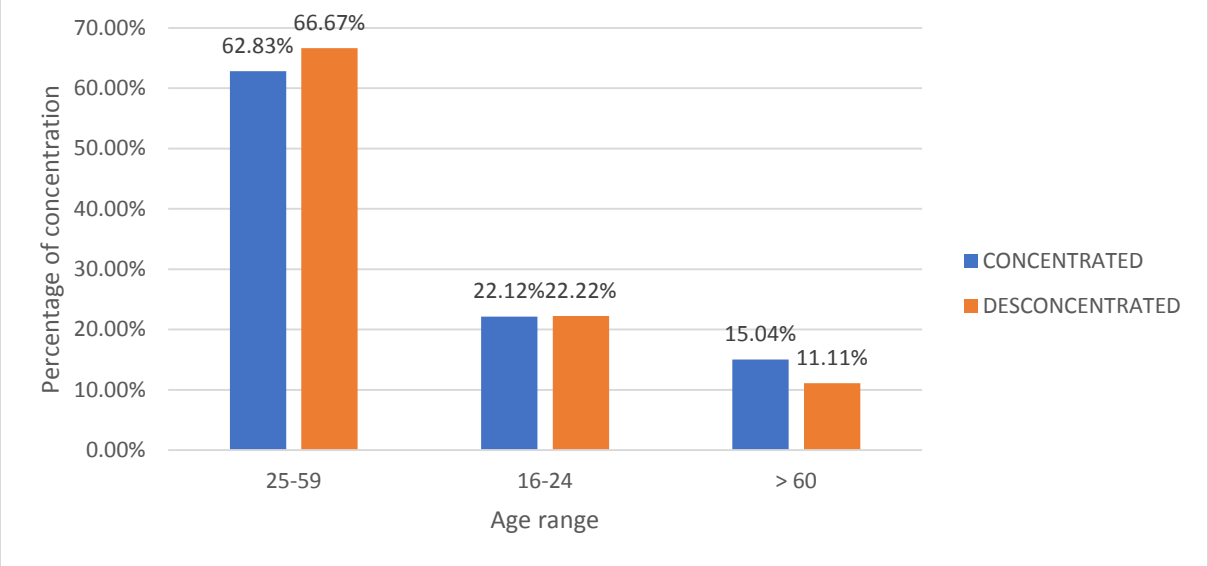

Figure 11: Concentration by percentage by age range

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Figure 12: presents the most common errors made by the participants by age range, during the execution of the experiment. Note that the most common mistake was not activating the turn signals and secondly, the accident.

