International ESRA Conference

Traffic Safety Culture and Performance Indicators

April 21, 2022, ONLINE

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

Nikolaou D., Zoupos A., Ziakopoulos A., Yannis G.
National Technical University of Athens

Université Gustave Eiffel

ESRA
www.esranet.eu
E-Survey of Road users' Attitudes

Presentation outline

- Background
- Objective
- ESRA2 data
- Descriptive Statistics
- Statistical Background
- Statistical Modelling Results
- Concluding Remarks

Background (1/2)

- 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).
- 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).
- Unlike other driving impairments, driver fatigue is very **hard to measure.**

Background (2/2)

- Several studies have addressed the problem of fatigue by aiming to detect it in drivers based on different features of **driver behavior** such as steering wheel angle, yaw angles, lateral and longitudinal accelerations of the vehicle.

- Apart from methods based purely on driving behavioral features, there have been studies that managed to detect fatigued driving using **physiological features** such as the response of the eyes.

- 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 and Anwar, 2018).

Objective

- The design of this study aims to **exploit a broad questionnaire sample** in lieu of driver simulators and eye-tracking devices, in order to:

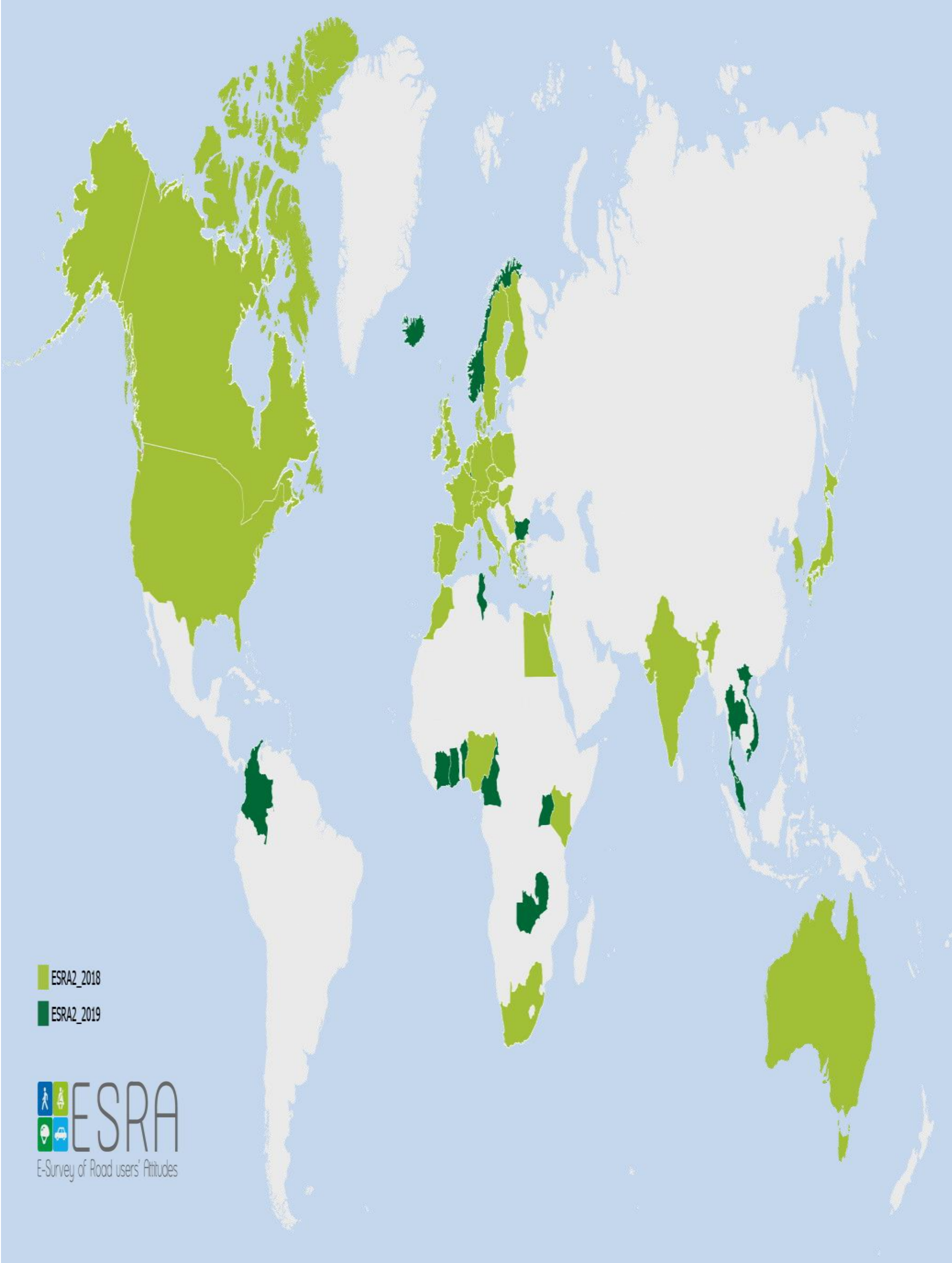
  i. gain a wider perspective from a large sample consisting of participants from 47 countries, which would not be feasible for an instrumented experiment,

