❐   756

Measuring driver cognitive distraction through lips and eyebrows

Afizan Azman[1], Mohd. Fikri Azli Abdullah[2], Sumendra Yogarayan[2], Siti Fatimah Abdul Razak[2], Hartini Azman[3], Kalaiarasi Sonai Muthu[2], Hani Suhaila[2]
[1]Research and Innovation Department, University College of Islam Melaka, Melaka, Malaysia
[2]Faculty of Information Science and Technology (FIST), Multimedia University, Melaka, Malaysia
[3]Faculty of Technology Management and Technopreneurship (FPTT), Universiti Technikal Malaysia, Melaka, Malaysia

Article Info

ABSTRACT

Cognitive distraction is one of the several contributory factors in road accidents. A number of cognitive distraction detection methods have been developed. One of the most popular methods is based on physiological measurement. Head orientation, gaze rotation, blinking and pupil diameter are among popular physiological parameters that are measured for driver cognitive distraction. In this paper, lips and eyebrows are studied. These new features on human facial expression are obvious and can be easily measured when a person is in cognitive distraction. There are several types of movement on lips and eyebrows that can be captured to indicate cognitive distraction. Correlation and classification techniques are used in this paper for performance measurement and comparison. Real time driving experiment was setup and faceAPI was installed in the car to capture driver's facial expression. Linear regression, support vector machine (SVM), static Bayesian network (SBN) and logistic regression (LR) are used in this study. Results showed that lips and eyebrows are strongly correlated and have a significant role in improving cognitive distraction detection. Dynamic Bayesian network (DBN) with different confidence of levels was also used in this study to classify whether a driver is distracted or not.

Corresponding Author:

Afizan Azman
Research and Innovation Department, University College of Islam Melaka
Batu 28, Kuala Sungai Baru 78200 Melaka, Malaysia
Email: afizan@kuim.edu.my

1. INTRODUCTION

World Health Organization (WHO) in 2009 has reported that approximately 1.3 million road users die every year on the world's roads and from 20 to 50 million were sustained with non-fatal injuries [1]. The report was the first broad assessment of the road safety situation done for 178 countries. In 2010 Great Britain Annual Report for Road Casualties reported that there were a total of 208,648 casualties in road accidents. These casualties were 6% lower than in 2009. Out of these, 1850 people were killed and 22,660 people were seriously injured [2]. Another key finding in this report was that 21% of fatalities were car occupants. Based on the number of casualties and fatalities reported in 2010, the UK government was estimated to spend around £ 15 billion in order to address this problem. Researchers all around UK have been encouraged to study causes and factors that caused road accidents. They have also been motivated to find solutions to reduce number of casualties and fatalities. Thus, active road safety research, has becomes very popular with many researchers looking for better prevention methods to reduce road accidents. Among contributory factors reported for road accidents, failed to look properly was the most frequently reported

contributory factor. Four out of five accidents reported under this factor indicated the reason to be driver or rider error or reaction. Failed to look properly factor is actually contributed by many factors. However, the most basic factor is from driver's or pedestrian's behavior. Many factors can cause a driver or a pedestrian to make an error and lead to functional failure whenever a decision has to be made. If a driver makes a mistake, it is usually because of both the external and internal conditions in which he was still performing his activity but did not enable him to cope with the demands appropriately [3]. Cognitive distraction is one such factor that can lead a driver into failure to look properly. It can reduce driver's capability to evaluate the current driving situation and notice sudden events and may lead to poor decisions when interacting with the road environment and with other road users [4]. Whenever a driver is required to make a quick decision while driving the car, he may accidently make a wrong decision and consequently take a wrong action. Cognitive distraction can make a driver enter a state calls looked but failed to see (LBFTS). A driver in this state might has his visual focuses on the road, but fails to see. Pedestrians, cyclists or other drivers, who are usually at the left and right sides, are at a high risk to get involved in an accident [5]. Therefore, it is important to find a way to detect this state and to prevent a driver from getting distracted cognitively. Previous study has showed that mouth and eye movements are highly correlated to each other when drivers were cognitively distracted [6]. These findings can be extended to looking specifically at new features such as lips and eyebrows. In this chapter, correlation between those two features were studied and presented in a linear regression model where their correlation coefficient was found to have a strong correlation. This strong correlation proved that lips and eyebrows are related to each other when a driver is cognitively distracted. A real time road experiment was conducted with 10 experienced drivers. Based on data collected four algorithms were investigated and compared in terms of average accuracy of distraction detection-support vector machine (SVM), static bayesian network (SBN) and logistic regression (LR) and dynamic bayesian network (DBN). The algorithms were also used to measure tp-rate, fp-rate, and precision, recall and f-measure or f-score which indicate amount or rate of classification on the driver's distraction.

2. PROPOSED METHOD

2.1. Looked but failed to see (LBFTS)

Driver's behaviors involve emotion, experience, attention, intention, vigilance and other factors involved during the driving activity. These behaviors are closely related to driving safety. As the surrounding world is far too rich for human perception capacities, hence in order to cope with this complexity, only relevant information is extracted by some specific filters in human beings [7]. Cognitive filter is the most important filter when it comes to a driving system. It impedes integration of information and the knowledge associated with them. Failing to do this might lead a driver into a state of LBFTS [8]. Research by the department for transport [9], has found that human perceptual errors were the predominant contributory factor where it contributed about 46% of all categories of contributory factor recorded. Within this perceptual factor, LBFTS was the third most frequently recorded contributory factor with 17% accounting from all other factors. Inattention was the highest contributory factor with 28% followed by misjudgment of others path or speed with 21%. There are two types of error that meet the criteria for LBFTS: i) when a driver searches the traffic environment over-selectivity for example looking for a large vehicles and overlooking the pedestrians or cyclist, ii) when a driver searches for features which distinguish hazardous from non-hazardous objects for instance location, orientation and speed but fail to integrate those features into a coherent danger image.

In this study, this scenario occurred several times. This state of mind made some of the drivers lost their attention and hence failed to recognize their surrounding and sometime even failed to respond to the question given during the experiment. After the experiment, the drivers were interviewed. Some of them mentioned that, they lost in focus on the surroundings or failed to respond appropriately to the asked questions since they were focusing on the shuttle bus moving in front of them or on the incoming one. They were struggling with the decision as to whether to overtake the bus or keep tailing behind it and some of them even failed to respond appropriately when stopping before the zebra crossing line or driving onto the road humps/bumps along the route due to the diversion of attention. The drivers were more concerned about the bus in front of them as they assumed it was the hazardous object for them and hence they tended to neglect their surroundings.

