

Modelling self-reported driver perspectives and fatigued driving via deep learning

Alexandros Zoupos¹, Apostolos Ziakopoulos^{2*}, George Yannis²

¹ School of Engineering, The University of Edinburgh, United Kingdom

² Department of Transportation Planning & Engineering, National Technical University of Athens, Greece

* Corresponding author: apziak@central.ntua.gr

Handling editor: **Aliaksei Laureshyn**, Lund University, Sweden

Reviewers: **Salvatore Damiano Cafiso**, University of Catania, Italy; **Paolo Intini**, Polytechnic University of Bari, Italy

Received: 25 June 2021; Accepted: 7 November 2021; Published: 16 November 2021

Abstract: Driving while fatigued is a considerably understudied risk factor contributing to car crashes every year. The first step in mitigating the respective crash risks is to attempt to infer fatigued driving from other parameters, in order to gauge its extent in road networks. The aim of this study is to investigate the extent to which declared fatigued driving behaviour can be predicted based on overall driver opinions and perceptions on that issue. For that purpose, a broad cross-country questionnaire from the ESRA2 survey was used. The questionnaire is related to self-declared beliefs, perception, and attitudes towards a wide range of traffic safety topics. Initially, a binary logistic regression model was trained to provide causal insights on which variables affect the likelihood that a driver engaged in driving while fatigued. Drivers reporting driving under the influence of drugs, fatigue, or alcohol, as well as speeding, safety, and texting while driving or drivers who were more acceptable of fatigued driving were more likely to have recently driven while fatigued. In contrast, acceptability of other hazardous behaviours, namely mobile phone use and drunk driving, was negatively correlated with fatigued driving behaviour, as were more responsible driver perspectives overall. To provide a more accurate detection mechanism, which would also incorporate non-linear effects, a Deep Neural Network (DNN) was subsequently trained on the data, slightly outperforming the binary logistic model. From the results of both models, it was concluded that declared fatigued driving behaviour can be predicted from questionnaire data, providing new insights to fatigue detection.

Keywords: driver fatigue, fatigue detection, multi-country survey, deep learning, binary logistic regression

1 Introduction

In the domain of road safety, significant progress has been made in the form of analytical methods, technological advancements, and intensive research (Zou *et al.*, 2018). Despite that progress, road crash numbers persist without visible reductions in recent years, barring the influence of the Covid-19 pandemic, which has greatly disrupted transport systems and caused changes in travel behavior (Aloi *et al.*, 2020). Road safety remains a crucial issue to countries around the world, as road crashes constitute the number one cause of death for people aged between 5 and 29 years old. Furthermore, it is estimated that approximately 1.35 million people

each year lose their lives due to involvement in crashes (*World Health Organization, 2018*). In a study conducted in the US with a sample greater than 2,000,000 crashes, it was found that in 94% ($\pm 2.2\%$) of crashes, the critical factor behind the crash lies with the drivers (*Singh, 2015; 2018*).

For road crashes to yield reductions once again, intensified focus needs to be placed on investigating individual causes of road crashes, with an emphasis on more subtle and non-obvious risk factors. Such a case is driving under the influence of driver fatigue.

Fatigue can refer to the tiredness an individual experiences as a result of mental or physical effort. An example of fatigue is the tiredness caused from driving for a long time (*Talbot & Filtness, 2017*). Driving is a complex task, requiring a variety of skills such as quick response time and attention. Due to fatigue all these skills may decrease, thus increasing the probability of road crashes (*Grossman & Rosenbloom, 2016*). It has been shown, for instance, that insufficient sleep, which is a major factor resulting in fatigue while driving, led to a significantly longer break reaction time in a harsh breaking test (*Miyata et al., 2010*).

Indicatively, past research reported that 58.6% of drivers admit to occasionally driving while fatigued or drowsy, and 14.5% admitted to falling asleep behind the wheel, while nearly 2% were involved in fatigue-related crashes during the span of a year (*Vanlaar et al., 2008*). On a similar note, the U.S. National Highway Traffic Safety Administration reported that during 2019, a drowsy driver was involved in almost 2% of the total crash fatalities (*NHTSA, 2020*). A comparable percentage of 1.7% was identified for Finnish drivers (*Radun et al., 2015*). In the U.S., 13.1% of collisions that results in a person being admitted to a hospital and 16.5% of fatal road crashes have been found as related to driving while fatigued (*Tefft, 2010*).

A study conducted in Ontario, in Canada, found that in a sample of 750 drivers, 60% of the drivers have driven while fatigued at least in some instances. Moreover, 14.5% of the participants reported that in the past year they have fallen asleep at least once while driving (*Vanlaar et al., 2008*). Another factor that indicates how often drivers might drive while fatigued is the fact that over 50% of the population in industrialized countries, do not receive sufficient sleep on a regular basis (*U.S. National Sleep Foundation, 2013*). Furthermore, *Smolensky et al. (2011)* highlighted how daylight fatigue is a non-negligible risk factor as well. Several prevalent sleep disorders as well as other medical disorders can results in daytime fatigue crashes. These conditions may vary from asthma and rheumatoid arthritis to insomnia and obstructive sleep apnea (*Smolensky et al., 2011*).

Although there has been evidence that sleep-deprived drivers may be as dangerous as drivers that drive under the influence of alcohol, awareness of this danger remains relatively low (*Grossman & Rosenbloom, 2016*). *Nordbakke & Sagberg (2007)* indicated that although many drivers experience symptoms of sleepiness, they do not take them seriously. Furthermore, drivers may continue to drive even after they recognize they are so sleepy or too tired to drive (*Nordbakke & Sagberg, 2007*).

Solutions to the problem of fatigue have been proposed with the introduction of blanket measures instead of driver-specific detection. Indicatively, *Zhang et al. (2016)* analyzed road crash data from 2006 to 2010 in China aiming for their study outcomes to have policy implications in order to manage the problem of fatigue crashes. The results of this study showed that attention should be paid to truck drivers, who have a high risk of fatigued driving. Moreover, it was found that another factor that might reduce fatigue crashes is road monitoring during rush and morning hours, as well as the improvements of road conditions. Lastly, it was suggested that if further measures are also applied, such as mandatory rests after long hours of driving, then fatigue crash cases would be further reduced (*Zhang et al., 2016*). Such measures

may vary from a 30 minute break after 8 hours of driving to a limit of maximum hours of driving every day (*U.S. Federal Motor Carrier Safety Administration, 2011*).

Unlike other driving impairments (e.g. driving under the influence of drugs or alcohol), which are measurable with various devices and imposed legal limits, driver fatigue is very hard to measure. Therefore, as a preliminary approach to this issue, it is imperative to clarify how overall driver perspectives and perceptions of issues regarding driving while fatigued influence their own behavior of fatigued driving. Such an endeavor can benefit from advances in analytical tools in the form of an investigation aided by machine/deep learning to the aggregated level, which, to the knowledge of the authors, is an unexplored research venue.

In light of the previous, the aim of this study is to investigate the extent to which fatigued driving behavior can be predicted based on overall driver opinions and perceptions on that issue. In other words, an examination of how various self-declared beliefs, perceptions and attitudes can influence road user choices on whether to drive under the influence of fatigue is conducted. To that end, a broad cross-country driver sample will be exploited for the exploration of driver fatigue. Data from the ESRA2 survey are utilized, encompassing responses from 47 countries. The input questions are related to self-declared beliefs, perception, and attitudes towards driving, all of which might affect a driver's choice of whether to drive under the influence of fatigue. Initially, a traditional binary logistic regression model is fitted for an exploratory correlation analysis. Subsequently, a Deep Neural Network (DNN) is considered to better capture non-linear effects or underlying patterns in the considered relationships, which may be reflected in better predictive performance by the DNN.

2 Literature relevant to fatigue and drowsiness detection

Several studies have addressed the problem of fatigue by aiming to detect it in drivers. Such studies typically base their analysis on different features of driver behavior, and have been successful on fatigue detection. Research conducted by *Feng et al. (2009)* found that fatigued drivers usually make less steering micro-corrections than normal ones. Furthermore, *Sayed & Eskandarian (2001)* were able to create an ANN, which used the steering angle as input, that classified drivers as drowsy or non-drowsy. The model had 88% accuracy for drowsy drivers, and a 90% for non-drowsy ones. Another interesting study was able to combine and analyze the steering wheel angles and the yaw angles, in order to detect fatigue with an accuracy of 88% (*Li et al., 2017*).