  ii. and to explore the feasibility and usefulness of using questionnaire data for the prediction of driving while fatigued amongst driver samples.

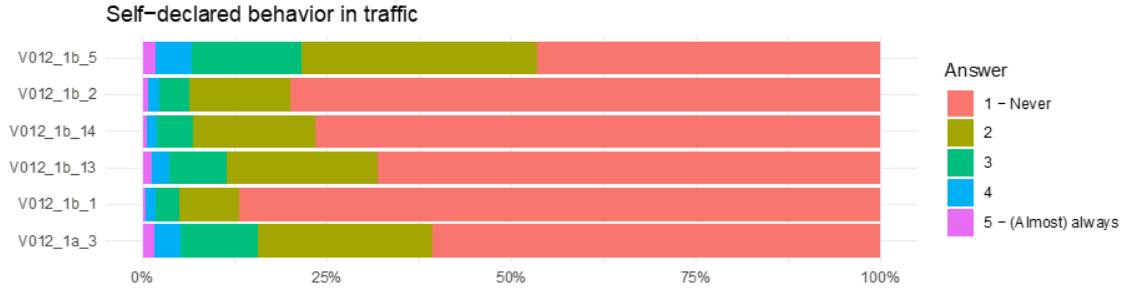
- Specifically, an examination of how various self-declared beliefs, perceptions and attitudes can **influence road user choices** on whether to drive under the influence of fatigue is conducted.

