Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting depression from driving data using machine learning algorithms

Dimitris Bobos

dbobos@uoa.gr

National and Kapodistrian University of Athens

Vagioula Tsoutsi

National and Kapodistrian University of Athens

Panagiotis Kazanis

Giannis Manesis

Evangelos Grinakis

University Hospital of Heraklion

Vasilios G. Masdrakis

National and Kapodistrian University of Athens

George Yannis

National Technical University of Athens

Dimitrios Vogiatzis

National Centre of Scientific Research "Demokritos"

Dimitris Dikeos

National and Kapodistrian University of Athens

Research Article

Keywords: Depression, Driving simulator, Machine learning, Time series analysis

Posted Date: September 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10660130/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title

Predicting depression from driving data using machine learning algorithms

Authors: D. Bobos (corresponding author), V. Tsoutsi, G. Manesis, P. Kazanis, E. Grinakis, V. Masdrakis, G. Yannis, D. Dikeos, D. Vogiatzis

Abstract

Novel machine learning algorithms can successfully analyze raw big data, such as the ones derived from driving simulators which record a plethora of parameters up to around 20 times per second. The aim of the present study is to provide proof of concept of applying machine learning algorithms on driving simulator data of patients suffering from depression. Driving data from patients (N=39) and control participants (N=30) included speed, distance from leading vehicle, lateral position, acceleration, deceleration, steering wheel movement, orientation of the car and brake pedal position. We integrated differential analysis with supervised machine learning classifiers and time series analysis to detect depression based on the driving data. The models created were evaluated using accuracy, precision, recall, and F1-score. In addition, the area under the receiver operating characteristic curve (AUC-ROC) was calculated to assess the discriminative performance of the models. To address the imbalance in the class distribution of participants, a penalized Support Vector Classifier was employed. Decision Tree Classification and Random Forest Classification were used on the consolidated driving data, and XGBoost Classifier and Random Forest Classifier on the time series. The consolidated test sample Random Forest Classifier metrics for predicting depression were 0.74, 0.76, 0.74 and 0.75. Using time series data, the respective metrics were 0.61, 0.66, 0.61 and 0.63; accuracy reached 0.79 when using feature-importance thresholds in XGBoost. The consolidated Random Forest Classifier results act as proof of concept of the ability of machine learning models to distinguish patient from control participants based on driving behavior.

Keywords

Depression

Driving simulator

Machine learning

Time series analysis

1. Introduction

Road traffic accidents are a leading cause of death and a major public health issue. Driving under a compromised state such as depression, which is characterized by executive, mental and emotional deficits, might have an adverse effect on driving behavior. According to the World Health Organization (WHO) road traffic accidents are the eighth leading cause of death, but by 2030 they will be the fifth (World Health Organization 2009). Such accidents can be detrimental to individuals who are seriously injured by them as well as family members of the injured and killed. The leading factor of road traffic accidents is the driver; more than 90% of crashes are attributed to human error (National Highway Traffic Safety Administration 2008). Previous studies on road traffic accidents have examined mainly the effects of the use of drugs such as anti-depressants and benzodiazepines (Brunnauer et al. 2015; Van der Sluiszen et al. 2017) on driving but not so much of the mental health disorders themselves. Such studies have collected driving data using cars equipped with cameras and sensors or by recording the controller area network (CAN) of the car (Fugiglando et al. 2019) or by using driving simulators. Driving simulators have been used to create various driving scenarios, which can simulate driver response to different driving conditions, such as heavy traffic, adverse weather, unfavorable road conditions and unexpected events. Driving simulators are preferable due to their accessibility and affordability and are often used to test the ability of a driver in professional settings, and in research on driving safety and driving ability under the influence of fatigue, alcohol and other substances (Rudin-Brown et al. 2009; Schreier et al. 2018); they also have been shown to successfully reflect similar results to real driving in assessing driving ability (Philip et al. 2005).