Other behavioral features, which can be obtained more easily and can be useful when it comes to detecting drowsy driving, are the lateral and longitudinal accelerations of the vehicle. *Wang et al. (2015)* used the random forest (RF) algorithm and managed to create a model that detects drowsy drivers with an accuracy of 84.8%. The main input parameters for their model consisted of three parameter types, including steering angle, lateral acceleration and longitudinal acceleration. The authors note that the RF algorithm showed increased resilience to data noise compared to an ANN algorithm developed for the task of drowsiness detection.

Apart from methods based purely on driving behavioral features, there have been studies that managed to detect fatigued driving using physiological features. A common physiological response used to detect fatigue is the response of the eyes. *Khan & Mansoor (2008)* managed to track the eyes of drivers in order to detect real time fatigue, with an accuracy of 90%. In order to make such a detection system more applicable, *Kong et al. (2015)* suggested a new system that which is based on machine vision and the AdaBoost algorithm. This system can be integrated in smart devices, such as a smartphones or tablets, and uses the percentage of duration of closed-eye state (PERCLOS). That system aims to detect eye openness/closeness and classify fatigue based on this rate, instead of direct detection of eye parameters.

Sometimes, visual systems used to monitor face and eye are susceptible to environmental factors and often have large errors in judgment since they use fixed thresholds for the face characteristics. To address this problem, *Kening et al. (2020)* proposed a new fatigue detection algorithm, which is based on a multi-feature facial fusion process. In contrast to ordinary fatigue driving detection algorithms, the proposed one takes into account the driver's characteristics and makes predictions based on the size of their mouth and eyes. The authors of the study claim that this feature adds high accuracy to their detection system. When tested, the algorithm was able to detect fatigue with an accuracy of 95.10%, using the time drivers' eyes were closed, blink frequency and yawning frequency as inputs.

As a different response, studies have tried using advanced physiological sensors in order to predict fatigued driving. Specifically, *Zhang et al. (2014)* created a real-time fatigue monitoring system, which used electroencephalogram (EEG), electromyography (EMG), and electrooculography (EOG) signals. Their model managed to obtain accuracies from 96.5% to 99.5%. Although the latter method of detecting real-time fatigue produces reliable and accurate results, it has been suggested that it is not the ideal system for implementation since the data acquisition is hard, with electrodes being needed. Instead, a combination of vehicular features, such as steering wheel angle and lane crossing, and physiological features (e.g. eye and head movement) has been suggested to produce the most applicable and reliable method for fatigue detection (*Sikander & Anwar, 2018*).

The above studies have achieved successful static or real-time fatigue detection. Fatigue prediction, however, has received less attention from researchers. A prediction process becomes crucial for the near future, as automated driving is expected to become an integral part of people's lives. In all likelihood, there will be instances where the system must request the driver to retake control of the vehicle, and they must be ready and able to assume control. Some studies have focused on how fatigue progresses in automated driving. For example, it was found that after 20 minutes of automated driving, drivers had fatigue signs (*Feldhütter et al., 2016*). *Gonçalves et al. (2016)* concluded that after 15 minutes of highly automated driving, participants were feeling fatigue. A more recent study managed to predict the time in which the driver, in highly automated driving, transitions from non-fatigued to fatigued, which was achieved using physiological features. The model managed to predict the transition 13.8 seconds ahead of time with an accuracy of 97.4% and 99.1%, for the two different models used (*Zhou et al., 2020*).

Most of the aforementioned studies were conducted in driving simulators, and although they produced high accuracy models, their transferability may become challenging. *Li et al. (2021)* proposed a simple method to detect fatigue driving in a real-world driving environment. This method works by measuring the driver's grip force on the steering wheel. The results showed that there is a high correlation between the grip force and fatigue, with fatigued drivers having a larger variation in their grip force. Although the recognition rate was 88.3%, the study lays the foundation for applicable fatigue detection systems. Another more readily applicable method is the one developed by *Parsa et al. (2021)*, which uses pressure sensors placed on the seat and the backrest of the driver. The fatigue index of the system is divided in five parts and the driver is considered to be tired when they have crossed 75% of the fatigue index. Drivers are assigned a specific fatigue index depending on their body position. For example, if the driver is in the "flopped" position, which means that they are immersed into the seat, then the system recognizes them as extremely fatigued.

Despite these notable results, a knowledge gap for more aggregate predictions and estimates for the driving population is detected in the literature. A critical problem related to fatigued driving is the fact that fatigue is an inherently difficult condition to regulate. Law enforcement

authorities are able to measure and detect driving under the influence of alcohol or drugs. Respective limits of acceptable thresholds are established by law, and their exceedance is considered a violation. However, such protocols and enforcement systems are not established for fatigue.

Therefore, to aid road management authorities and lawmakers, it is necessary to obtain an estimate of the extent of the phenomenon of driving while fatigued. The design of the present study aims to exploit a broad questionnaire sample in lieu of driver simulators and eye-tracking devices, in order to (i) gain a wider perspective from a large sample consisting of participants from 47 countries, which would not be feasible for an instrumented experiment, and to (ii) explore the feasibility and usefulness of using questionnaire data for the prediction of driving while fatigued amongst driver samples. Regarding the second point, adequate results promise more transferability, especially in countries/regions with less developed infrastructure, as a questionnaire can easily be administered and subsequently analyzed.

3 Methodology

3.1 The ESRA survey

ESRA (E-survey of Road users' Attitudes) is a joint initiative of road safety institutes from all over the world, aiming to collect and analyze data from road users on their opinions, attitudes and behavior towards traffic risks. With the aforementioned data ESRA aims to scientifically support road safety policy at international level, as well as to develop a series of road safety key performance indicators. Each ESRA iteration aims to include more countries. The ESRA edition used in this research, ESRA2, was completed in two waves: the first one was released in 2018 with 32 countries taking part, while the second was completed in 2020 and had an additional 16 countries, totaling 48 participant countries. More details about the methodology, data processing and questionnaires are provided in *Pires et al. (2020)*.

Several past studies have utilised the expertise of ESRA and conducted analyses exploiting ESRA data. Indicatively, *Yannis et al. (2020)* explored the road safety performance and attitudes of vulnerable road users (VRUs). *Goldenbeld et al. (2020)* found that in Europe drug driving was less frequently reported than drinking and driving, while in other continents it was reported as frequently. *Ziakopoulos et al. (2021a)* examined several unsafe powered-two-wheeler rider behaviours using ESRA2 data and found that they are correlated with a number of causal factors and with each other.

The ESRA2 survey addresses a wide range of thematic areas for most road users. The main thematic areas considered in this paper were (*Meesmann et al., 2021*):

- Self-declared behavior in traffic
- Acceptability of safe or unsafe traffic behaviors
- Attitude towards safe or unsafe traffic behaviors
- Subjective safety and risk perception
- Support for policy measures

Several variables with the ESRA questionnaire were considered for this research. Through numerous trials and iterations in the modelling process, and the evaluation of the questionnaire data with appropriate methodologies and specifically the LIME framework (*Ribeiro et al., 2016*), a final selection of input variables was conducted. Table 1 lists all the questions used as input in models of the present study, which are a subset of the greater ESRA questionnaire as presented in *Meesmann et al. (2021)*.

Table 1. Considered variables from the ESRA2 questionnaire.

| Abbreviation | Question | Scale |
|---|--|--|
| Self-declared behaviour in traffic | | |
| V012_1b_14 | <i>Over the last 30 days, how often did you as a CAR DRIVER drive when you were so sleepy that you had trouble keeping your eyes open?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1b_8 | <i>Over the last 30 days, how often did you as a CAR DRIVER drive without wearing your seatbelt?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1b_13 | <i>Over the last 30 days, how often did you as a CAR DRIVER read a text message/email or check social media (e.g. Facebook, twitter, etc.) while driving?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1b_2 | <i>Over the last 30 days, how often did you as a CAR DRIVER drive after drinking alcohol?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1b_9 | <i>Over the last 30 days, how often did you as a CAR DRIVER transport children under 150cm without using child restraint systems (e.g. child safety seat, cushion)?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1b_7 | <i>Over the last 30 days, how often did you as a CAR DRIVER drive faster than the speed limit on motorways/freeways?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1b_4 | <i>Over the last 30 days, how often did you as a CAR DRIVER drive after taking medication that carries a warning that it may influence your driving ability?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1b_1 | <i>Over the last 30 days, how often did you as a CAR DRIVER drive when you may have been over the legal limit for drinking and driving?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1b_5 | <i>Over the last 30 days, how often did you as a CAR DRIVER drive faster than the speed limit inside built-up areas?</i> | <i>At least once (2-5) – never (1)</i> |
| V012_1a_3 | <i>Over the last 12 months, how often did you as a CAR DRIVER read a text message or email while driving?</i> | <i>At least once (2-5) – never (1)</i> |
| Acceptability of traffic behaviour | | |
| V014_1 | <i>How acceptable do you, personally, feel it is for a CAR DRIVER to drive when he/she may be over the legal limit of drinking and driving?</i> | <i>Acceptable (4-5) – unacceptable/neutral (1-3)</i> |
| V014_9 | <i>How acceptable do you, personally, feel it is for a CAR DRIVER to talk on a hand-held mobile phone while driving?</i> | <i>Acceptable (4-5) – unacceptable/neutral (1-3)</i> |
| V014_12 | <i>How acceptable do you, personally, feel it is for a CAR DRIVER to drive when they're so sleepy that they have trouble keeping their eyes open?</i> | <i>Acceptable (4-5) – unacceptable/neutral (1-3)</i> |
| Support for policy measures | | |
| V018_5 | <i>Do you support or oppose legal obligation to install Dynamic Speed Warning signs (traffic control devices that are programmed to provide a message to drivers exceeding a certain speed threshold)?</i> | <i>Support (4-5) – oppose/neutral (1-3)</i> |