The situation above is well described in [10]. Cognitive processing monitors the environment to maintain awareness of other vehicles and makes decisions on when to change lanes based on the current mental state. There should be sufficient amount of attention needed for driving control, situation awareness and non-related driving tasks.

```
IF        my current goal is to check for a lane change
          & my current speed is less than the desired speed
          & there is a lead car blocking further acceleration
```

                 & there is enough room in the other lane

THEN   initiate a lane change for passing

If one or more of the condition in the IF statement above is not satisfied during lane changing the driver can be in an unsafe state. This could occur as the driver requires attention for more than one single task at that time and hence may ignore some of them. A theory was proposed about cognitive mechanisms that guide a deliberation process which involves making decisions under uncertainty [11]. Drivers with uncertainty usually require a lot of cognitive loading before a decision can be made. They need to integrate both information from the external environment and information from the individual's associative memory as determinants before a possible action to be taken and evaluated in a deliberation process.

2.2. Lips and eyebrows in cognitive distraction

A few papers have studied distractions based on lips and eyebrows. Rongben *et al.* [12] was found that lips movement can be used to detect a driver yawning or having a conversation while driving. A very significant finding can be found in this article that can be used to detect driver fatigue distraction. Eyebrows can also be used to determine whether a person is concentrating or focusing [13]. It was found that eyebrows can be useful to indicate human perception of focus. Nevertheless, the research did not take into account a driving environment. However, it was proved that perception of focus and eyebrows are correlated to each other.

When a person is cognitively distracted or thinking, lips and eyebrows are the two obvious moving features on a person's face and therefore these two features can be used to detect cognitive distraction [14]. Figure 1 as shown in was taken in a car where the participant was asked to give a normal expression and thinking expression (cognitive distracted) [15]. It can easily distinguish between non-distracted facial expressions with distracted facial expressions [16].

Figure 1. Non-distracted face and distracted face

People in cognitive distraction, often furrow or perplex their brows when concentrating and also purse their lips when conducting mental searches [17]. The thinking facial expression is usually related with the eyes looking up away or down at the ground and it usually combined with the use of biting the lower lip. A person with cognitive distraction usually will have behaviors as listed in Table 1.

Table 1. Lips and eyebrow behaviors when a person in cognitive distraction [18]

Lips	Eyebrows
Moving lips- Lips moving in the shape of words but without making sounds mean the person is thinking. Rolling in the lips sign of uncertainty and accompanying with lowered eyebrows.	Furrowed brow- Concentration may also be shown in the forehead as the eyebrows are brought together as the listener seeks to hear and understand the other person.
Protruding lips- Both lips pressed together and pushed out generally indicate doubt. If the finger touches them, it may indicate internal thinking.	Middle-lowered eyebrow- When the middle of the eyebrows is pulled down so they slope inwards, this shows that the person is intense concentration.

Matsumoto and Ekman [17] mentioned that facial musculature is fairly unique. It includes somatic muscles in the body attached on one side to bone and the other to skin and make facial movements to be specialized for expression. Face is one of the body places where some muscles are not attached to any bone at

all. For instance orbicularis oculi muscle is a muscle surrounding the eyes and orbicularis oris is the muscle in the lips. Thus, there is no one to one correspondence between structure and function in some facial muscles, meaning that, facial expression is only possible with a group of facial movements. In this study it has been proved that lips and eyebrows can be grouped together for a thinking expression. In our experiment, a toolkit called faceAPI [18] was used to capture movements on eyebrows and lips. Both right and left eyebrows have three different muscles representing three different face landmarks (FLM) [15] as shown in Figure 2: 300, 301, 302 for right eyebrows (presented by circles in red) and 400, 401, 402 for left eyebrows (presented by circles in black). Lips can be categorized into two groups of FLM. Outer lip contour (presented by circles in brown): 100, 101, 102, 103, 104 and 105. Inner lip contour (presented by circles in pink): 200, 201, 202, 203, 204, 205, 206 and 207. Each group represents as one node or feature for data classification later. Thus, the model from this study will have 7 nodes with groups of face landmarks. Those 7 nodes are: i) right eyebrow, ii) left eyebrow, iii) inner lips, iv) outer lips, v) head position, and vi) head rotation and distraction as the last node.
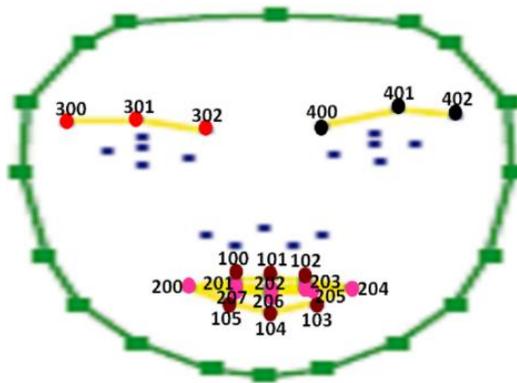

Figure 2. Face landmarks (FLM)

3. RESEARCH METHOD

In the experiment participants were required to drive the car in Loughborough University, from Holywell Park at one side of the campus to the Haslegrave Building at the center of the campus and turn back to Holywell Park as shown in Figure 3. The route is around 3.1 miles and one lane per direction. It consists of 9 humps, 7 zebra crossing and 4 shuttle bus stops.

Figure 3. Experiment's route

During the experiments, oncoming traffic was present, however the overall traffic density was moderate. Drivers were driving during daytime under various weather conditions. The experiments were

conducted during early spring and the temperatures were around 5 to 10 degree celsius. Contextual information like weather, temperature, traffic condition and road were considered in this study. Every participant who was involved in the experiment had a similar period of time and road condition. The experiment was run during day time from 9 am to 3 pm. The weather during the experiment was also consistent where the temperature was ranged between 5 to 10 degrees celsius. Traffic condition during the experiment was also not very heavy. Therefore, the contextual information among participant is considered similar. 10 experienced drivers were involved in this experiment. All participants are male. Their ages are between 21 years old to 36 years old. On average they have been driving for at least 4 years. Each driver participated on one run without distraction and another one run with distraction. Overall, 10 runs were made with the driver being attentive and 12 runs while the driver were distracted. 2 runs were wrong runs because the drivers were mistakenly took a wrong route. The wrong runs affected the experiments because the situational awareness questions have to be changed. For instance, the questions about shuttle bus stop, humps, signboard and zebra crossing were required to be changed due to the wrong route taken by the drivers. Therefore those two runs were eliminated. Non-distracted experiment lasted around 13 to 18 minutes for each run and the distracted experiment lasted around 24 to 32 minutes depending on the drivers' speed. The average speed for each driver was around 12 to 20 miles per hour. Speed limit on campus is 15 miles per hour. The car used in this experiment was an automatic transmission car.