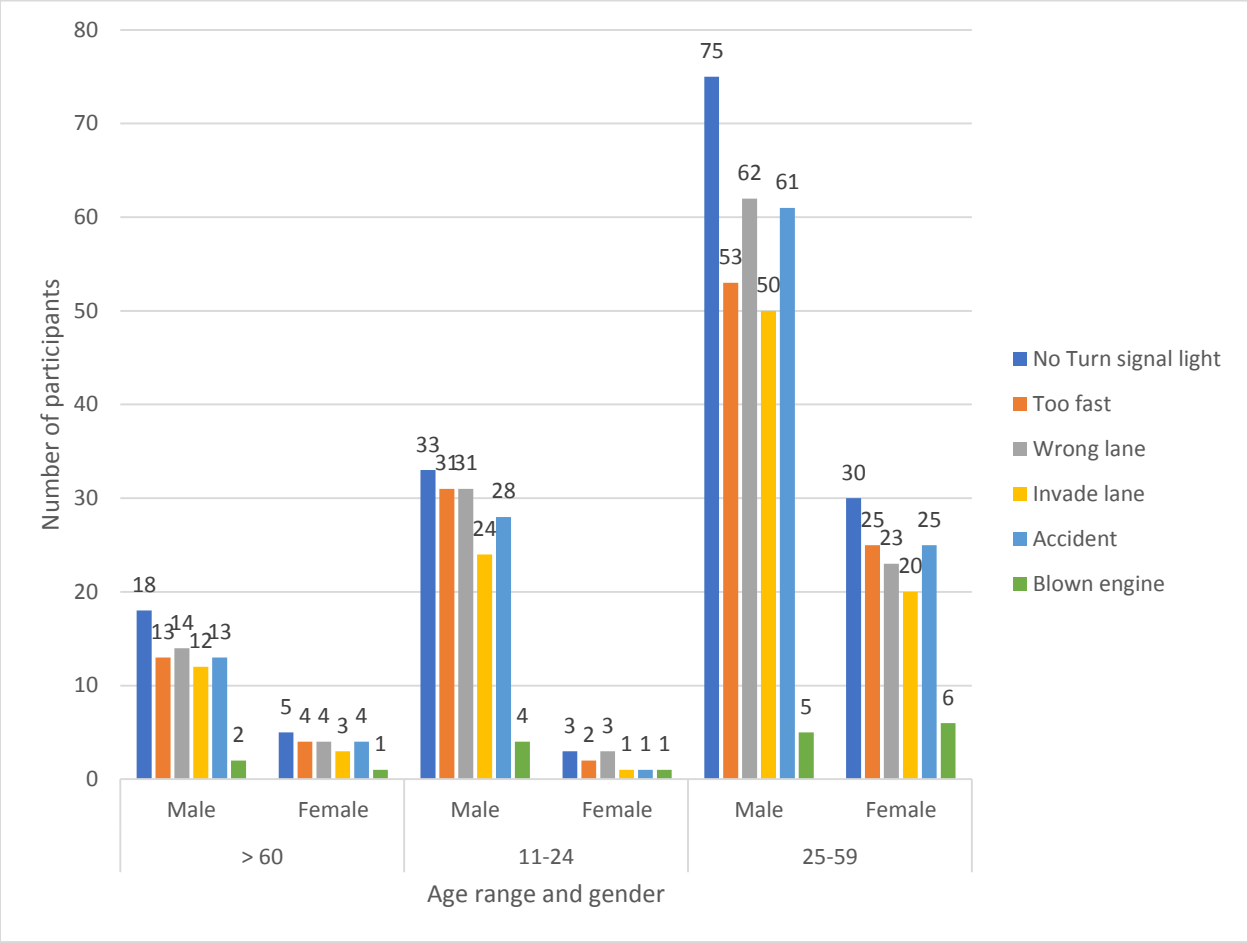

Figure 12: Most common errors by age, by gender

Figure 13: presents the number of concentrated and deconcentrated participants who had accidents during the execution of the experiment.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