Table 1 (cont.). Considered variables from the ESRA2 questionnaire.

| Abbreviation | Question | Scale |
|--|---|--|
| Subjective safety and risk perception | | |
| V017_2 | How often do you think each of the following factors is the cause of a road crash involving a car? Drive after taking drugs (other than medication) | Often/frequently (4-6) – not that often/not frequently (1-3) |

Figure 1 shows the distribution of the answers to each question of Table 1, after the elimination of incomplete or unknown entries.

**Figure 1. Distribution of answers to ESRA2 questions.**

3.2 Data collection & sample characteristics

The ESRA2 online survey provides a dataset from many countries around the world, for all road users. The total sample size was 45,000 road users across 48 countries (men: 49.6%, women: 50.1%, other: 0.3%) (Meesmann *et al.*, 2021).¹ From this dataset, 30,683 participants were regular car drivers, defined as a person who uses their car a few days a month or more. The initial aim of the survey was to have a representative sample of the adult population of at least 1,000 respondents in each country. However, the minimum sample of 1,000 respondents

¹ Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Poland, Portugal, Serbia, Slovenia, Spain, Sweden, Switzerland, United Kingdom, Canada, USA, Australia, India, Israel, Japan, Republic of Korea, Egypt, Kenya, Morocco, Nigeria, South Africa, Bulgaria, Iceland, Luxembourg, Norway, Colombia, Lebanon, Malaysia, Thailand, Vietnam, Benin, Cameroon, Ghana, Ivory Coast, Tunisia, Uganda, Zambia.

could not be met in all countries, due to the size of the available online panels in certain countries (Ziakopoulos *et al.*, 2021a). Table 2 summarizes the sample size, gender, and age group distribution by regions in the ESRA2 study (Meesmann *et al.*, 2021).

Table 2. Sample size, gender, and age group distribution by region.

| Region | Sample size | Gender | | | Age group | | | | | |
|--------------|-------------|--------|--------|-------|-----------|--------|--------|--------|--------|------|
| | | Male | Female | Other | 18-24y | 25-34y | 35-44y | 45-54y | 55-64y | 65y+ |
| Europe | 25,987 | 48% | 52% | 0% | 10% | 16% | 17% | 18% | 16% | 23% |
| Asia/Oceania | 8,590 | 50% | 49% | 1% | 26% | 31% | 19% | 12% | 6% | 6% |
| America | 3,009 | 48% | 51% | 0% | 12% | 17% | 16% | 17% | 16% | 20% |
| Africa | 7,528 | 49% | 51% | 0% | 26% | 28% | 19% | 13% | 7% | 6% |

As the goal of the survey was to have a representative sample of the population, hard quotas were implemented for gender and age distribution per country, along with other quotas described in previous research (Meesmann *et al.*, 2021; Ziakopoulos *et al.*, 2021b). In the present study, data from Iceland had to be discarded due to methodological differences, translation difficulties and other barriers which led to an overwhelming number of missing entries, therefore 47 countries remained in total.

4 Statistical Background

As per the aim of the present study, statistical methodologies are employed to investigate how various self-declared beliefs, perceptions and attitudes can influence road user choices on whether to drive under the influence of fatigue, as self-reported by car drivers. Within the ESRA2 questionnaire, this variable is reported in a binary format (0 for not even once driving under the influence of fatigue during the last 30 days, 1 in the opposite case). Both a traditional statistical method and a deep learning method are considered, in order to tackle the issue of fatigued driving with a linear modelling and a non-linear modelling approach. Initially, a common binary logistic regression model is implemented to provide a basis for causal interpretation as well as a benchmark for measuring the performance of a deep neural network (DNN).

4.1 Binary Logistic Regression

Binary logistic regression models are well established statistical functional-based methods, widely applied in order to model binary outcomes. A brief outline is provided here; the reader can refer to Washington *et al.* (2020) for more in-depth explanations. The model begins by considering a linear predictor, as expressed by Equation (1):

$$y_i = b_0 + \sum_{k=1}^n b_k * x_{ik} + \varepsilon_i \quad (1)$$

Where:

y_i is the dependent (or response) variable of observation i

x_{ik} are the independent (or explanatory) n variables of observation i

b_k is the coefficient of a particular x_k

b_0 is the constant term

ε_i is the error term of the model at observation i .

If a utility function is considered by expressing the predictor without considering the error term, as given by Equation (2):

$$U = b_0 + \sum_{k=1}^n b_k * x_{ik} \quad (2)$$

Then the probability P that the dependent variable belongs in a specific class is given by Equation (3):

$$P = \frac{e^U}{e^U + 1} \quad (3)$$

The reader is referred to several available sources which describe the underlying statistical processes of (binary) logistic regression in great detail (e.g. *Harrell, 2015; Tranmer & Elliot, 2008*). Model selection process between models including different independent variable subsets is conducted by the examination of the corrected Akaike Information Criterion (AICc). The sign of the coefficients ought to be examined to ensure that results are consistent with the real mechanisms that are modelled and thus reasonable.

Since residuals cannot be used as a test for goodness of fit of the model, the Hosmer and Lemeshow Test was used. The test calculates whether observed event rates match the expected ones in subgroups of the model population (*Hosmer & Lemeshow, 1989*). The output of the test is a chi-squared value and a p-value. A p-value higher than 0.05 indicates a good fit, and an acceptable model.

In the present research, binary logistic regression was selected as a baseline model for benchmarking purposes in order to measure DNN performance. Random-effects models were also considered during the model training phase and investigations were made following *Bates et al. (2011)*. However, results indicated that random intercept effects were not statistically significant, and the implementation of random slopes is inappropriate for questionnaire replies. Therefore, the methodological details of random effects models are omitted here as these models were discarded.

4.2 Deep Neural Network

The second method used to predict whether a driver has recently driven while fatigued is through a deep artificial neural network (DNN/ANN). The DNN family of algorithms was partly inspired by the biological neural networks, which are able to process information in parallel. Each artificial neuron has a number of inputs that produce a single output, which is then sent to other neurons. A connection between neurons is associated with a weight, which serves to express different relative importance. DNN methods have been widely used in quantitative research and encompass a large variety of techniques; indicatively, see *Samarasinghe (2006)*.

The DNN consists of a number of different layers. The first one is the input layer of units (or neurons, nodes), afterwards there is a number of hidden layers of units (usually two or three), and the final (output) layer has the output neurons. Each layer has an activation function, computes outputs by combining the weights with the inputs (*Dawson & Wilby, 1998*). The activation function not only introduces nonlinearity into the DNN, but also bounds the value of the neuron so that the DNN is not paralyzed by divergent neurons (*Wang, 2003*). The activation functions used in this paper were the Rectified Linear Unit (ReLU) and the Sigmoid. ReLU has been found to allow for rapidly converging DNNs; it is a pseudo-linear function that does not

become trapped in local minima and allows for unhindered backpropagation (Hara *et al.*, 2015). The Sigmoid function was used in the output layer since the output of the function returns a result between 0 and 1, which is what is needed in a binary classification model. DNNs can be applied to supervised and unsupervised tasks and have been reported to overcome the necessity of feature engineering (Schmidhuber, 2015). Figure 2 is a simplified version of the DNN used in this research, where there are two hidden layers, and the output layer has two neurons, since the outcome variable is binary.