Depression has a global prevalence of 4.7% in the general population and is characterized by mental, emotional, and executive dysfunction (Ferrari et al. 2012; WHO 2017). The relationship between driving ability and depression has been examined in previous studies (Aduen et al. 2015; Alavi et al. 2017; Brunnauer et al. 2008; Brunnauer et al. 2015; Buckley et al. 2015, 2017; Bulmash et al. 2006; Ferrari et al. 2012; Liang and Yang 2022; Perrier et al. 2014; Shen et al. 2009; Tsoutsi et al. 2019, 2023; Van der Sluiszen et al. 2017; WHO 2017; Wickens et al. 2014; Wingen et al. 2006). A higher risk of road traffic accidents has been shown to be associated with compromised driving ability due to depression through forensic analysis of recorded road traffic accidents (Alavi et al. 2017; Wickens et al. 2014). Many studies have shown the association between depression and compromised driving ability; however, the findings are contradictory to each other. Some studies have shown that there is a relationship between depression and declined driving ability (Brunnauer et al. 2008; Shen et al. 2009), and others have not (Buckley et al. 2015, 2017; Liang and Yang 2022; Tsoutsi et al. 2023).

Advances in technology, intelligent transportation systems and vehicular technologies have spawned new innovative concepts, which have not been previously investigated. Sources such as driving simulators and autonomous driving provide a plethora of data, which can be used – based on powerful machine learning algorithms – to evaluate driving capabilities and, most significantly, to ensure the safety and comfort of drivers. Due to the novelty of the field, there has been limited evidence in the literature on how to utilize micro-driving data from driver simulators using machine learning methods to provide insight on driving behaviors (Birchler et al. 2023; Elassad et al. 2020)

and no such analysis has been conducted on samples of patients with psychiatric disorders such as depression.

This paper aims to provide a further understanding of the relationship between driving behavior and depression by employing machine learning algorithms, through which we aim to detect distinct driver behavior characteristics distinguishing depressed individuals from healthy controls. Identifying driving patterns in depressed individuals could be significant in ascertaining a level of driving ability in drivers suffering from depression and other mental disorders and developing informed guidelines for healthcare users and transportation authorities to use (Afghari et al. 2022; Buckley et al. 2017; WHO 2025).

The structure of this paper is as follows: The Methodology section contains the description of data acquisition and of the driving simulators which have been used, the presentation of the data obtained, and the statistical analysis methods employed to analyze the data. In the Results section, the sample characteristics of the participants are presented, followed by analysis of the data obtained from the driving simulators, both using the aggregated approach of data and time series data. In a similar format, the Discussion first refers to the raw driving data, and then to the findings derived after analysis by the machine learning algorithms. The paper closes with the Conclusion and Future Work of the study, an overall conclusion and possibilities for future work.

2. Methodology

2.1 Data Acquisition

Participants with a currently valid driver's license, were recruited within the scope of the PhD thesis of one of the authors (Vagioula Tsoutsi). All patient participants enrolled in the study had been diagnosed with Major Depressive Disorder by a psychiatrist. Healthy participants were used as controls. Informed consent was obtained from both patients and control participants, and the study had the approval of the Ethics Committee of Eginition Hospital, Athens, Greece (464/5-7-2019) and of the University Hospital of Heraklion, Crete, Greece (798/20/25-7-2018).

2.2 Driving Simulators

Two different driving simulators were used to obtain driving data for the participants. One of the two simulators used was the Virage Driving Simulator "Virage VS500M", which was set at the Laboratory of Health and Road Safety (LaHeRS), Department of Social Work, Hellenic Mediterranean University (HMU) in Heraklion, Crete, Greece, and the second was the Foerst Driving Simulator "FPF F10P", which was set at the Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University (NTU) of Athens, Greece. Both simulators used an open cabin equipped with a steering wheel, gearshift and pedals, dashboard, light switches, steering wheel with horn, ignition, and window wipers. The simulators were created using real car parts, which would more closely simulate the experience of being in a

real car. In addition, to mimic the external validity of a driving environment, the simulators used graphics for a 180-degree field of view and cameras for the mirrors, driving sounds and movement vibrations.