Each driver was required to drive along the route twice. First round was the control experiment where the driver was not distracted with any type of distraction. In this first round, the driver was asked to familiarize with the car and the route. This round has been labeled as control/baseline/non-distracted experiment. Second round was the task experiment where the driver was distracted with both manual and visual distractions. As in [18], driving is a multitasking activity that requires the drivers to divide their attention with various driving and non-driving related tasks. Thus, to operate this multitasking activity, the drivers need to be experienced, well in emotion, motivated, vigilance and highly attentive while driving as in Figure 1. This second round was also denoted as distracted/task experiment. After the drivers familiarized themselves with the car and route showed to them, they were asked to drive down the same route while performing some additional tasks in addition to driving. Ten typical tasks were given as distraction conditions and these tasks were applied to all drivers: i) identify pedestrians, ii) recognize the road signboard, banner and poster, iii) identify road speed limit and respond accordingly, iv) converse with the experimenter, v) using mobile phone (iPhone) to find words, contact numbers or used the phone applications like weather, calculator or messages, vi) recognize model and registration number plate from the car in-front. Read from a pamphlet, vii) find contact number and price on the pamphlet. answer some simple arithmetic questions, viii) collect a bottle on the floor of the car and pass it to the experimenter, and ix) adjust the radio settings.

Cognitive distraction was triggered from the above tasks. The car was equipped with the face tracking and feature extraction faceAPI toolkit. A normal web camera was used and it was attached to the dashboard of the car as in Figure 4. Since the drivers were allowed to adjust the seats accordingly to their preferences, the web camera was also adjusted to ensure that it can capture the full face of a driver. The experiment setup for this study was shown in Figure 4(a) and Figure 4(b). The camera was mounted on the dashboard and the laptop for collecting the data was set at the passengers' seat. Figure 4(c) and Figure 4(d) showed the participant involved in the experiment. Instructions were read to the participant and every movement in the vehicle was recorded. For each driver, as shown in Figure 1, a snapshot picture was taken for every round of driving. These pictures can be used to demonstrate the differences between facial expression during distracted and not distracted. From these pictures, obvious movements can be found on the drivers' facial features in distracted picture in comparison with the not distracted picture. In this paper, the proposed approach detected facial features and further distinguished the driver' status: whether he/she is distracted.

| (a) | (b) | (c) | (d) |

Figure 4. Experiment setup: (a) faceAPI toolkit setup, (b) faceAPI initializing, (c) driver is driving, and (d) after the experiment

4. RESULTS AND DISCUSSION

This section discussed about the results received from the experiment. Several algorithms are used to analyze the data from the experiment by computing regression, correlation and accuracy rate.

4.1. Linear regression model and correlation

Correlation between lips and eyebrows were analyzed with linear regression method. Linear regression is an approach to modelling the relationship between a scalar variable Y and one or more explanatory variables denoted X. The equation for linear regression is [19]:

$$Y = mX + b$$

where m is the slope and b is the intercept.

A driver's distraction or non-distraction was the scalar variable whereas lips and eyebrows were the explanatory variables. Each driver had an average of 1380 seconds total of data frames which were divided into two groups: distracted and non-distracted. In our previous study, correlation between height of eye movement and mouth movement and width of eye movement and mouth movement were analyzed with pearson-r correlation [15]. On an average, the correlation r value between those features was around 0.7635 and gave a very strong relationship between them.

In this study, linear regression model was used for two reasons. First, it was used to obtain the correlation coefficient between type of experiments (distracted or non-distracted) and the features (lips and eyebrows). Thus, it can be used to analyze how a lip data varies with the eyebrows data when drivers were distracted and not distracted. Secondly it was used to predict the value of lips from eyebrows or value of eyebrows from lips. For example it can be used to predict the movement of lips when the eyebrows move at certain value. Thus, if the movement on lips is increasing, the movement on eyebrows will also increase. Significantly, lips and eyebrows are proportionally moving to each other.

For the first objective, each driver set of data was trained with linear regression model and the correlation coefficient (r-value) was taken. On an average, during the distracted experiment, the correlation coefficient between distraction and the selected features was 0.70598 and during the non-distracted experiment it was only 0.30527 as shown in Table 2. The difference between distracted and non-distracted was around 40.071% which is a big difference to distinguish whether a driver is distracted or not. Again, 0.70598 indicated that lips and eyebrows have a strong positive linear relationship with cognitive distraction via a firm linear rule.

Secondly, the linear regression model also exhibited a linear regression equation between lips and eyebrows. This equation is useful to predict the value for lips from eyebrows or vice versa. Since the second objective was to find the best line that fits both features (lips and eyebrows), the data in the distracted experiment was combined all together and to gave a total of 6900 data frames (690 frames x 10 drivers). If lip is X variable and eyebrow is Y variables, the best line fits both variables is:

$$Eyebrow = (0.9002 \times Lips) + 0.5109$$

the line is a positive line with a slope equal to 0.9002. Correlation coefficient for this second purpose is 0.7162 which again depicts a strong correlation between those two features.

Table 2. Linear regression model

Brows and Lips	Distracted-2				No Distracted-1			
	Correlation Coefficient	Mean Absolute Error	Root Mean Squared Error	Total Number of Instances	Correlation Coefficient	Mean Absolute Error	Root Mean Squared Error	Total Number of Instances
Driver1	0.6492	0.1829	0.2537	690	0.4381	0.0254	0.0682	690
Driver2	0.6904	0.1654	0.2448	690	0.5832	0.0283	0.0617	690
Driver3	0.7394	0.185	0.2609	690	0.2293	0.0151	0.064	690
Driver4	0.5928	0.2224	0.3092	690	0.155	0.0158	0.075	690
Driver5	0.6344	0.2332	0.2977	690	0.3306	0.0223	0.0716	690
Driver6	0.7113	0.2453	0.3145	690	0.2102	0.0175	0.0742	690
Driver7	0.8251	0.1972	0.2666	690	0.45	0.0256	0.0757	690
Driver8	0.6546	0.317	0.3754	690	0.2133	0.0182	0.0742	690
Driver9	0.7943	0.1443	0.2148	690	0.2262	0.018	0.074	690
Driver10	0.7683	0.154	0.2244	690	0.2168	0.0172	0.0741	690
AVERAGE	0.70598	0.20467	0.2762	690	0.30527	0.02034	0.07127	690

4.2. Distraction detection using different algorithms

All features captured by faceAPI including head position, head rotation, lips and eyebrows were used. Scatter diagram was also used for correlation. However, for the scatter diagram, only lips and eyebrows under distracted are used. Data was sub-divided into two groups. The first group of data consisted of lips, eyebrows, head rotation and head position, with lips and eyebrows. The second group of data contained only head position and head rotation, i.e. without lips and eyebrows.