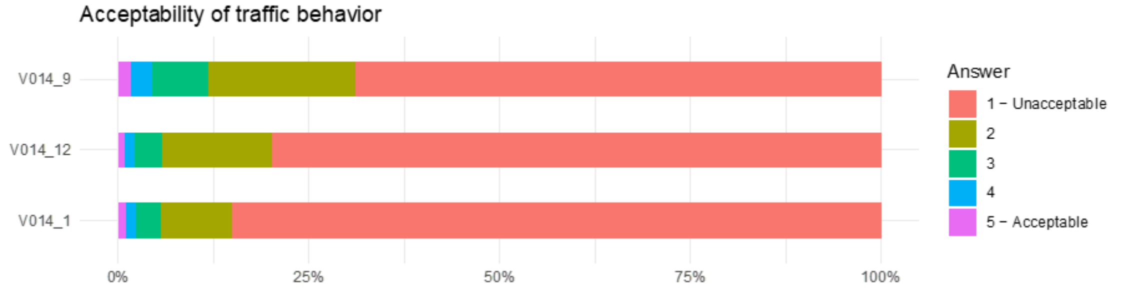
ESRA2 data

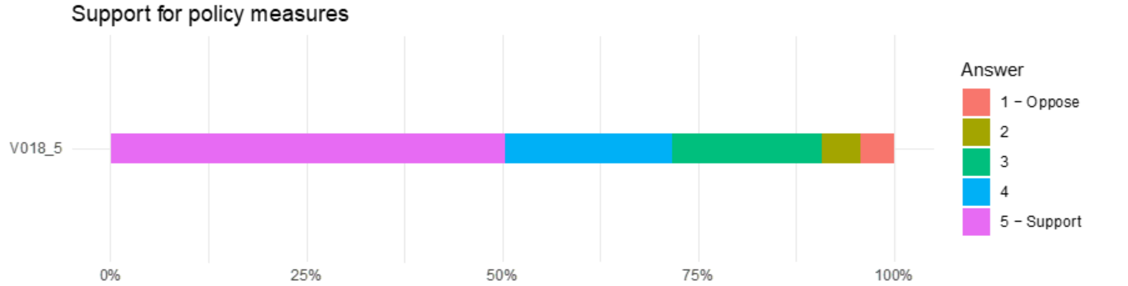
- The ESRA2 online survey provides a dataset from many countries around the world, for all road users. The total sample size was **45,000 road users across 48 countries** (men: 49.6%, women: 50.1%, other: 0.3%).
- From this dataset, **30,683 participants were regular car drivers**, defined as a person who uses their car a few days a month or more.
- In this study, data from Iceland had to be discarded due to methodological differences, translation difficulties and other barriers, therefore **47 countries** remained in total.
- 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.