Different driving scenarios were used in the simulators. In the Virage Driving Simulator "Virage VS500M", participants drove in a motorway environment and an urban environment with low traffic, whereas in the Foerst Driving Simulator "FPF F10P", participants drove in a motorway environment with low traffic, urban environment with high and low traffic and rural environment with high and low traffic.

2.3 Data Pre-processing

Each driving scenario's recording average duration was about 5-6 minutes. Since the simulator recorded data with a mean rate of 60 times per second for the Foerst Driving Simulator "FPF F10P" and about 8.5 times per second for the Virage Driving Simulator "Virage VS500M" (each successive row of data comes about every 17 ms or 120ms, respectively, after the previous one), there were about 30-40 thousand rows of data per participant and driving scenario on the Foerst simulator and 2-3 thousand rows of data on the Virage simulator. Following data acquisition, in order to control for variability between the two driving simulators, each data subset was normalized and converted to be comparable between the simulators. Descriptive statistics were obtained, and any outliers or missing data were removed prior to analysis.

2.4 Feature Engineering

Participant means and standard deviations values were used as statistics of importance for participants, in the aggregated model approach.

For the extraction of acceleration and deceleration values from the data acquired, using the speed and time columns, the rate of change of speed was calculated, with the assumption that in the 17ms or 120ms time intervals (for the Foerst and Virage simulator, respectively) the car is moving in a straight line, performing a uniformly accelerated motion.

The lane offset had to be calculated based on the x,y coordinates of the track, which were derived by the participants from the control group to calculate the mean values of x and y relative positions in bins of cumulative distance. It was assumed that the (route_x_i, route_y_i) pair of coordinates represent the center of the lane along the bin i. By considering the same cumulative distance bins and by examining the mean values of x and y of each driver i.e. (mean_x_i, mean_y_i), we calculated the mean lane offset of each participant in each one of these bins, using:

- either the Pythagorean Theorem when:

$$\left(mean_{x_i} \neq route_{xi}\right) AND \left(mean_{yi} \neq route_{yi}\right)$$

- or a subtraction when at least one of the participant's coordinates was equal to the respective route coordinate, i.e. in case of route_x_i = mean_x_i the lane offset was calculated as the absolute value of route_y_i - mean_y_i and vice versa.

2.5 Performance evaluation

The performance of the predictive models was evaluated using accuracy, precision, recall, and the F1-score, the harmonic mean of precision and recall. In addition, the area under the curve (AUC) of the receiver operating characteristic (ROC) statistics (Fawcett 2006) was calculated to assess the discriminative performance of the models. To address the imbalance between control and patient participants, a penalized Support Vector Classifier was additionally employed, which acted as a baseline against which the models were compared.

2.6 Modeling on aggregated data

The modeling on aggregated data used the raw data from each participant from the Foerst Driving Simulator "FPF F10P" and the Virage Driving Simulator "VS500M". Means and standard deviations were calculated for each participant based on their motorway driving data and urban low traffic data, chosen since these two driving conditions were derived from both driving simulators used. Each participant was represented as a vector point value based on those statistics. Half the study participants randomly selected from both groups (control and patient) were used to train the model (N = 34) and the rest (N = 35) were used to test it.

The purpose of the aggregated data set approach was to create vectors for the control participants and the patient participants. Those vectors were created to identify each participant driving characteristics based on the driving simulator data, with the aim being able to predict drivers suffering from depression based on total aggregated model datasets. Through those vectors, a statistical table was created to distinguish the driving characteristics of the control and patient participants as seen in table 1. The control and patient participant data vectors were used to train and test models which would predict whether a participant was in the healthy control subset or in the patient subset.

The models used were the Decision Tree Classifier, the Random Forest Classifier and the Support Vector Classifier, which was used as a threshold of performance for the model. Those models were used due to their acceptable performance in small data sets when tuned correctly. The model was evaluated based on the ROC statistics obtained and on Support Vector Classifier which used a penalizing algorithm.