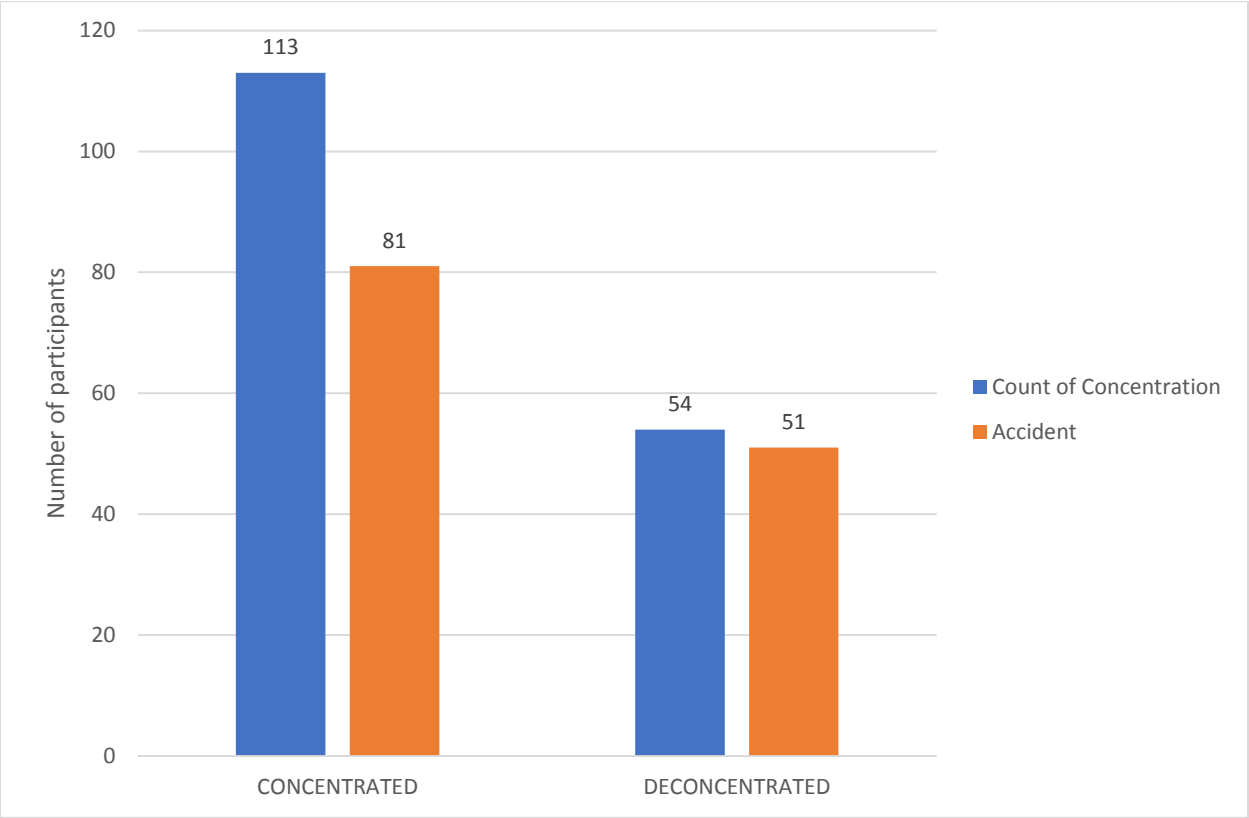

Figure 13: Concentration by accident

Use of Auto machine Learning.

When applying a Stepwise forward regression model, using Auto machine learning, executing logistic regressions, the following values were obtained, where a relationship was made between the intercept and the concentration, finding a significant association or p-value of 0.0211, finding that concentration is a factor that explains the accident rate. In the same way, the significance of the model is shown and observed with the p-value using the Real Statistics V.7.0 software of March 2020, which is the R software V.3.2.6 of March 2020 for Excel. Table 4 contains the values previously explained plus the confidence interval (lower and upper)

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Table 4. Stepwise Forward: Accident Probability

	coeff b	*s.e.*	*Wald*	*p-value*	*OR*	*lower*	*Upper*
Intercept	1.18	0.26	21.25	0.000	3.25		
conc. Norm.	1.05	0.45	5.37	0.021	**2.85**	1.17	6.90

Table 5 shows the values obtained for X^2, p-value, the area under the curve and the precision model.

Table 5. Values for the concentration model.

Chi-Sq	5.84
p-value	0.016
AUC	0.87
Accuracy	0.83

The equation obtained for this model is presented below as well as Figure 9 showing the area under the curve for the model obtained. This equation represents the curve of the model proposed in this research, where the model has a precision of 83% with 87% of the total data obtained.

The equation corresponding to the concentration model is determined by

$$Y= -1.18x_1-1.05x_2$$

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Figure 14: show the curve (true positive rate vs. false positive rate) and equation for this model.