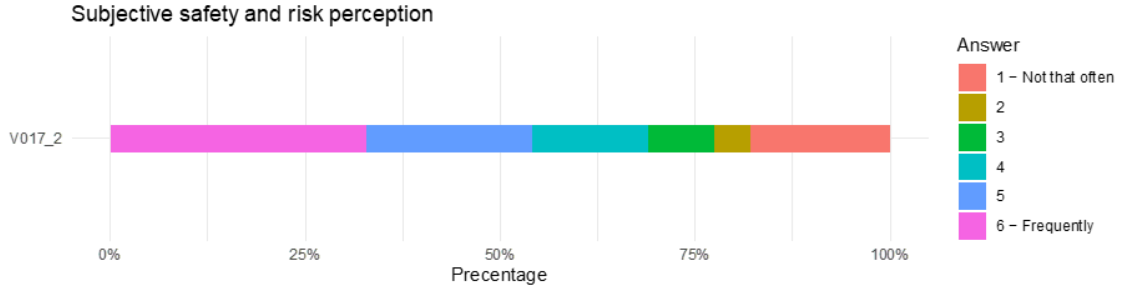


Descriptive Statistics









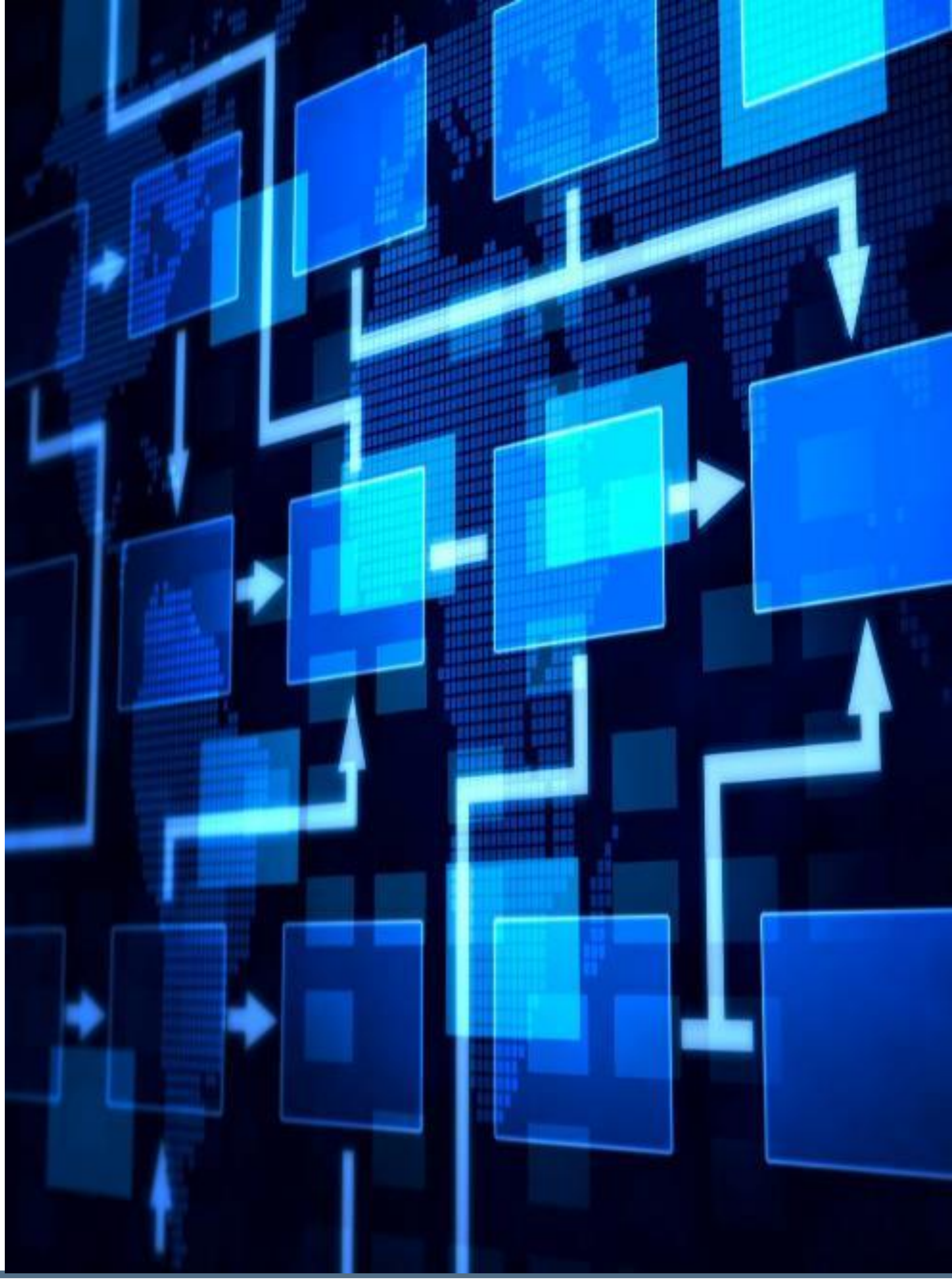
Abbreviation	Question	Scale
	Self-declared behavior in traffic	
V012_1b_14	Over the last 30 days, how often did you as a CAR DRIVER drive when you were so sleepy that you had trouble keeping your eyes open?	At least once (2-5) – never (1)
V012_1b_8	Over the last 30 days, how often did you as a CAR DRIVER drive without wearing your seatbelt?	At least once (2-5) – never (1)
V012_1b_13	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?	At least once (2-5) – never (1)
V012_1b_2	Over the last 30 days, how often did you as a CAR DRIVER drive after drinking alcohol?	At least once (2-5) – never (1)
V012_1b_9	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)?	At least once (2-5) – never (1)
V012_1b_7	Over the last 30 days, how often did you as a CAR DRIVER drive faster than the speed limit on motorways/freeways?	At least once (2-5) – never (1)
V012_1b_4	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?	At least once (2-5) – never (1)
V012_1b_1	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?	At least once (2-5) – never (1)
V012_1b_5	Over the last 30 days, how often did you as a CAR DRIVER drive faster than the speed limit inside built-up areas?	At least once (2-5) – never (1)
V012_1a_3	Over the last 12 months, how often did you as a CAR DRIVER read a text message or email while driving?	At least once (2-5) – never (1)
	Acceptability of traffic behavior	
V014_1	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?	Acceptable (4-5) – unacceptable/neutral (1-3)
V014_9	How acceptable do you, personally, feel it is for a CAR DRIVER to talk on a hand-held mobile phone while driving?	Acceptable (4-5) – unacceptable/neutral (1-3)
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?	Acceptable (4-5) – unacceptable/neutral (1-3)
	Support for policy measures	
V018_5	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)?	Support (4-5) – oppose/neutral (1-3)
	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)

Statistical background

- Within the ESRA2 questionnaire, the variable on fatigued driving 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).**



Binary Logistic Regression (1/2)

- Binary logistic regression models are well established statistical functional-based methods, widely applied in order to model **binary outcomes**.

- The model begins by **considering a linear predictor**, as expressed by:

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

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

Binary Logistic Regression (2/2)

- If a **utility function** is considered by expressing the predictor without considering the error term, as given by:

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

- Then the **probability P** that the dependent variable belongs in a specific class is given by:

$$P = \frac{e^U}{e^U + 1}$$

- **Hosmer and Lemeshow Test** was used s a test for goodness of fit of the model and model selection process between models including different independent variable subsets is conducted by the examination of the corrected **Akaike Information Criterion** (AICc).