Three different types of classification algorithms were used: SBN, SVM, and LR, for accuracy or successful rate comparison. Later in this section, these three algorithms will be compared with DBN. Classification is the process of generalizing known structure to apply to new data. Classification accuracy is the number of instances which are correctly classified. It is a ration of the successfully classified data divides with the total number of the data.

Figure 5 as shown there is a significant difference between the group with lips and eyebrows and the one without lips and eyebrows in every classification algorithms. SVM was the best classification method in these three algorithms to classify whether the drivers were cognitively distracted. With SVM, data with lips and eyebrows was 79.58% correctly classified and without lips and eyebrows was 71.19% of the data was correctly classified. Thus, it clear that data pertaining to lips and eyebrows has higher classification accuracy (of the order of 8.39%) as compared to the data set without lips and eyebrows. Logistic regression had an accuracy of 75.85% with lips and eyebrows and 68.41% without lips and eyebrows. It can be seen that, LR is less accurate. Finally, SBN accuracy was 77.57% with lips and eyebrows and 69.99% without lips and eyebrows.
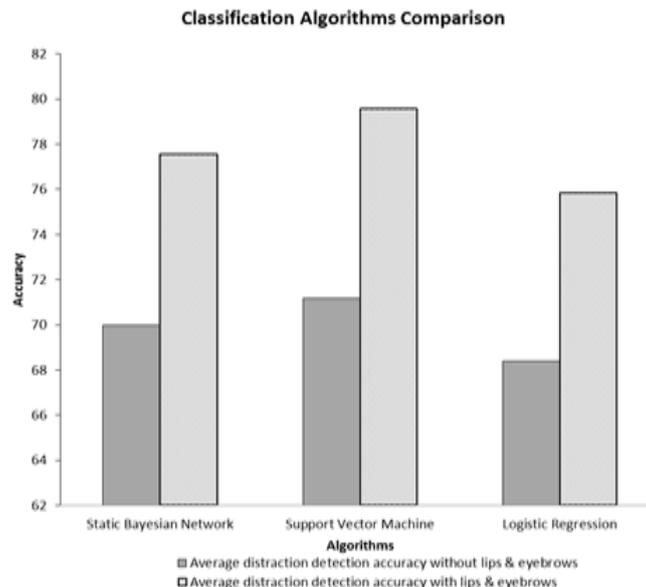

Figure 5. Accuracy of distraction detection comparison

The comparison between data with lips and eyebrows and data without lips and eyebrows was made from distracted and non-distracted groups. The data set consists of 50% from distracted group and 50% from non-distracted group. The training was conducted with 2/3 of the data and the testing was tested with rest of the data.

Data with lips and eyebrows was also analyzed for its true positive rate (TP), false positive rate (FP), f-score (F-measure), precision and recall. TP rate calculates the true positive rate with respect to a particular class and this is defined as correctly classified positives divided by the total number of positives [20]. FP rate for a particular class is computed as incorrectly classified negatives divided by the total number of negatives. FP rate can also be calculated as 1-TP. Precision is calculated as correctly classified positives divided by the total values predicted as positive. Recall is also equal to TP rate value and its calculation is the same as TP rate. Finally, F-measure or f-score is calculated as:

$$f - measure = 2 \left[\frac{recall \times precision}{recall + precision} \right]$$

f-measure weighted average of the precision and recall values from this cognitive distraction detection. The f-measure maximum value is 1 and minimum value when it is 0.

From Figure 6 it can be seen that, SVM's TP rate, F-measure, precision and recall values are the highest among those three whereas its FP rate is the lowest. Basically, precision value in cognitive distraction tells the proportion of the cognitive distraction which truly have classified over the total classified cognitive distraction. Recall was also sometime referred as TP rate is a proportion classified as cognitively distracted over the actual total in cognitive distraction. Therefore, F-measure is the weighted average from both precision and recall in this study [21].
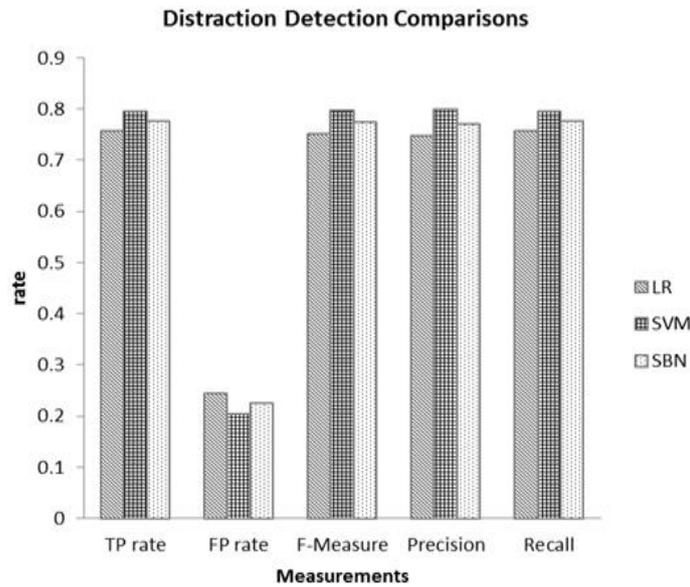