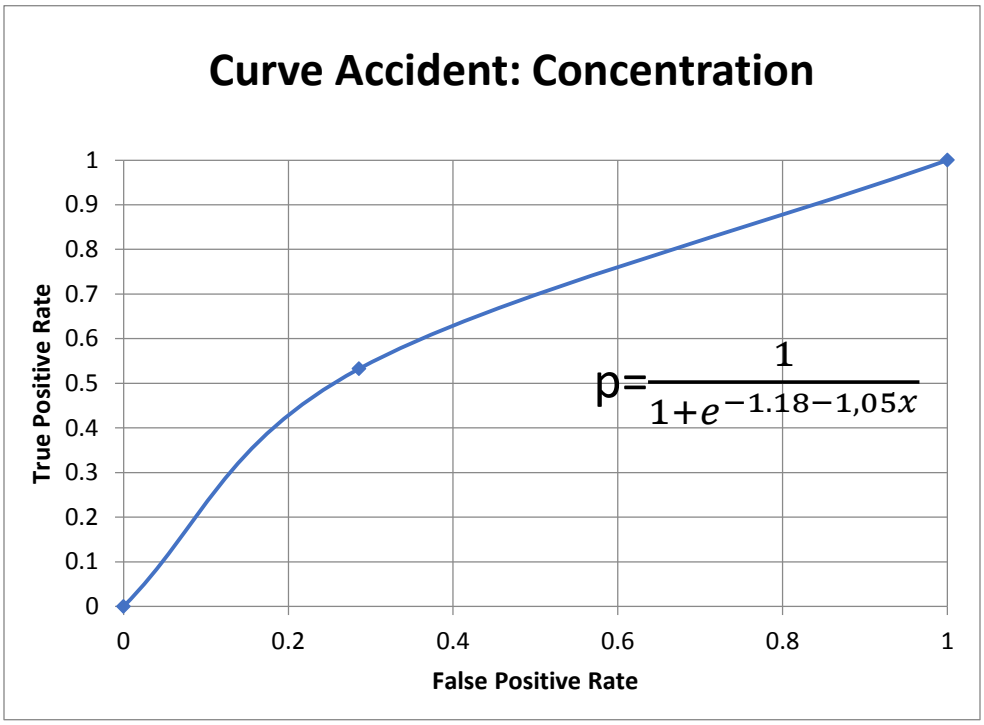

Figure 14. Area under the curve for the model obtained

Models built with a neural network

In this type of models, which function as a black box type, specifically for this doctoral research, a model of 7 input neurons, 2 hidden layers of 20 neurons each and an output of 2 neurons, was elaborated, which was predicted with one 98.2% precision, as presented below in figures 15, 16 and 17, highlighting that the most important variables are again the average concentrations.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

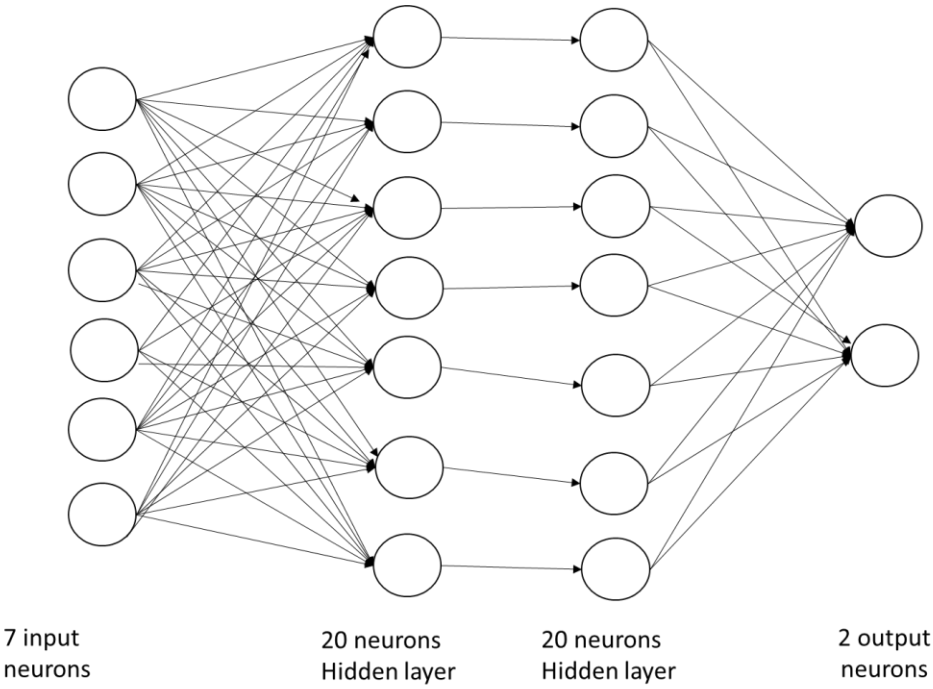

Figure 15 Multilayer neural network

		0	1	Error	Rate
0	0	25.0	3.0	0.1071	(3.0/28.0)
1	1	0.0	139.0	0.0	(0.0/139.0)
2	Total	25.0	142.0	0.018	(3.0/167.0)