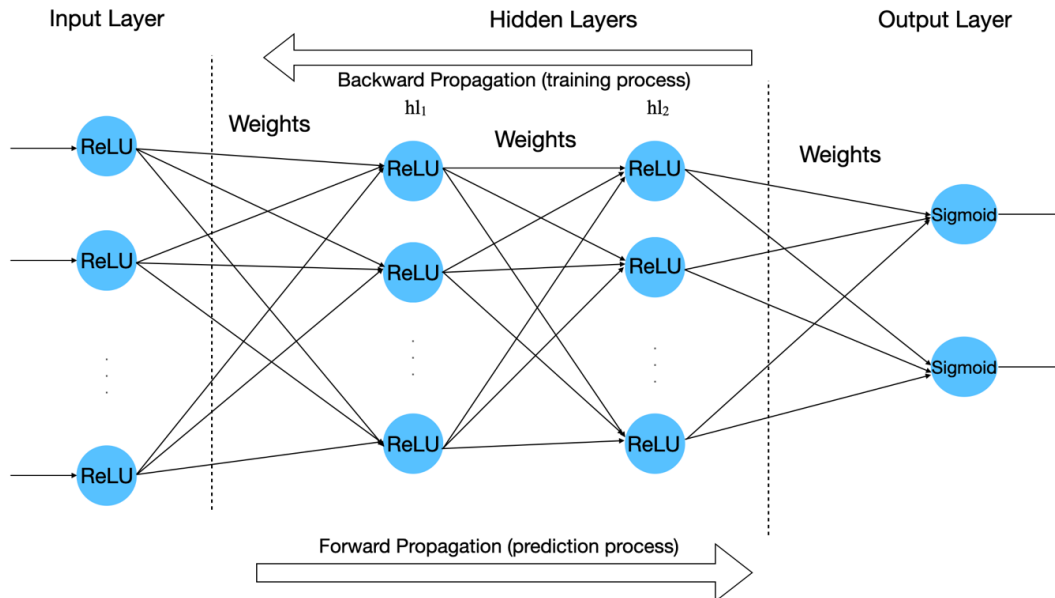


Figure 2. Theoretical DNN structure.

The training of the DNN (backward propagation) is performed by calculating the weights using the training set. This is conducted in many iterations, depending on the size of the dataset and the number of variables. Using the complete, trained DNN, data from the test set are used as input and predictions are obtained from the output layer (forward propagation).

Some of the most widely used goodness-of-fit indicators for the DNN model are the accuracy, the loss, and the mean square error (MSE). The term accuracy measures the fraction of observations that are correctly classified, while loss measures the difference between predicted and actual values. MSE is the average of the square of the distances between the predicted and actual values (Arnold, 2017).

Two additional important parameters of the model are the epoch number (i.e. the number of times that the learning algorithm will process the dataset) and the batch size (i.e. the number of samples to process before updating the internal model parameters) (Gulli & Pal, 2017). The number of nodes in each layer along with the epoch number and batch size are called hyperparameters (in other words, hyperparameters are numeric quantities that govern the DNN structure). To determine the optimal values for the hyperparameters, hyperparameter tuning was conducted. This process trains a number of DNN models with different specification combinations and selects the optimal combination for a given dataset based on better classification performance, higher accuracy and lower MSE.

4.3 Comparative model performance evaluation

Comparative metrics for model performance evaluation were established, since the two modelling processes are too dissimilar to allow for direct comparison. The commonly used ROC (receiver operating characteristic) curves were adopted, which help to visualize the

performance of the model and provide quantitative assessment by measuring the area under the curve (AUC); a ROC-AUC value closer to 1 indicating better distinction between positive and negative classes (Washington *et al.*, 2020). Confusion matrixes were also created in order to inspect the proportions of true positives (TP) and true negatives (TN) as compared to the total sample, in other words, the percentage of correct classification.

5 Model Development and Results

As per the aforementioned, the outcome variable for the models referred to their driving under the influence of fatigue. Specifically: “Over the last 30 days, how often did you as a CAR DRIVER drive when you were so sleepy that you had trouble keeping your eyes open?” (V012_1b_14 abbreviation in the ESRA2 questionnaire).

The rest of the variables listed in Table 1 were used as input for both models. Since the dependent variable question addresses only car drivers, all other road users were removed from the dataset. After filtering and data cleaning for the elimination of incomplete or unknown entries, the resulting dataset consisted of observations from 31,606 respondent drivers, removing 13,437 observations from the dataset. The dependent variable responses were slightly imbalanced (24,416 respondents replied never (coded as 0) and 7,190 respondents replied at least once or more times (coded as 1)), resulting to an imbalance ratio of 30% between classes.

As per standard good practice, models were trained in a subset of the entire usable dataset, while their predictions were tested in data which was completely new for the model. The chosen ratio was 70% for the train subset and 30% for the test subset, and the two sets were kept identical for the binary logistic model and the DNN. All analyses are conducted using R-studio (R Core Team, 2019) and a number of packages.

5.1 Binary logistic regression results

Binary logistic regression models were fitted using backward elimination. It should be noted that upsampling and downsampling techniques were explored but they did not produce any noticeable gains in the final models and were ultimately not adopted. When examining a model, all variables with a p-value of 0.05 are considered statistically significant (95% significance level). The final model results appear on Table 3.

Table 3. Binary logistic regression model for fatigued driving.

| Independent variable | Coefficients | | | |
|----------------------|---------------|------------|---------|---------|
| | Beta Estimate | Std. Error | z value | p-value |
| <i>Intercept</i> | -4.479 | 0.114 | -39.460 | < 0.001 |
| V014_12 | 0.891 | 0.030 | 29.259 | < 0.001 |
| V012_1b_8 | 0.150 | 0.021 | 7.159 | < 0.001 |
| V014_1 | -0.230 | 0.034 | -8.726 | < 0.001 |
| V012_1b_13 | 0.324 | 0.031 | 10.619 | < 0.001 |
| V012_1b_2 | 0.189 | 0.033 | 5.811 | < 0.001 |
| V012_1b_7 | 0.169 | 0.019 | 9.114 | < 0.001 |
| V012_1b_4 | 0.489 | 0.028 | 17.431 | < 0.001 |
| V017_2 | -0.053 | 0.010 | -5.301 | < 0.001 |
| V012_1b_1 | 0.256 | 0.038 | 6.680 | < 0.001 |

Table 3 (cont.) Binary logistic regression model for fatigued driving.

| Independent variable | Coefficients | | | |
|----------------------|---------------|------------|---------|---------|
| | Beta Estimate | Std. Error | z value | p-value |
| V012_1b_5 | 0.187 | 0.022 | 8.339 | < 0.001 |
| V012_1a_3 | 0.107 | 0.028 | 3.864 | < 0.001 |
| V018_5 | -0.039 | 0.016 | -2.371 | 0.018 |
| V014_9 | -0.058 | 0.024 | -2.467 | 0.014 |

The ROC curve produced by the binary logistic regression model had an AUC of 0.793 and is shown in Figure 3.

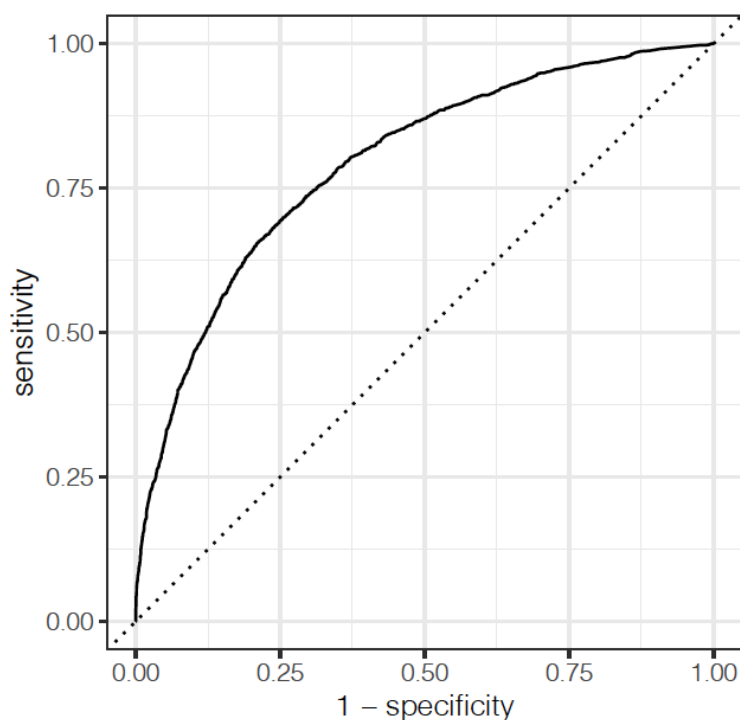
**Figure 3. Binary logistic regression ROC curve [AUC 0.793].**

Table 4 shows the confusion matrix for the test data using this model, with a classification probability threshold of 0.30 (in other words, if an observation had a probability ≥ 0.30 of being 1, then it would be classified as 1). The correct classifications (TP+TN), lying on the diagonal, amount to 79.0%.