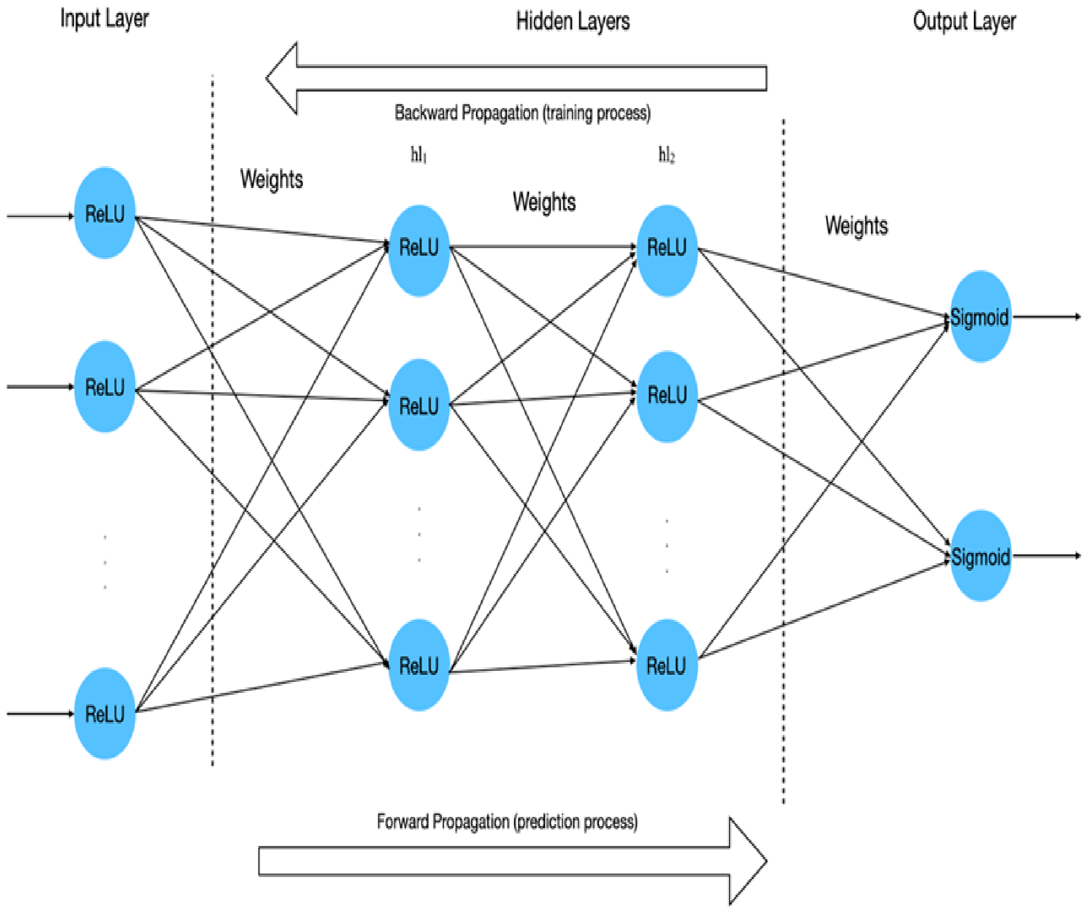
Deep Neural Network (DNN) (1/2)

- 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 functions used in this study were the **Rectified Linear Unit (ReLU) and the Sigmoid**.

- The training of the DNN (**backward propagation**) is performed by calculating the weights using the training set. Using the complete, trained DNN, data from the test set are used as input and predictions are obtained from the output layer (**forward propagation**).