The settings used for the Decision Tree Classifier aimed to avoid issues of high bias/overfitting of the majority class. Thus, the splitting criterion of the nodes of the Decision Tree was Gini. Gini is a measurement of impurity of the values in the dataset. The Decision Tree created from the Gini measurement aims to create a tree which has the lowest possible feature as the root node and decreasing it towards the leaf nodes. The Gini measurement derived from AUC measures the power of a factor to distinguish between classes. Gini is measured by subtracting the sum of the squared probabilities for each occurring class, from one.

$$Gini = 1 - \sum_{i=1}^{n} (p_i)^2$$

The max depth of the tree created was set to 5, meaning that only 5 splits can occur after the root node.

The Random Forest is an ensemble classifier which uses the output of several Decision Trees to reach a single output. The settings used for the Random Forest Classifier aimed to avoid overfitting and balancing the uneven classes.

We used 200 estimators, meaning that 200 'trees' were created on our Random Forest, which were used to predict the outcome of the classifier. Each tree had a maximum depth of 2, in order to avoid overfitting the model. In addition, class weight was set to "balanced" which would adjust the weights inversely proportional to the number of participants per class.

2.7 Modeling on tabular time series data

The raw data used for the model creation of the time series approach included: time, cumulative distance, speed, acceleration, gear, brake, steering wheel movement, headway (which represents the distance from leading vehicle), lane offset (which shows how far the driver is from the center of the lane) and the unexpected events that occurred during the course, which was split in before events, during events and between events.

Time series data modeling enables us to create predictive models based on the location and timing of the participants. Modeling on time series data used the raw data from each participant from the Foerst Driving Simulator "FPF F10P" on the urban low traffic simulation. The aim of the tabular time series data approach is to predict drivers suffering from depression throughout the driving simulator path of the participants.

The time series data of each participant were processed using noise reduction and aggregation every 100 rows of raw data; the simulator recorded data every 17ms and was aggregated to data every 1.7 seconds due to computational restrictions, treating each time point in the series as an independent tabular observation, classifying every time point separately, and then aggregating those per-point predictions into a total prediction. Each participant's time series data were tested using the models of Random Forest Classifier, which makes use of the bagging algorithm to create new training subsets from the original one, and the XGBoost Classifier (Chen and Guestrin 2016). Two control participants' data were omitted, as they crashed in their course events. Participants of the Foerst Driving Simulator were split as 70% training (for the XGBoost analysis the training set was composed of 40% controls and 60% patients; the Random Forest split was not stratified) and 30% testing. We used 250 estimators in both classifiers to predict the outcome. For the Random Forest Classifier, each tree was grown without a depth limit using the Gini criterion, and class weight was set to "balanced_subsample", which adjusts the weights inversely proportionally to the class frequencies within each subsample. The Random Forest Classifier is able to use continuous data for classification, such as time series data. The SMOTE-ENN under-sampling method was used to balance the sample. SMOTE-ENN is the ensemble method of Synthetic Minority Sampling (SMOTE) and Edited Nearest Neighbor under-sampling (ENN) (Batista et al. 2004). The SMOTE oversampling technique chooses a random minority class instance, and through the use of k nearest

neighbors, an instance of another minority class is found. Then a synthetic, new instance of the minority class is created between those two points (Chawla et al. 2002). ENN is a hybrid technique where a number of observations are removed from the sample. In addition, the nearest neighbor of each of the majority class is estimated, and if it is misclassified, then the instance at which the misclassification occurred is deleted. (Lemaître et al. 2016; Nishat et al. 2022).