Table 4. Binary logistic confusion matrix.

| Predicted value | Actual value | |
|-----------------|--------------|--------|
| | 0 | 1 |
| 0 | 67.0 % | 10.8 % |
| 1 | 10.3 % | 12.0 % |

The binary logistic regression model can be insightful when it comes to showing how answers impact the prediction of the model. In Table 3 the negative signs of the coefficients indicate that the specific variable contributes to the driver being less likely to have recently driven while fatigued. The opposite happens with positive coefficients.

The first category of questions is the “Self-declared behavior in traffic”, which addresses self-reported behavior towards topics related to driving under the influence of drugs, fatigue, or alcohol, as well as speeding, safety, and texting while driving. All of the variables in this category are positively correlated with recent fatigued driving. In other words, the more often drivers have reported such behavior, the more likely it is that they have recently driven while fatigued.

The second category of questions is the “Acceptability of traffic behavior”, which has questions related to the acceptability of drivers towards impaired driving, speeding, and texting. Variables in this category have both positive and negative coefficients. The variable with the highest positive coefficient is V014_12: “How acceptable do you, personally, feel it is for a CAR DRIVER to drive when they’re so sleepy that they have trouble keeping their eyes open?”. It is not surprising for this question to have the highest positive coefficient since it is a variable denoting a close a priori qualitative relation with outcome variable. Results suggest that the higher personal acceptability of driving while fatigued leads to more likely engagement in such behavior.

Variable V014_9 (i.e. “How acceptable do you, personally, feel it is for a CAR DRIVER to talk on a hand-held mobile phone while driving?”) and variable V014_1 (i.e. “How acceptable do you, personally, feel it is for a CAR DRIVER to drive when he/she may be over the legal limit of drinking and driving?”) both have negative coefficients. In other words, this finding denotes that mobile use – and the related distraction it causes – or alcoholic inebriation are less likely to lead to driving while fatigued. These outcomes are likely interpreted by overcompensating effects on the part of the drivers. Despite potentially committing violations when driving while using mobile phone or while drunk, drivers realize that they are driving with compromised driving ability. As a result, they avoid engaging in such behavior while fatigued to avoid further deterioration of driving ability.

The third category of questions is the “Support for policy measures”, which is related to questions regarding the support of measures for safety and against drunk driving, and speeding. The higher the numeric number for the answer the higher the support of the driver for the specific measure. The respective question, V018_5, (i.e. “Do you support or oppose legal obligation to install Dynamic Speed Warning signs (traffic control devices that are programmed to provide a message to drivers exceeding a certain speed threshold)?”) is negatively correlated with recent fatigued driving. This finding is expected since drivers who support these measures are more likely to be more responsible when it comes to driving overall, and, as such, avoid driving while fatigued.

The last category of questions is the “Subjective safety and risk perception”, where drivers were asked how often they believe certain factors are the cause of a road crash involving a car. The higher the numeric answer, the more frequent drivers believe these factors are causes of a road crash. The respective question, V017_2, (i.e. “How often do you think each of the following factors is the cause of a road crash involving a car? Drive after taking drugs (other than medication)”) is negatively correlated with recent fatigued driving, once again hinting at responsible driver perspectives reflecting to more responsible driver behavior.

When examining the model overall, and despite an adequate classification performance, the Hosmer & Lemeshow Test was statistically significant: $\chi^2_{[df=8]} = 125.58$, $p < 0.001$. This outcome may indicate a subpar model fit for some of the strata of the sample, and provides further incentive to examine non-linear modelling such as the DNN.

As per the aforementioned, random-effects models were also considered, trained following *Bates et al. (2011)*. A configuration of random intercepts for each respondent was considered. However, when conducting the ANOVA test (log-likelihood test) between the fixed-effects and

the random intercepts binary logistic models, the outcome was not statistically significant, denoting that the inclusion of random effects on an individual level is not justified. The baseline fixed-effects model is thus retained.

5.2 Deep Neural Network results

The DNN utilized the exact same dataset and variables as binary logistic regression, including the random partition between training and test dataset. As described, the DNN created in this study comprised four layers in total. The input layer featured 124 neurons, the two hidden layers featured 64 and 32 neurons respectively, and the output layer featured 2 neurons.

After hyperparameter tuning, the combination with the highest accuracy and the smallest loss was obtained. Thus, the epoch number of the model was 7, since for higher values overfitting was observed, and the batch size was 32. It should be noted that due to the nature of DNNs there might be slight fluctuations between two executions, even if the model was trained with the same dataset, due to randomizations in the process. It should be emphasized, however, that the executions did not have a big difference between them. Figure 4 shows the ROC curve produced by the DNN model, which had an AUC of 0.801. The run had an accuracy of 0.815, a loss of 0.423, and MSE of 0.134.

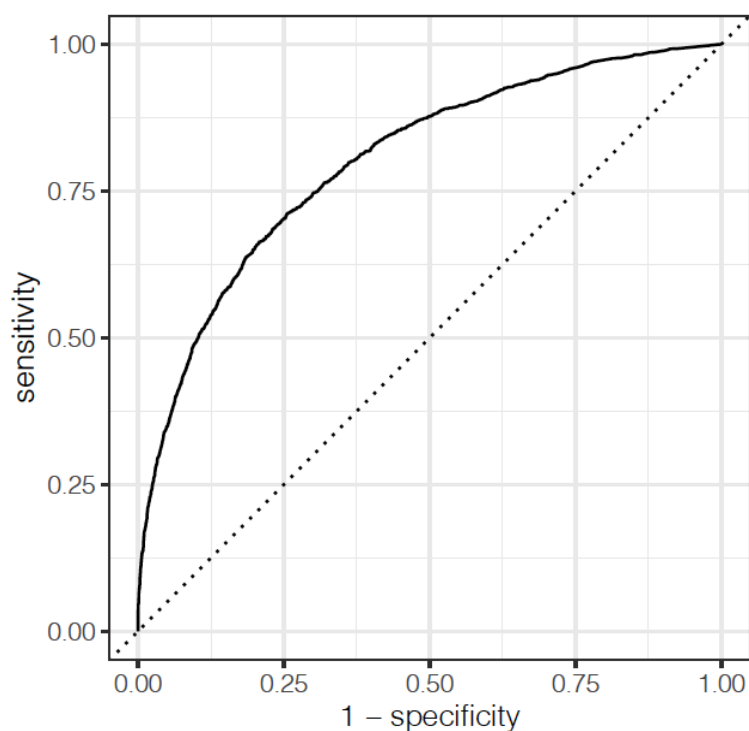


Figure 4. DNN ROC curve [AUC 0.801].

The DNN outperformed the binary logistic regression in terms of ROC-AUC, and provided comparable results overall. The correct classifications (TP+TN), lying on the diagonal, amount to 80.1%, as shown on Table 5.

Table 5. DNN confusion matrix.

| | Actual values | |
|------------------|---------------|-------|
| | 0 | 1 |
| Predicted values | | |
| 0 | 67.6% | 10.4% |
| 1 | 9.5% | 12.5% |

The examination of the influence of each independent variable in the DNN output is also fruitful. The test subset is again considered. For each independent variable, the test subset is separated in two subsets based on the values of the considered variable (since they are all in a binary format). The aggregate estimate of the DNN and the respective accuracy of predictions for each subject is then calculated. Results are shown on Table 6.