Figure 16 :Confussion matrix

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

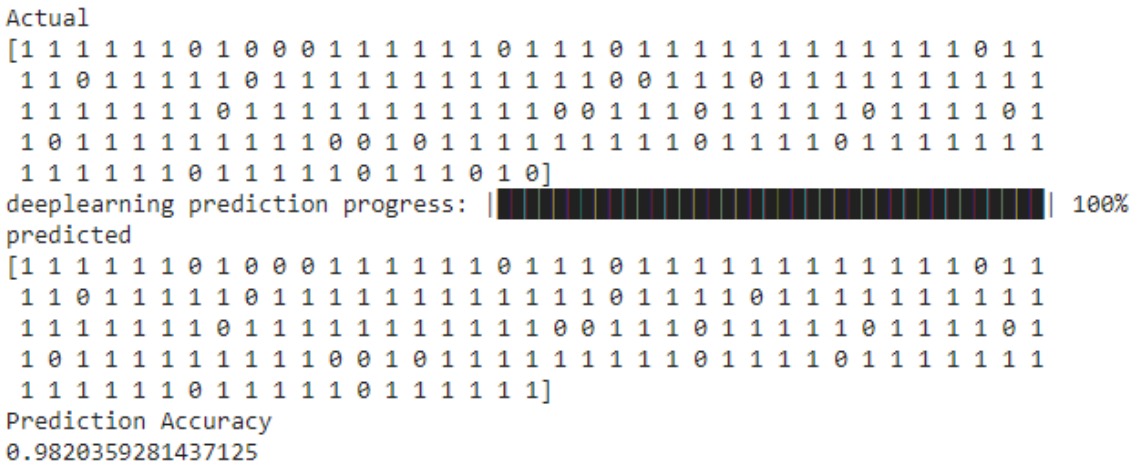

Figure 17: Current number of accidents versus predicted number of accidents

Conclusions

Deconcentration contributes significantly to the prediction of the probability of a traffic accident for drivers, because it was observed that the cumulative effect of the text messages that generated deconcentration affects the generation of accidents. The models developed show that they are influenced by the most important variable, which is concentration. A predictive model based on Machine Learning was successfully implemented. The concentration level, which is one of the most difficult to measure in this type of experiment, was successfully measured. It was observed that when considering the simple behavior model, the only variables that influence are the age, the degree of concentration of the drivers, obtaining details of the model from 83 to 100%. 70.1% of the total male participants were concentrated and 26.9% deconcentrated, the female gender, 58.9% of the participants were concentrated and 42.04% deconcentrated during the execution of the experiment. The factor that is not considered in the study of behavior and that is the most important and most difficult to study is the level of concentration, which has been measured and analyzed in this doctoral research. The best model obtained is a decision tree of 20 nodes with a precision of prediction (accurracy) of 100% and whose generalization is 90% and 91% measured with x5 and x10 cross validations, in this model the two most important variables are the average concentration.

Risky behaviors of drivers when subjected to a distraction factor were classified and characterized, regardless of gender or age, and that in some way influenced the behavior at the time of driving the recreated scenario.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

It was observed that the drivers who participated in the experiment driving a vehicle with mechanical transmission presented greater difficulties and, therefore, were more easily distracted when receiving *WhatsApp* messages on their cell phones, making more recurrent mistakes than drivers who preferred automatic transmission. It was observed that people over 50 years old tend not to check the cell phone while driving, while people in the range of 18 to 45 years old, checked the cell phone and on several occasions, answered messages while driving, affecting the degree concentration at the time of the test and presenting more recurrent driving errors and crashes.

The mathematical model obtained depends on the degree of concentration of the drivers and varies, depending on the scenario, weather conditions, type of road, time of driving, type of vehicle and type of vehicle transmission.

Acknowledgment

The authors are grateful to the National University of Colombia for support and Colciencias for providing funding for this research

References

AASHTO. 2004. A Policy on Geometric Design of Highways and Streets. American Association of State Highway and Transportation Officials. Washington, D.C., USA.

Af Wåhlberg, A.E., 2012. Changes in driver acceleration behavior over time: Do drivers learn from collisions? *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(5), pp.471–479.

Alonso G. et al (2004), Entrenamiento de una red neuronal artificial usando el algoritmo simulated annealing, Revista Scientia et Technica. Año X, No 24, Mayo 2004. UTP. ISSN 0122-1701 .