The XGBoost Classifier utilizes gradient boosting, a technique with which new models were created to predict the residuals or errors of the prior ones. Boosting uses trees with fewer splits than bagging approaches like Random Forest, which grow trees to the fullest extent possible. These small trees have little depth and are quite interpretable. An excessive quantity of trees may cause overfitting. The sampling method used for the model was the equalization of the classes as samples, meaning that we took an equal amount of the majority and minority class in the model training. The parameters of the XGBoost Classifier creation which were chosen are: gbtree as the booster, which uses tree-based models, with binary:logistic as the objective, since the task is essentially a binary classification between patients with depression and control subjects. As evaluation metric, the area under the ROC curve (AUC) was chosen. Moreover, approx. was used, which approximates a greedy algorithm using sketching before building a tree using all the rows; it was utilized as the tree method, for computational power efficiency purposes since we were dealing with quite large datasets. The values of the rest of the parameters, such as the number of trees or iterations, the rate at which the gradient boosting learns, and the depth of the tree, were selected and optimized by running multiple iterations and comparing results on our train set of participants, to create a model with predictive abilities. From the XGBoost classification, we obtained F-scores of feature importance for the 12 factors used in our model. Those F-scores were then used as a model threshold to optimize our model.

3. Results

3.1 Participant characteristics

In this dataset, we have data from Athens and Crete driving simulations (Foerst and Virage, respectively). In the Athens dataset we had 12 control and 26 patient participants. In the Crete dataset we had 18 control and 13 patient participants. This means that we have an unbalanced dataset of 30 control and 39 patient participants.

In this study 65.2% of the participants were women; 66.7% of the patients and 63.3% of the controls. The mean age of the participants was 45.22 with the youngest being 24 and the oldest being 74. The mean age of the control group was 42.93 and 46.97 of the patient group.