Figure 6. Distraction detection comparisons of LR, SVM and SBN

4.3. DBN for driver cognitive distraction

DBN is used to describe how variables influence each other over time based on the model derived from past data. A DBN contains a pair of information that models the temporal process by specifying a probability distribution for:

$$X^0, \ldots, X^T; P(X^0, \ldots, X^T | G, \Theta)$$

where

$$X^t = \{X_1^t, \ldots, \quad X_1^t\}$$

is a set of discrete random variables that represents the state of a temporal process at a discrete time point t [22]. G is a directed acyclic graph (DAG) with the nodes corresponding to the random variables in X^0 and X^1 X^1. Θ is a set of parameters specifying a conditional probability distribution for each node X_1^t in G given its parents $Pa(X_1^t)$ in G,

$$P(X_1^t | Pa(X_1^t), G, \Theta)$$

DBNs can also be defined by a pair of where is a prior knowledge or network and is a transition network. In DBN, joint distribution is important, given two random variables X and Y defined on the same probability space, the joint distribution for X and Y defines the probability of events related to both X and Y. For a given DBN model the joint distribution for X over X [0], X [1], …, X [T] is:

$$P_{DBN}(x[0], x[1], x[2], \ldots x[T]) = P_{B_0}(X[0]) \prod_{t=0}^{T-1} P_{B_\rightarrow}\big((x[t+1]) | x[t]\big)$$

where $P_{B_\to}$ is a transition model. DBN in this paper was developed in MATLAB by using a Bayesian network toolbox in [23]. As stated previously, 7 groups of face landmarks are the nodes in the DBN model like in the Figure 7.
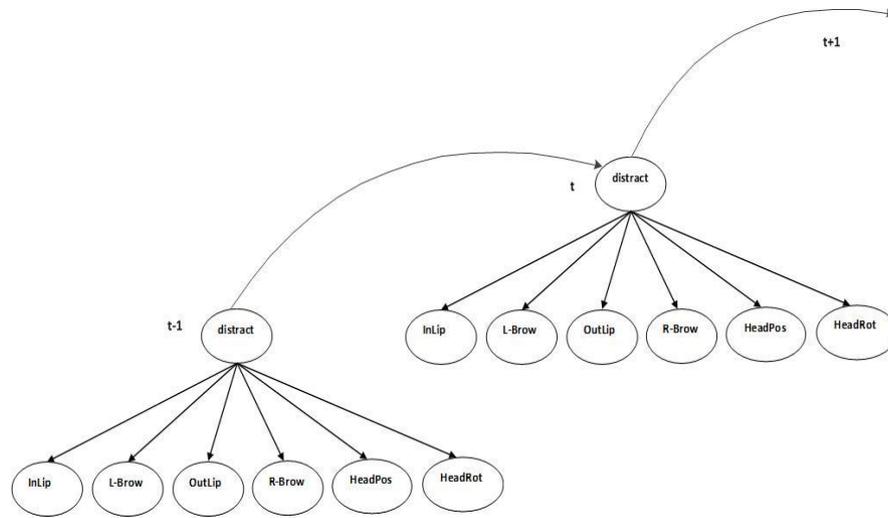
Figure 7. DBN model with face and head parameters

The DBN model was created with three different confidence levels: 50%, 70% and 90%. Data from all the ten drivers was used to check the accuracy of cognitive distraction detection. Features used to develop this DBN model include inner and outer lips, right and left eyebrows, head position and head rotation which are similar to SVM, SBN and LR features.

On average, with 50% of confidence level, DBN detected the driver's cognitive distraction with an accuracy of 84.82%. At a confidence level of 70%, the accuracy of detection is only 74.80%. At 90%, the accuracy for cognitive distraction with DBN was even lesser, at only 59.59%.

Confidence level is a statement about likelihood that the interval obtained actually has the true parameter value. For an instance, assume there is a true, fixed value of θ. Given by the assumption, data can be used to get an estimation of θ ( ). Once the estimated value has been assessed, the location where the true value is in relation to the estimation can be found.

In this ayesian approach, assume that the true value is a random variable. Thus, the uncertainty about the true parameter value is imposed by a prior distribution on the true parameter vector. By constructing the posterior distribution for the parameter vector, the prior distribution and the data can be blended together. To know the extent of uncertainty at point of estimation, the interval can be constructed as:

$$P(lower\ bound\ (\theta)\ \leq\ \theta\ \leq\ upper\ bound\ (\theta))\ =\ 0.95$$

The value of 0.95 is changeable depending on the interval required for the true parameter value located. Almost every bayesian network has a confidence level at ≥0.5. Most of the time, the confidence level in bayesian network is depending on the distribution of the data values [24]. In this study data values are mostly distributed around or above 0.5. Standard deviation between drivers on the accuracy was also measured. This standard deviation showed a difference between accuracy of detection among drivers.

As shown in Figure 8, at confidence level of 70%, the standard deviation on the accuracy rate between drivers is the lowest, at 3.62. However at 90% confidence level, the difference of accuracy rate between the drivers is the highest with 8.49. As can be seen in Table 3, accuracy rates between drivers in CL 50% and CL 90%, the differences of each rate are bigger compared to the accuracy rate in CL 70%. Clearly in the table, range for accuracy rate in CL 70% for each driver is in between 70 to 84. However, range for CL 50% is in between 72 to 93 and range for CL 90% is in between 50 to 72. Since the range gap in CL 70% is the smallest, thus, the standard deviation for CL 70% is the lowest among these three confidence levels.

DBN was then compared with the other three types of classification algorithms used earlier: SVM, SBN and LR. Same features were again employed in the DBN model: inner lips, outer lips, right eyebrow, left eyebrow, head position and head rotation. Previously, SVM was the best performance algorithm with the

most correctly classified driver cognitive distraction. SBN being the next and LR last. However, with DBN, the accuracy rate or correctly classified data was closer to 85% when the DBN model was built with a 50% confidence level and it is the only classification algorithm that gave an accuracy rate above 80%. DBN with confidence level of 50% is more important compared to 70% or 90% confidence level because, at 50% of confidence level, many true data can be classified as true at which the accuracy rate reached 84.82. These results are presented as shown in the Figure 9.

Table 3. DBN distraction detection accuracy

Accuracy of Cognitive Distraction Detection		Confidence Level (CL)		
		50%	70%	90%
Drivers	Driver 1	72.5362	70.6522	50.2899
	Driver 2	75.942	72.6812	50.5072
	Driver 3	78.0435	72.8986	50.7246
	Driver 4	78.4058	73.0435	52.7536
	Driver 5	80.2174	73.2609	54.7826
	Driver 6	91.1594	73.6232	62.5362
	Driver 7	92.4638	74.9275	63.5507
	Driver 8	92.5362	75.4348	65.2174
	Driver 9	93.2609	76.9565	72.6812
	Driver 10	93.6232	84.4928	72.8986