Alicea, L. 2004. Analysis and Evaluation of Crashes Involving Pedestrians in Puerto Rico. Tesis de Maestría en Ingeniería Civil, Recinto Universitario de Mayagüez, Universidad de Puerto Rico.

Alonso, M., 2016. La integración del factor humano en el ámbito técnico de la gestión de las carreteras y la seguridad vial: Un enfoque investigativo. Available at: http://roderic.uv.es/handle/10550/51943.

Arias, W., Colucci, B., 2006. Road Safety Audit. , 19(3), p.28.

Benekohal, R.F., Hashmi, A.M. 1992, Procedures for estimating accident reductions on two-lane highways, Journal of Transportation Engineering, 118 (1), pp. 111-129.

Engelbrecht, A.P., 2007. Computational Intelligence. An Introduction, second ed. Wiley, NY. Evans, L., 2004. Traffic safety. Bloomfield Hills, MI: Science Serving Society.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

Ferrer, A., Smith, R. & Cuellar, M., 2013. Análisis de la Capacidad de Gestión de la Seguridad Vial. *Banco Mundial*, pp.92–93.

Figueroa, A., 2005. Speed factors on four lane highways in free flow conditions *Doctoral thesis*. Purdue University

Figueroa, A., Kong, S., A. Tarko. 2005. Roadway and Driver Factors of Risk Perception on Four-Lane Highways, International: Road Safety on Four Continents, Warsaw, Poland.

Figueroa, A., Colucci, B. Arias, W. 2006, Sistema de Gerencia en Seguridad Vial: Integrando la Planificación, el Diseño Geométrico y la Auditoria de las Carreteras Primera cumbre puertorriqueña de seguridad vial, San Juan, , Puerto Rico, USA, 2006.

Gibson, J.J., 1986. The ecological approach to visual perception. Hillsdale, NJ: Lawrence Erlbaum.

Glymour, C., Madigan, D., Pregibon, D., Smyth, P., 1997. Statistical themes and lessons for data mining. Data Mining and Knowledge Discovery 1 (1), 11–28.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Macmillan, NY.

Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J., Brown, J., 2006. Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. Accident Analysis and Prevention, n.38, pp.185-191...

Jamson, A.H., 2001. Image characteristics and their effect on driving simulator validity. Proceedings of the first international driving symposium on human factors in driver assessment, training and vehicle design (pp.190-195). Aspen, CO.

Karlaftis, M.G, Vlahogianni E.I., 2010 Statistical methods versus neural networks in transportation research, Transportation Research part C,

Kesting, A., Trieber, M. Helbing, D. (2009). Agents for traffic Simulation. in Uhrmacher A. & Weyns, D. (Ed.), Multi-agent systems. Simulation and applications (pp. 325-356), Crc Press.

MATLAB, User's Manual, Mathworks Inc, 2017

National Center for Statistics and Analysis (NCSA). 2006a. Race and Ethnicity in Fatal Motor Vehicle Traffic Crashes 1999-2004. National Highway Traffic Safety Administration, Washington, D.C.

National Center for Statistics and Analysis (NCSA). 2006. Puerto Rico Toll of Motor Vehicle Crashes, 2005. National Highway Traffic Safety Administration, Washington, D.C.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:10:2021
DOI: 10.17605/OSF.IO/C4KY2

National Center for Statistics and Analysis (NCSA). 2005. Traffic Safety Facts 2004. National Highway Traffic Safety Administration, Washington, D.C.

Pollatsek, A., Fisher, D.L., Pradhan, A.K., 2006. Identifying and remediating failures of selective attention in younger drivers. Current directions in Psychological Science, 15, 255-259.

Rumar, K. 1985. The Role of Perceptual and Cognitive Filters in Observed Behavior. En: Human Behavior in Traffic Safety. Evans and Schwing, Plenum Press.

Sadek, A.W., spring, G., Smith, B.L., 2003. Toward more effective transportation applications of computational intelligence paradigms. Transportation Research Record 1836, 57–63

Vaa, T. 1997, Increased police enforcement> effects on speed, Accident Analysis and Prevention. 29(3) pp. 373-85.