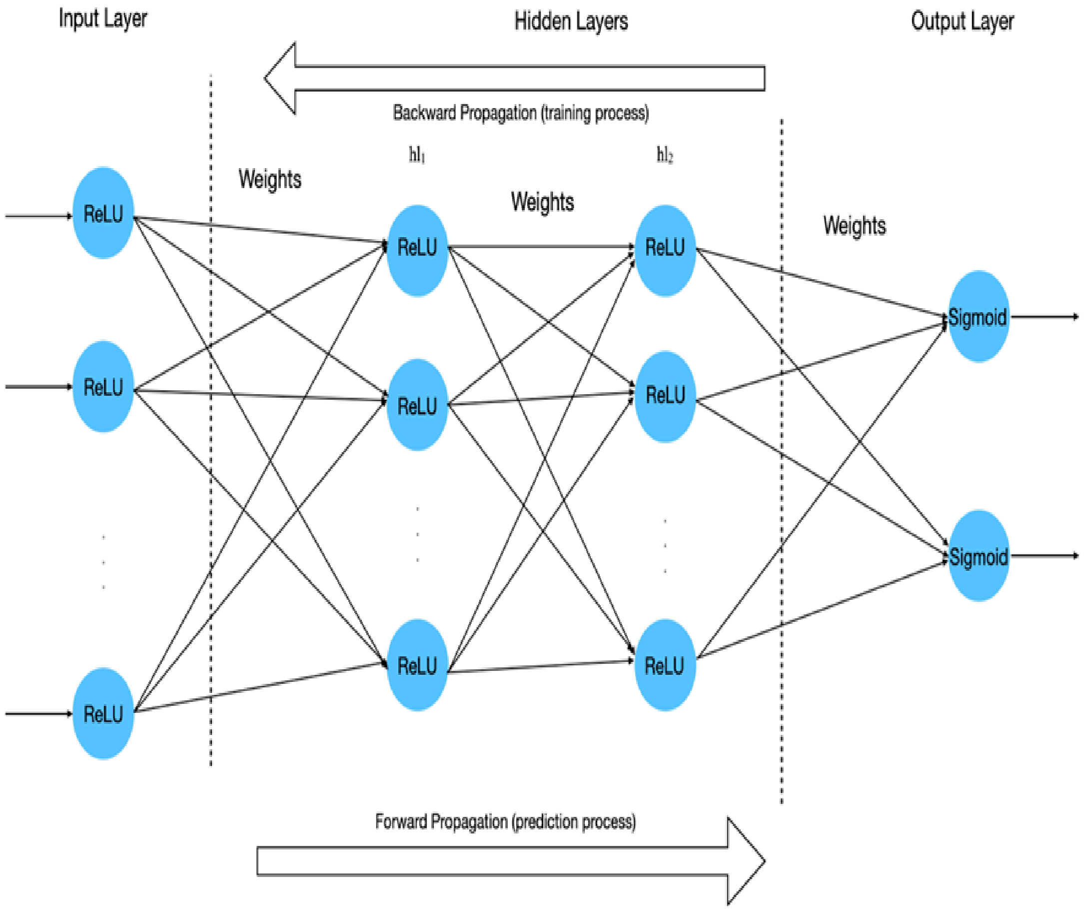
Deep Neural Network (DNN) (2/2)

- 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).

- 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**.

- 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.



Comparative model performance evaluation

- 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.

Models development

- The outcome variable for the models referred to **fatigued driving**. 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).

- 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)).

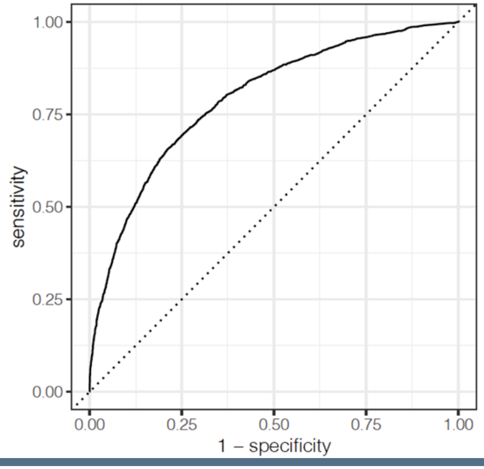
- As per standard good practice, models were **trained in a subset** of the entire usable dataset, while their predictions were tested in data which was completely new for the model.

- 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.

Binary Logistic Regression results (1/2)

- The ROC curve produced by the binary logistic regression model had an **AUC of 0.793**.
- The correct classifications (TP+TN) amount to **79.0%**.
- All of the variables in the category of "Self-declared behavior" (DUI of drugs or alcohol, speeding, non-seatbelt use, texting while driving) are **positively correlated** with recent fatigued driving.
- 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?" suggesting that **the higher personal acceptability of driving while fatigued leads to more likely engagement** in such behavior.
- Acceptability of 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.