Table 6. Influence of each variable on DNN predictions.

| <i>Independent variable</i> | <i>Average prediction</i> | <i>Accuracy</i> | <i>Average prediction</i> | <i>Accuracy</i> | <i>Prediction difference</i> |
|-----------------------------|--|-----------------|-------------------------------|-----------------|------------------------------|
| | <i>Never (1)</i> | | <i>At least once (2-5)</i> | | |
| <i>V012_1b_8</i> | <i>0.092</i> | <i>69.30%</i> | <i>0.165</i> | <i>73.96%</i> | <i>0.073</i> |
| <i>V012_1b_13</i> | <i>0.090</i> | <i>67.18%</i> | <i>0.154</i> | <i>71.60%</i> | <i>0.064</i> |
| <i>V012_1b_2</i> | <i>0.120</i> | <i>67.59%</i> | <i>0.200</i> | <i>74.19%</i> | <i>0.080</i> |
| <i>V012_1b_9</i> | <i>0.113</i> | <i>69.19%</i> | <i>0.179</i> | <i>73.60%</i> | <i>0.066</i> |
| <i>V012_1b_7</i> | <i>0.083</i> | <i>70.66%</i> | <i>0.142</i> | <i>74.67%</i> | <i>0.059</i> |
| <i>V012_1b_4</i> | <i>0.107</i> | <i>67.77%</i> | <i>0.206</i> | <i>67.59%</i> | <i>0.099</i> |
| <i>V012_1b_1</i> | <i>0.108</i> | <i>72.31%</i> | <i>0.217</i> | <i>82.22%</i> | <i>0.109</i> |
| <i>V012_1b_5</i> | <i>0.091</i> | <i>73.30%</i> | <i>0.144</i> | <i>73.79%</i> | <i>0.053</i> |
| <i>V012_1a_3</i> | <i>0.089</i> | <i>75.94%</i> | <i>0.145</i> | <i>72.56%</i> | <i>0.056</i> |
| | <i>Unacceptable/Neutral (1-3)</i> | | <i>Acceptable (4-5)</i> | | |
| <i>V014_1</i> | <i>0.139</i> | <i>74.63%</i> | <i>0.433</i> | <i>82.50%</i> | <i>0.294</i> |
| <i>V014_9</i> | <i>0.135</i> | <i>74.39%</i> | <i>0.268</i> | <i>72.80%</i> | <i>0.133</i> |
| <i>V014_12</i> | <i>0.145</i> | <i>74.57%</i> | <i>0.274</i> | <i>69.23%</i> | <i>0.129</i> |
| | <i>Oppose/Neutral (1-3)</i> | | <i>Support (4-5)</i> | | |
| <i>V018_5</i> | <i>0.161</i> | <i>75.14%</i> | <i>0.124</i> | <i>75.46%</i> | <i>-0.037</i> |
| | <i>Not that often/Not frequently (1-3)</i> | | <i>Often/Frequently (4-6)</i> | | |
| <i>V017_2</i> | <i>0.152</i> | <i>75.50%</i> | <i>0.128</i> | <i>76.26%</i> | <i>-0.024</i> |

Table 5 shows the confusion matrix for the test data for the DNN. The classification probability threshold for the DNN was set to 0.35 after several trials as it yielded the best overall results.

It can be observed that, by examination of the respective subsets, most variables affect aggregate predictions in the same manner as in the binary logistic model. In other words, changes in the binary categories of most variables (moving from lower to higher scores) appear to increase average predictions if that variable had a positive binary logistic coefficient (e.g. *V014_12*) or to decrease average predictions if that variable had a negative binary logistic coefficient (e.g. *V018_5*). There are two exceptions: *V014_1* and *V014_9*, which switch to a positive contribution in the DNN. The explanation for this discrepancy is, most likely, that the effect of these two variables is not sufficiently isolated from the other variables. This may be exacerbated by the fact that, as shown in the distributions of Figure 1, there is a considerable imbalance of the negative against positive cases for these variables, and thus their contributions are less straightforward to interpret.

6 Discussion

From the results of the ROC curve and the confusion matrix the binary logistic regression model seems to perform at a satisfactory level when it comes to predicting whether drivers recently drove while fatigued. However, the results from the Hosmer & Lemeshow test showed that the model did not achieve an adequate fit on the data, and random-effects models were not statistically significant from the fixed-effects model. For that reason, a more advanced method of modelling was used, the DNN. The DNN managed to outperform the binary logistic regression, yielding a small gain in ROC-AUC and correct classifications, probably due to non-linearity in the model.

Considering the two utilized models, it can be concluded that driving behaviour can be predicted from questionnaire data on overall driver opinions and perceptions on that issue. Specifically, using the answers of the ESRA dataset, fatigued driver behaviour, in terms of driving when a driver was so sleepy that they had trouble keeping their eyes open, was classified in a binary format. By covering a wide range of topics, from support to legal measures to self-declared behaviour, it was possible to detect driver behaviours towards fatigued driving to an adequate degree. This would hint that the present variables can provide a basis for further expansion for the investigation of the extend of fatigue, which is an elusive trait to measure in road users.

Arguably, comparison of the present results with existing literature is not a straightforward process, as not many studies have been conducted for fatigued driving prediction based on self-reported driver perspectives from an aggregate multi-country driver sample. Nonetheless, certain parallels can be drawn from previous research results for specific variables. Higher acceptability of traffic driving under the influence of fatigue was positively correlated with self-declared fatigued driving. This finding is in line with the results reported by *Watling (2014)* and those of *Jackson et al. (2011)* regarding perceived social norms. In turn, lower acceptability can be reasonably linked with higher perception of risk, therefore it can be considered that this result is also in line with the findings of *Watling et al. (2014)*, who reported that higher risk perception was negatively correlated with self-reports of continuing to drive while drowsy. Similarly, links have been made by *Radun et al. (2015)*, especially towards explaining the less frequent engagement of women in fatigued driving, as they are reported to have better perceptions of driving behaviours with increased risk. Nonetheless, women had been reported as having greater relative risk of sleepiness in earlier research (*Wilson et al., 2006*). Finally, increased support for policy measures was associated with less frequent fatigued driving in this study. This can be considered an extension of the results reported by *Beck et al. (2018)*, where participants supported messages of the graphic consequences of drowsy driving as they believed it reduced such instances based on their own behaviours.

Naturally, a number of limitations exist in the current approach. Most variables considered were not directly related to fatigued driving, as ESRA is a more general, high-level survey. A critical issue is that the analysed data are self-reported questionnaire data revealing stated behaviour as opposed to observed behaviour. It is well known that this kind of data are often less accurate and include inherent biases (*Backer-Grøndahl & Sagberg, 2011*). The bias may be from a misunderstanding of the question to desirability or recall bias (*Ziakopoulos et al., 2021a*). However, because most of the answers are drawn from a time period of a month recall error can reasonably be assumed to be limited. Another source that might have caused slight inaccuracies was the translation of the survey to 61 languages. An additional point to consider is that the ESRA survey includes general response tendencies and biases, as it is an international respondent survey, which may affect accuracy of conclusions (*Torfs et al., 2016*). Lastly, numerous observations with missing values had to be removed, amounting to a loss of 13,437

observations, comprising almost 30% of the original raw dataset. This is crucial for machine learning tasks, where more observations lead to a larger training set, which in turn typically produces more accurate models.

7 Conclusions

Unlike other driving impairments, such as drugs or alcohol, driver fatigue is very hard to measure. Evidence suggests that driving while fatigued may be as dangerous as driving while inebriated. There remains little awareness on the issue, however, or attempts to quantify and measure it. The present study aimed to investigate the extent to which fatigued driving behaviour can be predicted based on overall driver opinions and perceptions on that issue. To that end, data from the third iteration of the ESRA2 study are utilized, encompassing a broad sample of more than 31,606 respondent drivers from 47 countries. The input questions are related with self-declared beliefs, perception, and attitudes towards driving, all of which might affect a driver's choice of whether to drive under the influence of fatigue.

A binary logistic regression model and a Deep Neural Network (DNN) were trained on identical data in order to classify and predict driver behaviour as driving while fatigued or not. The DNN slightly outperformed the binary logistic model, yielding a small gain in ROC-AUC and correct classifications. Judging from the binary model coefficients, drivers reporting driving under the influence of drugs, fatigue, or alcohol, as well as speeding, safety, and texting while driving were more likely to have recently driven while fatigued. Acceptability of fatigued driving was also found to be positively correlated with actual fatigued driving behaviour. In contrast, acceptability of other hazardous behaviours, namely mobile phone use and drunk driving, was negatively correlated with fatigued driving behaviour. More responsible driver perspectives, in the form of support for road safety policy measures or the perspective that drugs is a crash causation factor were found to have a negative influence on driving while fatigued.

While imperfect, the present model outcomes hint at new capabilities for fatigue detection. The present results can be expanded by creating international or internationally comparable national surveys that would be purpose-made to investigate the extent of fatigued driving in road networks. Extrapolations of numbers could formulate the basis for new policies or regulations to mitigate fatigued driving, road enforcement and effective awareness raising of all road users, while emphasis could be placed on the needs of fatigued-prone groups, such as long-haul truckers or other professional drivers.