Std Dev on DBN's Accuracy Rate for Different CL

▨ CL 50%   ▤ CL 70%   ▥ CL 90%

CL 90% — 8.49
CL 70% — 3.62
CL 50% — 8.03
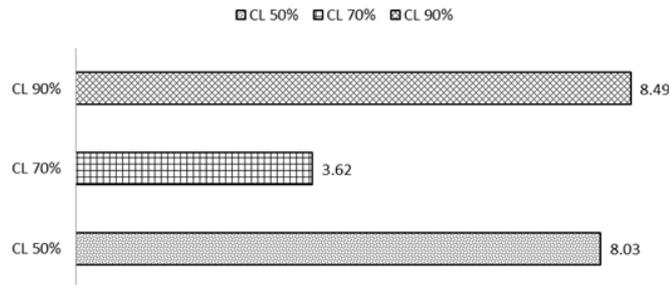
Figure 8. Standard deviation at different confidence level (CL)

Accuracy Comparison from Every Classification Algorithms

▨ DBN 50% ▦ DBN 70% ☐ DBN 90% ▥ SVM ☐ SBN ▥ Logistic Regression

Logistic Regression — 75.85
SBN — 77.57
SVM — 79.58
DBN 90% — 59.59
DBN 70% — 74.8
DBN 50% — 84.82
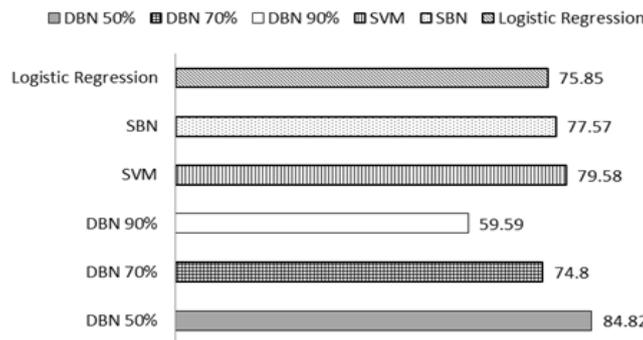
Figure 9. Comparison on average of accuracy rate for every algorithm used in this study

Features from DBN model earlier were basically captured by faceAPI and they are more physical and their movements can be captured normal web camera. However, the features in the extended version model are only able be captured with infrared camera. The additional features captured by faceLAB machine and they are gaze rotation (left and right eyes) and blinking frequency. Those features are proposed in [24] and [25]. The extended model contained 9 nodes: distraction, blinking frequency, right eye gaze rotation, left eye gaze rotation, head rotation, outer lip, inner lip, right eyebrow and left eyebrow as in Figure 10. From this

new DBN model, head position has been removed because its information was not really useful to detect cognitive distraction. There are a few studies were also used head rotation for their detection [24], [25]. Table 4 as shown in give a full description on each feature used in the experiment.

The vehicle setup to capture the final DBN model was set just like in Figure 11. The route mapping for this experiment was just like in Figure 3. Results in this final experiment were compared. Two groups of features are received as in Table 5. First group was imitating the features taken as in the previous experiment where only lips, eyebrows and head are taken. Second group contained all the features captured from the combination of the machines: lips, eyebrows, head, blinking and gaze.
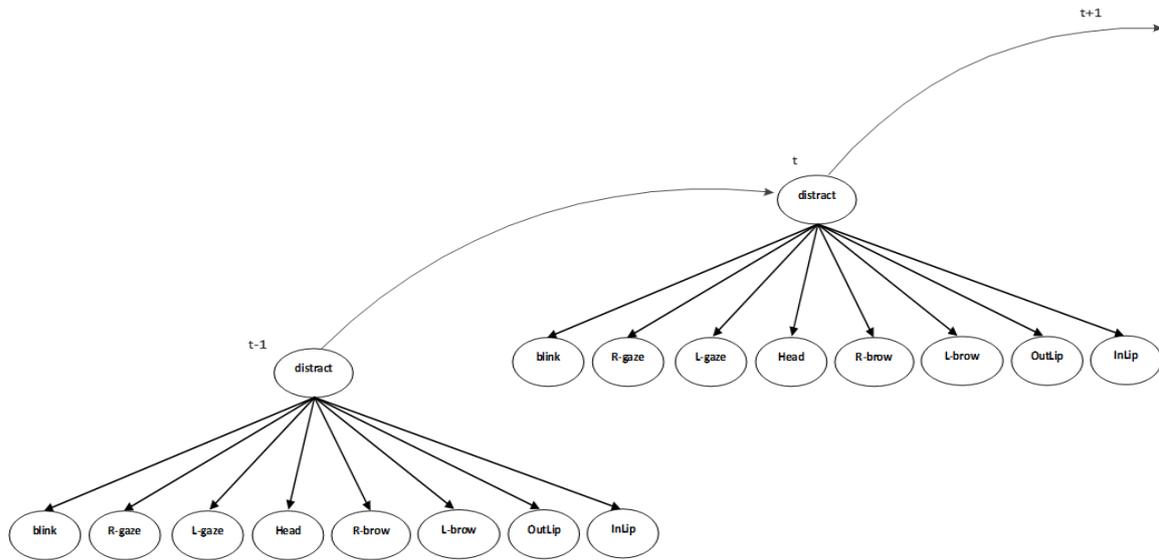

Figure 10. Extended DBN model (final experiment)

Table 4. DBN model's (final experiment) notation and description

Notation	Description
t	Time
distract	Distraction
blink	Blinking frequency
R-gaze	Right eye gaze rotation
L-gaze	Left eye gaze rotation
Head	Head rotation
R-Brow	Right eyebrow
L-Brow	Left eyebrow
InLip	Inner lip
OutLip	Outer lip

Figure 11. Vehicle setup for final experiment

Table 5. DBN model accuracy for final experiment (1st and 2nd groups)

| | 1st Group | | | 2nd Group | | |
| | Lips, Eyebrows, and Head | | | Lips, Eyebrows, Head, Gaze, and Blinking | | |
	CL 50%	CL 70%	CL 90%	CL 50%	CL 70%	CL 90%
Driver 1	75.6412	67.1457	54.3112	81.7412	73.3466	65.4112
Driver 2	75.911	68.3241	55.556	82.6173	73.355	65.4221
Driver 3	76.4345	68.766	56.0472	82.9721	73.8667	65.6385
Driver 4	77.8821	69.043	56.4325	83.334	74.5812	66.5415
Driver 5	78.315	69.5619	56.8671	83.8577	75.0113	67.1443
Driver 6	79.9246	70.2367	58.7265	84.6413	75.7222	69.166
Driver 7	80.0613	71.5433	59.3557	84.8666	76.443	69.4312
Driver 8	80.7556	72.7778	59.6781	85.514	76.512	69.8556
Driver 9	80.9861	72.8125	60.0024	86.6678	77.0001	69.8913
Driver 10	83.0334	72.9766	60.3225	87.011	77.8999	69.8934
Driver 11	83.5446	73.334	61.4804	89.4778	78.2454	69.9255
Driver 12	84.222	74.6125	61.5146	89.899	78.6888	70.043
Driver 13	85.6813	74.8064	62.004	90.5728	79.0055	70.3691
Driver 14	85.799	75.0367	62.3581	90.8714	79.4567	70.432