Independent variable	Coefficients			
	Beta Estimate	Std. Error	z value	p-value
Intercept	*-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***
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***



	Actual value	
Predicted value	0	1
0	*67.0 %*	*10.8 %*
1	*10.3 %*	*12.0 %*

Binary Logistic Regression results (2/2)

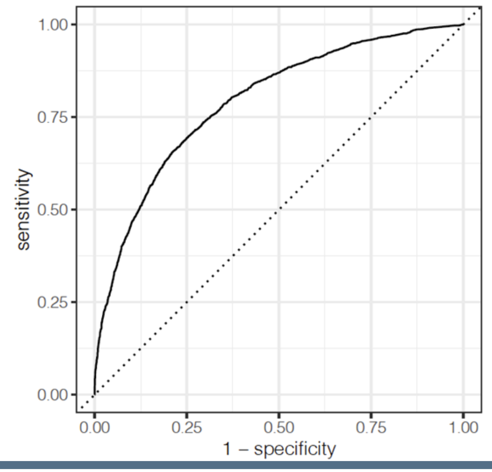
- Supporting legal obligation to install Dynamic Speed Warning signs is **negatively correlated** with recent fatigued driving.

- The respective question of the category “Subjective safety and risk perception” 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.

Independent variable	Coefficients			
	Beta Estimate	Std. Error	z value	p-value
Intercept	*-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***
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***

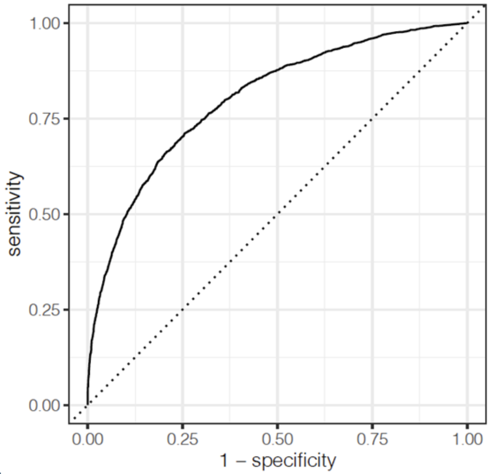


	Actual value	
Predicted value	0	1
0	*67.0 %*	*10.8 %*
1	*10.3 %*	*12.0 %*

Deep Neural Network results (1/2)

- 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**.
- The ROC curve produced by the DNN model had an **AUC of 0.801**.
- The run had an accuracy of 0.815, a loss of 0.423, and **MSE of 0.134**.
- The correct classifications (TP+TN) amount to **80.1%.**

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

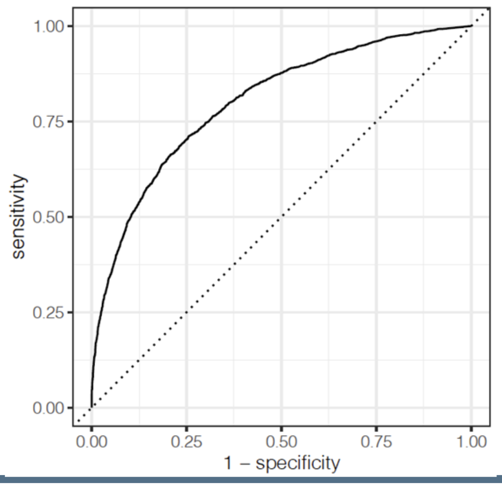


	Actual values	
Predicted values	**0**	**1**
0	*67.6%*	*10.4%*
1	*9.5%*	*12.5%*

Deep Neural Network results (2/2)

- 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, there is a considerable imbalance of the negative against positive cases for these variables, and thus their contributions are less straightforward to interpret.

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



	Actual values	
Predicted values	**0**	**1**
0	*67.6%*	*10.4%*
1	*9.5%*	*12.5%*

Concluding remarks (1/2)

- 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. 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.

Concluding remarks (2/2)

- By covering a wide range of topics, from support to legal measures to self-declared behavior, it was possible to detect driver behaviors towards fatigued driving to an **adequate degree.**

- While imperfect, the present model outcomes hint at **new capabilities** for fatigue detection.

- **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.

International ESRA Conference

Traffic Safety Culture and Performance Indicators

April 21, 2022, ONLINE

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

Nikolaou D., Zoupos A., Ziakopoulos A., Yannis G.

National Technical University of Athens

Université Gustave Eiffel

www.esranet.eu

ESRA

E-Survey of Road users' Attitudes