CRedit authorship contribution statement

Alexandros Zoupos: Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **Apostolos Ziakopoulos:** Investigation, Data curation, Validation, Writing - original draft, Writing - review & editing. **George Yannis:** Conceptualization, Supervision, Writing - review & editing.

Declaration of competing interests

The authors report no competing interests.

Acknowledgement

This research was conducted as part of the second edition of the ESRA project (E-Survey of Road users' Attitudes), a joint initiative of road safety institutes, research organisations, public services, and private sponsors. The project was coordinated by VIAS Institute and funded

through the contributions of the partner organisations, either from their own resources or from sponsoring.

References

- Aloi, A., Alonso, B., Benavente, J., Cordera, R., Echániz, E., González, F., ... & Sañudo, R. (2020) *Effects of the COVID-19 lockdown on urban mobility: empirical evidence from the city of Santander (Spain)*. Sustainability, 12(9), p. 3870 <https://doi.org/10.3390/su12093870>
- Arnold, T. B. (2017) *kerasR: R interface to the keras deep learning library*. Journal of Open Source Software, 2(14), 296 <https://doi.org/10.21105/joss.00296>
- Backer-Grøndahl, A., & Sagberg, F. (2011) *Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones*. Safety Science, 49(2), pp. 324-330 <https://doi.org/10.1016/j.ssci.2010.09.009>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Grothendieck, G. (2011) *Package 'lme4'. Linear mixed-effects models using Eigen and Eigen++*. R package version, 1(6) <http://dk.archive.ubuntu.com/pub/pub/cran/web/packages/lme4/lme4.pdf>
- Beck, K.H., Lee, C.J. and Weiner, T. (2018) *Motivational factors associated with drowsy driving behavior: a qualitative investigation of college students*. Sleep Health, 4(1), pp.116-121 <https://doi.org/10.1016/j.sleh.2017.10.007>
- Dawson, C. W., & Wilby, R. (1998) *An artificial neural network approach to rainfall-runoff modelling*. Hydrological Sciences Journal, 43(1), pp. 47-66. <https://doi.org/10.1080/02626669809492102>
- Feldhütter, A., Gold, C., Schneider, S., & Bengler, K. (2016) *How the Duration of Automated Driving Influences Take-Over Performance and Gaze Behavior*. Advances in Ergonomic Design of Systems, Products and Processes, pp. 309–318 https://doi.org/10.1007/978-3-662-53305-5_22
- Feng, R., Zhang, G., & Cheng, B. (2009) *An on-board system for detecting driver drowsiness based on multi-sensor data fusion using Dempster-Shafer theory*. In 2009 International Conference on Networking, Sensing and Control (pp. 897-902). IEEE. <https://doi.org/10.1109/ICNSC.2009.4919399>
- Goldenbeld, C., Torfs, K., Vlakveld, W., & Houwing, S. (2020) *Impaired driving due to alcohol or drugs: International differences and determinants based on E-Survey of Road Users' Attitudes first-wave results in 32 countries*. IATSS Research, 44(3), pp. 188–196 <https://doi.org/10.1016/j.iatssr.2020.07.005>
- Gonçalves, J., Happee, R., & Bengler, K. (2016) *Drowsiness in conditional automation: proneness, diagnosis and driving performance effects*. In 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC) (pp. 873-878). IEEE. <https://doi.org/10.1109/ITSC.2016.7795658>
- Grossman, E.S., & Rosenbloom, T. (2016) *Perceived level of performance impairment caused by alcohol and restricted sleep*. Transportation Research Part F, 41: pp. 113–123 <https://doi.org/10.1016/j.trf.2016.06.002>
- Gulli, A., & Pal, S. (2017) *Deep learning with Keras*. Packt Publishing Ltd. ISBN: 978-1-78712-842-2.
- Hara, K., Saito, D., & Shouno, H. (2015) *Analysis of function of rectified linear unit used in deep learning*. In 2015 IEEE International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). <https://doi.org/10.1109/IJCNN.2015.7280578>
- Harrell, F. E. (2015) *Binary logistic regression*. In Harrell, F. E. *Regression modeling strategies* (pp. 219-274). Springer, Cham. ISBN: 978-3-319-19425-7. <https://doi.org/10.1007/978-3-319-19425-7>
- Hosmer DW, Lemeshow S. (1989) *Applied logistic regression*. New York: John Wiley & Son. ISBN: 9781118548387. <https://doi.org/10.1002/9781118548387>
- Jackson, P., Hilditch, C., Holmes, A., Reed, N., Merat, N. and Smith, L., (2011) *Fatigue and road safety: a critical analysis of recent evidence*. UK Department for Transport, Road Safety Web Publication, 21. ISBN: 978-1-84864-110-5.
- Khan M.I., Mansoor A.B. (2008) *Real Time Eyes Tracking and Classification for Driver Fatigue Detection*. In: Campilho A., Kamel M. (eds) *Image Analysis and Recognition*. ICIAR 2008. Lecture Notes in Computer Science, vol 5112. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69812-8_72
- Kong, W., Zhou, L., Wang, Y., Zhang, J., Liu, J. and Gao, S., (2015) *A system of driving fatigue detection based on machine vision and its application on smart device*. Journal of Sensors, 2015. <https://doi.org/10.1155/2015/548602>
- Li, R., Chen, Y. V., & Zhang, L. (2021) *A method for fatigue detection based on Driver's steering wheel grip*. International Journal of Industrial Ergonomics, 82, 103083. <https://doi.org/10.1016/j.ergon.2021.103083>
- Li, K., Gong, Y., & Ren, Z. (2020) *A fatigue driving detection algorithm based on facial multi-feature fusion*. IEEE Access, 8, pp. 101244-101259 <https://doi.org/10.1109/ACCESS.2020.2998363>

- Li, Z., Chen, L., Peng, J., & Wu, Y. (2017) *Automatic Detection of Driver Fatigue Using Driving Operation Information for Transportation Safety*. *Sensors*, 17(6), p. 1212 <https://doi.org/10.3390/s17061212>
- Meesmann, U., Torfs, K., Wardenier, N. & Van den Berghe, W. (2021) *ESRA2 methodology*. ESRA2 report Nr. 1 (updated version). ESRA project (E-Survey of Road users' Attitudes). Brussels, Belgium: Vias institute. Available: <https://www.esranet.eu/storage/minisites/esra2-methodology-report-updatewave2-def.pdf>
- Miyata, S., Noda, A., Ozaki, N., Hara, Y., Minoshima, M., Iwamoto, K., Takahashi, M., Iidaka, T., & Koike, Y. (2010) *Insufficient sleep impairs driving performance and cognitive function*. *Neuroscience letters*, 469(2), pp. 229–233 <https://doi.org/10.1016/j.neulet.2009.12.001>
- Nordbakke, S., & Sagberg, F. (2007) *Sleepy at the wheel: Knowledge, symptoms and behaviour among car drivers*. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(1), pp. 1–10 <https://doi.org/10.1016/j.trf.2006.03.003>
- Parsa, M.J., Javadi, M. and Mazinan, A.H. (2021) *Fatigue level detection using multivariate autoregressive exogenous nonlinear modeling based on driver body pressure distribution*. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, p. 09544070211014290 <https://doi.org/10.1177/09544070211014290>
- Pires, C., Torfs, K., Areal, A., Goldenbeld, C., Vanlaar, W., Granie, M.A., Stürmer, Y.A., Usami, D.S., Kaiser, S., Jankowska-Karpa, D. and Nikolaou, D. (2020) *Car drivers' road safety performance: A benchmark across 32 countries*. *IATSS research*, 44(3), pp.166-179 <https://doi.org/10.1016/j.iatssr.2020.08.002>
- R Core Team (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Radun, I., Radun, J., Wahde, M., Watling, C.N. and Kecklund, G. (2015) *Self-reported circumstances and consequences of driving while sleepy*. *Transportation Research Part F: Traffic Psychology And Behaviour*, 32, pp. 91-100 <https://doi.org/10.1016/j.trf.2015.05.004>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016) "Why should I trust you?" *Explaining the predictions of any classifier*. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135-1144 <https://doi.org/10.1145/2939672.2939778>
- Samarasinghe, S. (2006) *Neural Networks For Applied Sciences And Engineering: From Fundamentals To Complex Pattern Recognition*. CRC Press. ISBN 9780849333750
- Sayed, R., & Eskandarian, A. (2001) *Unobtrusive drowsiness detection by neural network learning of driver steering*. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 215(9), pp. 969–975 <https://doi.org/10.1243/0954407011528536>
- Schmidhuber, J. (2015) *Deep learning in neural networks: An overview*. *Neural networks*, 61, pp. 85-117 <https://doi.org/10.1016/j.neunet.2014.09.003>
- Sikander, G., & Anwar, S. (2018) *Driver fatigue detection systems: A review*. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), pp. 2339-2352 <https://doi.org/10.1109/TITS.2018.2868499>
- Singh, S. (2018) *Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey*. (*Traffic Safety Facts Crash Stats*. Report No. DOT HS 812 506). Washington, DC: U.S. National Highway Traffic Safety Administration.
- Singh, S. (2015) *Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey*. NHTSA Report (No. DOT HS 812 115).
- Smolensky, M. H., Di Milia, L., Ohayon, M. M., & Philip, P. (2011) *Sleep disorders, medical conditions, and road accident risk*. *Accident Analysis & Prevention*, 43(2), pp. 533–548 <https://doi.org/10.1016/j.aap.2009.12.004>
- Talbot, R., Filtness, A. (2017) *Fatigue – Not Enough Sleep/Driving While Tired*, European Road Safety Decision Support System, developed by the H2020 project SafetyCube. Retrieved from www.roadsafety-dss.eu on 11 June 2021
- Tefft, B.C. (2010) *Asleep at the wheel: The prevalence and impact of drowsy driving*. AAA Foundation for Traffic Safety. Available: <https://aaafoundation.org/wp-content/uploads/2018/02/2010DrowsyDrivingReport.pdf>
- Torfs, K., Meesmann, U., Van den Berghe, W., & Trotta, M. (2016) *ESRA 2015 – The results. Synthesis of the main findings from the ESRA survey in 17 countries*. ESRA project (European Survey of Road users' safety Attitudes). Brussels, Belgium. Available: <https://www.esranet.eu/storage/minisites/esra2017-en.pdf>
- Tranmer, M., & Elliot, M. (2008) *Binary logistic regression*. Cathie Marsh for census and survey research, paper, 20. Available: <http://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2008/2008-20-binary-logistic-regression.pdf>