Just like in previous experiment, three different confidence levels are used: 50%, 70% and 90%. From the first group results, on average, the 50% confidence level is the highest accuracy compares to the other confidence level. In 50% confidence level the average accuracy from each driver is 80.59%. In 70% confidence level the average accuracy is 71.49% and in 90% confidence level the average accuracy is 58.90% which is the lowest average accuracy compared to the other two levels. From the second group, in 50% of confidence level the average accuracy is 86% which is again the highest accuracy. In 70% confidence level the average accuracy is 76.37% and in 90% of confidence level the average accuracy is 68.51%. By comparing both groups, the average accuracy in 50% of confidence level is always the highest accuracy and the 90% of confidence level is always the lowest. From the results it can be found that, many data distributions are captured around 0.50 and above and a few are captured above 0.90. Results received above also showed that, when the features were fused with other promising features (blinking and gaze), the average accuracy is able to be enhanced. By comparing both groups in 50% of confidence level, the average accuracy for the first group is only 80.59% but when more features are combined, the average accuracy is boosted up to 86%. This is also happened in 70% and 90% of confidence level.

Basically the results from first and second groups can be presented as in shown the Figure 12. As can be seen, whenever features from first group combined with blinking and gaze rotation, the averages accuracy are getting higher for every confidence level. Next a comparison was made between those features with lips and eyebrows and without lips and eyebrows as shown in Table 6. Group A contained head, blinking and gaze whereas group B contained lips, eyebrows, head, blinking and gaze.

Again, on average of accuracy, with confidence level of 50%, the accuracy is the highest among the confidence levels. From group A, the average accuracy in 50% of confidence level is 81.3%. In 70% of confidence level the average accuracy is 74.5% and in 90% of confidence level the average accuracy is 65.71%. This shown that, with lips and eyebrows added in the cognitive distraction the average accuracy will get higher. With lips and eyebrows, in 50% of confidence level, the average accuracy is improved by 4.7%. In 70% of confidence level the improvement is 1.9% and in 90% of confidence level the accuracy has been improved by 2.8%. Comparison between group A and group B is as shown in Figure 13.
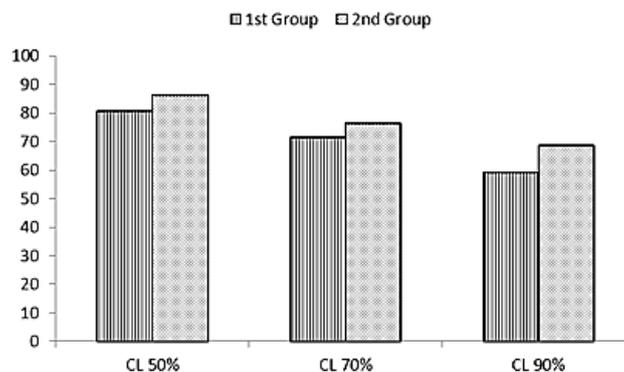
Figure 12. Average accuracy comparison between 1st group and 2nd group

Table 6. DBN model accuracy for final experiment (group A and group B)

| | Group A | | | Group B | | |
| | Head, Gaze, and Blinking | | | Lips, Eyebrows, Head, Gaze, and Blinking | | |
	CL 50%	CL 70%	CL 90%	CL 50%	CL 70%	CL 90%
Driver 1	77.552	69.7778	58.1042	81.7412	73.3466	65.4112
Driver 2	78.8311	69.9071	60.1889	82.6173	73.355	65.4221
Driver 3	79.2983	71.0419	61.4223	82.9721	73.8667	65.6385
Driver 4	79.6513	71.8756	62.7563	83.334	74.5812	66.5415
Driver 5	79.9031	72.4592	63.9	83.8577	75.0113	67.1443
Driver 6	80.055	73.53	65.0731	84.6413	75.7222	69.166
Driver 7	80.5777	74.8724	66.4892	84.8666	76.443	69.4312
Driver 8	80.9486	75.8251	66.889	85.514	76.512	69.8556
Driver 9	81.1844	75.98	67.2145	86.6678	77.0001	69.8913
Driver 10	82	76.5814	68.6667	87.011	77.8999	69.8934
Driver 11	83.7392	77.0716	68.9104	89.4778	78.2454	69.9255
Driver 12	84.5644	77.3667	69.5861	89.899	78.6888	70.043
Driver 13	84.7919	77.9641	70.0042	90.5728	79.0055	70.3691
Driver 14	85.1165	78.751	70.7183	90.8714	79.4567	70.432
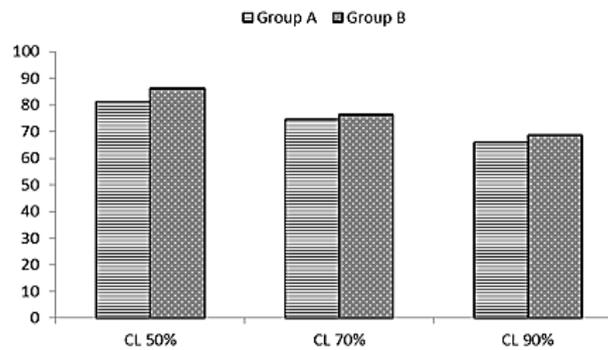
Figure 13. Average accuracy comparison between group A and group B

Based on the comparisons made, it is clearly showed that with lips and eyebrows added to blinking, gaze and head rotation can improve the performance of driver cognitive distraction detection. This improvement showed that lips and eyebrows are easily to be detected and also useful to determine whether a driver is cognitively distracted or not. Nevertheless, compared to SVM, SBN and LR, DBN is a better classification algorithm than other three with 50% of confidence level. Generally, Bayesian networks has a number of advantages over other type of classification algorithms:

a. Bayesian networks visually represent all the relationships between the features used to detect driver cognitive distraction. The relationships were shown with connecting arcs.
b. It is easy to recognize the dependence and independence between various nodes. For an instance, nodes distract as in Figure 10 are depending to the previous nodes distract in the system.
c. Bayesian networks can handle situations where the data set either from distracted or no distracted is missing or incomplete since the model accounts for dependencies between all features.
d. Bayesian networks can even maps driver cognitive distraction where it is not feasible or practical to measure all features used due to system constraints (costs, not enough sensors).
e. Help to model noisy systems.
f. Can be used for any system model - from all known parameters to no known parameters.