- United States National Sleep Foundation (2013) *2013 International Bedroom Poll*:
<https://www.sleepfoundation.org/professionals/sleep-america-polls/2013-international-bedroom-poll>
 (Accessed 11 June 2021)
- United States Federal Motor Carrier Safety Administration (FMCSA) (2011) *Hours of service of drivers*. U.S. Federal Register, vol. 76(248)
- United States National Center for Statistics and Analysis (NCSA). (2020) *Overview of motor vehicle crashes in 2019*. (Traffic Safety Facts Research Note. Report No. DOT HS 813 060). National Highway Traffic Safety Administration (NHTSA).
- Vanlaar, W., Simpson, H., Mayhew, D., & Robertson, R. (2008) *Fatigued and drowsy driving: A survey of attitudes, opinions and behaviors*. Journal of Safety Research, 39(3), pp. 303-309
<https://doi.org/10.1016/j.jsr.2007.12.007>
- Wang, M. S., Jeong, N. T., Kim, K. S., Choi, S. B., Yang, S. M., You, S. H., ... & Suh, M. W. (2016) *Drowsy behavior detection based on driving information*. International Journal of Automotive Technology, 17(1), pp. 165-173. <https://doi.org/10.1007/s12239-016-0016-y>
- Wang, S.C. (2003) *Artificial Neural Network in Wang, S.C., Interdisciplinary Computing in Java Programming*, 81–100. ISBN: 978-1-4615-0377-4. https://doi.org/10.1007/978-1-4615-0377-4_5
- Washington, S., Karlaftis, M. G., Mannering, F., & Anastasopoulos, P. (2020) *Statistical and Econometric Methods for Transportation Data Analysis*. CRC Press. ISBN: 9780429244018.
<https://doi.org/10.1201/9780429244018>
- Watling, C.N., Armstrong, K.A., Obst, P.L. and Smith, S.S., (2014) *Continuing to drive while sleepy: The influence of sleepiness countermeasures, motivation for driving sleepy, and risk perception*. Accident Analysis & Prevention, 73, pp.262-268 <https://doi.org/10.1016/j.aap.2014.09.021>
- Watling, C.N. (2014) *Sleepy driving and pulling over for a rest: Investigating individual factors that contribute to these driving behaviours*. Personality and Individual Differences, 56, pp. 105-110
<https://doi.org/10.1016/j.paid.2013.08.031>
- Wilson, R.J., Fang, M., Cooper, P.J. and Beirness, D.J. (2006) *Sleepiness among night-time drivers: relationship to blood alcohol concentration and other factors*. Traffic Injury Prevention, 7(1), pp. 15-22
<https://doi.org/10.1080/15389580500412929>
- World Health Organization – WHO. (2018) *Global status report on road safety 2018*. ISBN: 9789241565684. Available: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/
- Yannis, G., Nikolaou, D., Laiou, A., Stürmer, Y. A., Buttler, I., & Jankowska-Karpa, D. (2020) *Vulnerable road users: Cross-cultural perspectives on performance and attitudes*. IATSS Research, 44(3), pp. 220–229
<https://doi.org/10.1016/j.iatssr.2020.08.006>
- Zhang, C., Wang, H., & Fu, R. (2014) *Automated Detection of Driver Fatigue Based on Entropy and Complexity Measures*. IEEE Transactions on Intelligent Transportation Systems, 15(1), pp. 168–177
<https://doi.org/10.1109/tits.2013.2275192>
- Zhang, G., Yau, K. K., Zhang, X., & Li, Y. (2016) *Traffic accidents involving fatigue driving and their extent of casualties*. Accident Analysis & Prevention, 87, pp. 34-42 <https://doi.org/10.1016/j.aap.2015.10.033>
- Zhou, F., Alsaid, A., Blommer, M., Curry, R., Swaminathan, R., Kochhar, D., Talamonti, W., Tijerina, L., & Lei, B. (2020) *Driver fatigue transition prediction in highly automated driving using physiological features*. Expert Systems with Applications, 147, p. 113204 <https://doi.org/10.1016/j.eswa.2020.113204>
- Zou, X., Yue, W. L., & Le Vu, H. (2018) *Visualization and analysis of mapping knowledge domain of road safety studies*. Accident Analysis & Prevention, 118, pp. 131-145 <https://doi.org/10.1016/j.aap.2018.06.010>
- Ziakopoulos, A., Theofilatos, A., Laiou, A., Michelaraki, E., Yannis, G., & Rosenbloom, T. (2021a) *Examining the relationship between impaired driving and past crash involvement in Europe: Insights from the ESRA study*. International Journal of Injury Control and Safety Promotion, pp. 1–11
<https://doi.org/10.1080/17457300.2021.1928234>
- Ziakopoulos, A., Nikolaou, D., & Yannis, G. (2021b) *Correlations of multiple rider behaviors with self-reported attitudes, perspectives on traffic rule strictness and social desirability*. Transportation Research Part F: Traffic Psychology and Behaviour, 80, pp. 313–327 <https://doi.org/10.1016/j.trf.2021.05.011>

About the authors



Alexandros Zoupos is currently studying Electronics and Electrical Engineering at The University of Edinburgh. He was a two-month intern at the National Technical University of Athens at the Department of Transportation Planning and Engineering at the School of Civil Engineering. His interests include Data Analysis research in R, as well as scientific programming in Python. Recently, he has engaged in machine learning programming.



Dr. **Apostolos Ziakopoulos**, PhD, MSc DIC, is a Research Associate at the Department of Transportation Planning and Engineering at the School of Civil Engineering of the National Technical University of Athens (NTUA). He holds a Civil Engineering Diploma from NTUA majoring in Transportation Engineering (2013). He holds a Master of Science – DIC in Transport from Imperial College London and University College London (2014). His Ph.D. dissertation was in spatial analysis of road safety and traffic behavior from NTUA (2020). His main research interests involve road crash analyses, statistical and spatial modelling and traffic engineering. He has participated in 8 research and engineering projects and has published 55 papers in scientific journals, textbooks and conferences.



Professor **George Yannis** is a Full Professor in Traffic Safety and Management with particular focus on data management and analysis at the Department of Transportation Planning and Engineering of the School of Civil Engineering of the National Technical University of Athens. He has contributed extensively in more than 245 research and engineering projects and studies and in several scientific committees of the European Commission and other International Organizations (UN/ECE, OECD, WHO, World Bank, EIB, CEDR, ERF, UITP, ETSC, ECTRI, WCTR, TRB). He has published more than 640 scientific papers (more than 200 in scientific journals) widely cited worldwide.



All contents are licensed under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).