5.   CONCLUSION

DBNs provide a better and viable detection method for driver cognitive distraction. DBN models developed from this paper are created based on the real time experiment. Results comparison made between DBN and other traditional algorithms like SVM, logistic regression and static Bayesian network has showed that DBN are outperformed them in every confidence levels. Physiological features are captured from two different machines: faceAPI and faceLAB and both machines provide different type of features. faceAPI provides lips, eyebrows and head data whereas faceLAB captures head, blinking and gaze data. Data from captured from both machines are the fused and compared. Data are collected in a real car and in a real environment where up to 14 drivers are volunteered to participate. On road contextual information in a more

diverse set of conditions is considered. Major contribution from this paper is when lips and eyebrows are combined with head, gaze and rotation it improved the average accuracy for driver cognitive distraction detection. This showed that lips and eyebrows are not physically and easily to detect but they are also capable to improve the existing system in detecting driver cognitive distraction.

ACKNOWLEDGEMENTS

REFERENCES

[1]   "Global status report on road safety: Time for action," World Health Organization, Geneva, Switzeland, 2009. Accessed Feb, 2015. [Online]. Available: www.who.int/violence_injury_prevention/road_safety_status/2009

[2]   P. Kilbey, D. Wilson, O. Beg, G. Goodman, and Bhagat, "A reported road casualties in Great Britain: 2012 annual report," Department of Transport,United Kingdom, 2012 RRCGB, 2013.

[3]   J. J. Rolison and A. Feeney, "What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records," *Accident Analysis and Prevention*, vol. 115, pp. 11-24. 2018, doi: 10.1016/j.aap.2018.02.025.

[4]   H. Farah and D. Bankosegger, "Evaluation of the effect of cooperative infrastructure-to-vehicle systems on driver behavior," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 42-56, 2012, doi: 10.1016/j.trc.2011.08.006.

[5]   A. Azman, Q. Meng, and E. Edirisinghe, "Non intrusive physiological measurement for driver cognitive distraction detection: Eye and mouth movements," *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 2010, pp. V3-595-V3-599, doi: 10.1109/ICACTE.2010.5579547.

[6]   P. V. Elslande, C. Naing, and R. Engel, "Analyzing human factors in road accidents: TRACE WP5 summary report," Deliverable, 2008.

[7]   M. B. Herslund and N. O. Jørgensen, "Looked-but-failed-to-see-errors in traffic," *Accident Analysis and Prevention*, vol. 35, no. 6, pp. 885-891, 2003, doi: 10.1016/s0001-4575(02)00095-7.

[8]   K. Rumar, "The role of perceptual and cognitive filters in observed behavior," *In Human Behavior and Traffic Safety*, pp. 151-170, 1985, doi: 10.1007/978-1-4613-2173-6_8.

[9]   I. D. Brown, "A review of the'looked but failed to see'accident causation factor," In *Behavioural Research in Road Safety: Eleventh Seminar*, 2002.

[10]   D. D. Salvucci, "Modeling driver distraction from cognitive tasks," In *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, pp. 792-797, 2019, doi: 0.4324/9781315782379-171.

[11]   E. Stern, "Reactions to congestion under time pressure," *Transportation Research Part C: Emerging Technologies*, vol. 7, no. 2-3, pp. 75-90, 1999, doi: 10.1016/S0968-090X(99)00013-3.

[12]   W. Rongben, G. Lie, T. Bingliang, and J. Lisheng, "Monitoring mouth movement for driver fatigue or distraction with one camera," *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749)*, 2004, pp. 314-319, doi: 10.1109/ITSC.2004.1398917.

[13]   E. Krahmer, Z. Ruttkay, M. Swerts, and W. Wesselink, "Eyebrows and the perception of focus," *2002 International Conference of Speech Prosody*, 2002.

[14]   H. Kawanaka, M. Miyaji, Md. S. Bhuiyan, and K. Oguri, "Identification of cognitive distraction using physiological features for adaptive driving safety supporting system," *International Journal of Vehicular Technology*, vol. 2013, 2013, Art. no. 817179, doi: 10.1155/2013/817179.

[15]   A. Azman, Q. Meng and E. Edirisinghe, "Non intrusive physiological measurement for driver cognitive distraction detection: Eye and mouth movements*," 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE),* 2010, pp. V3-595-V3-599, doi: 10.1109/ICACTE.2010.5579547.

[16]   M. Iwasaki and Y. Noguchi, "Hiding true emotions: micro-expressions in eyes retrospectively concealed by mouth movements," *Scientific Reports*, vol. 6, 2016, Art. no. 22049, doi: 10.1038/srep22049.

[17]   D. Matsumoto and P. Ekman, "Facial expression analysis," *Scholarpedia*, vol. 3, no. 5, 2008, doi: 10.4249/scholarpedia.4237.

[18]   D. Straker, "Changing minds," *In detail*, Second Edition, Syque Press, 2010.

[19]   D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*, 5th Ed., John Wiley and Sons, 2012.

[20]   I. Russell and Z. Markov, "An introduction to the weka data mining system," In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 2017, pp. 742-742, doi: 10.1145/3017680.3017821.

[21]   Q. Ji, P. Lan, and C. Looney, "A probabilistic framework for modeling and real-time monitoring human fatigue," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 36, no. 5, pp. 862-875, Sep. 2006, doi: 10.1109/TSMCA.2005.855922.

[22]   K. Murphy, "Software packages for graphical models/bayesian networks," *International Society for Bayesian Analysis*, 2007.

[23]   L. Hespanhol, C. S. Vallio, L. M. Costa, and B. T. Saragiotto, "Understanding and interpreting confidence and credible intervals around effect estimates," *Brazilian Journal of Physical Therapy*, vol. 23, no. 4, pp. 290-301, 2019.

[24]   M. Miyaji, M. Danno, H. Kawanaka and K. Oguri, "Driver's cognitive distraction detection using AdaBoost on pattern recognition basis," *2008 IEEE International Conference on Vehicular Electronics and Safety*, 2008, pp. 51-56, doi: 10.1109/ICVES.2008.4640853.

[25]   M. H. Kutila, M. Jokela, T. Mäkinen, J. Viitanen, G. Markkuula, and T. W. Victor, "Driver cognitive distraction detection: Feature estimation and implementation," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 2007, vol. 221, no. 9, pp. 1027-1040, doi: 10.1243/09544070JAUTO332.