3.2 Driving characteristics

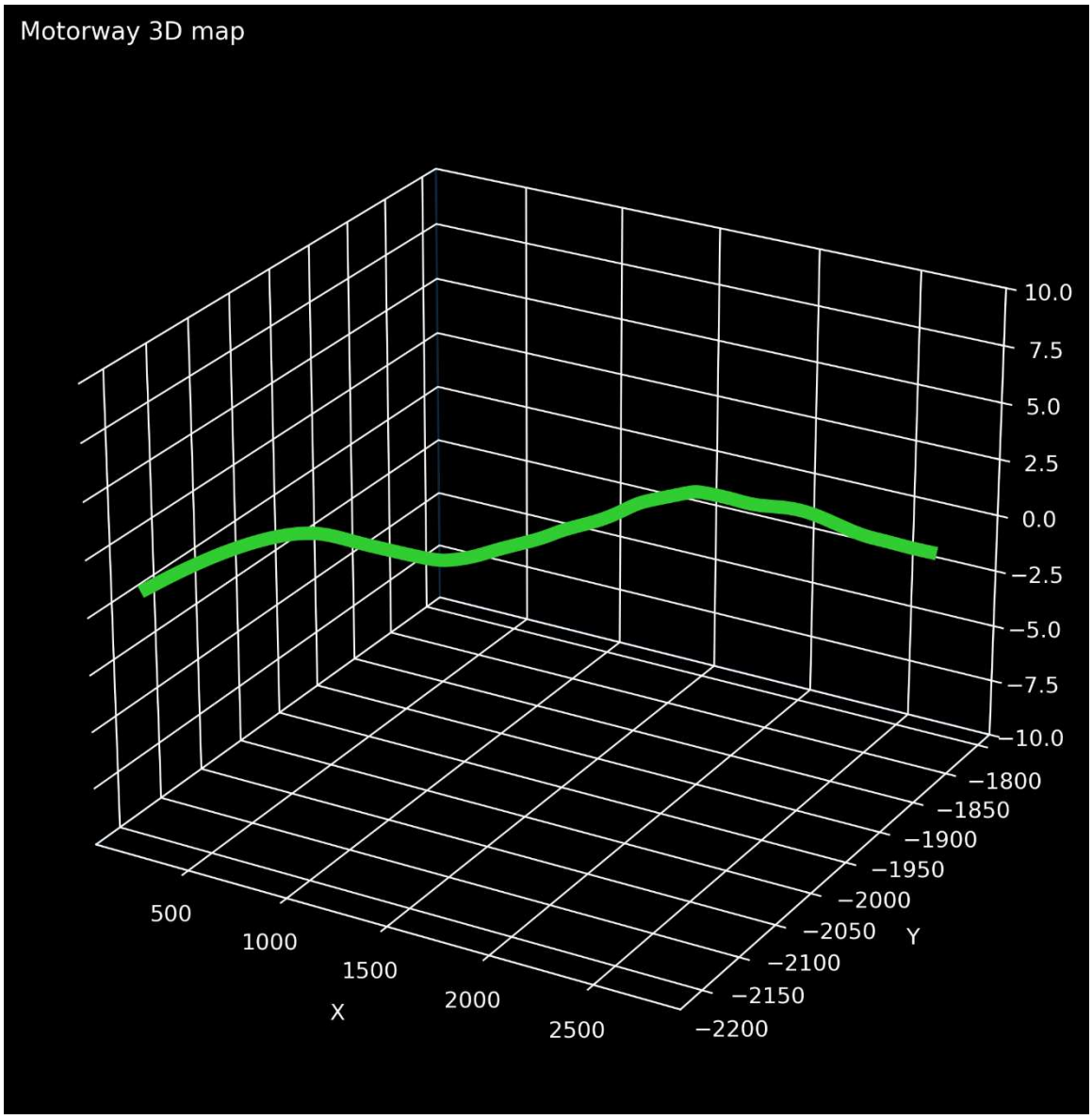


Figure 1: Motorway simulation path.

Figure 1 shows the path the participants took during their motorway simulation in both the Foerst and Virage simulators. The motorway path was created using the x and y coordinates present in the raw data. The x and y coordinates are measurements of cardinal directions and z coordinates measure elevation.

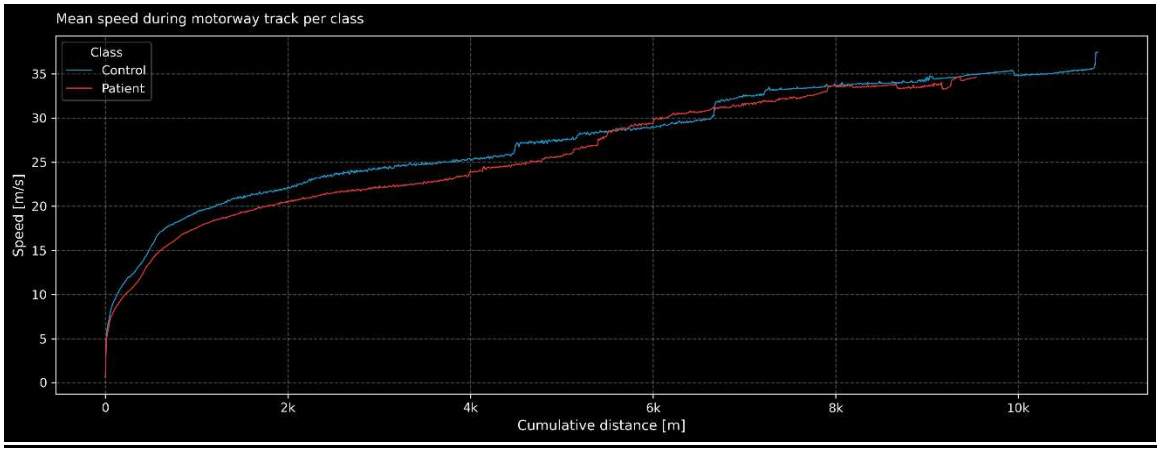


Figure 2: Mean speed during the motorway simulation.

Figure 2 shows the mean speed during the motorway simulation of each class (group of patients or controls) through the Foerst motorway track. The control class is shown in blue, and the patient class in red. The y-axis shows the speed in meters per second and the x-axis shows the cumulative distance in meters. The data collection did not start from speed 0 and the end point of the track was calculated using the speed of the participants at the end of the timed data collection. The speed of the control group is higher than the speed of the patient group, especially between the 2-kilometer and 4-kilometer marks.

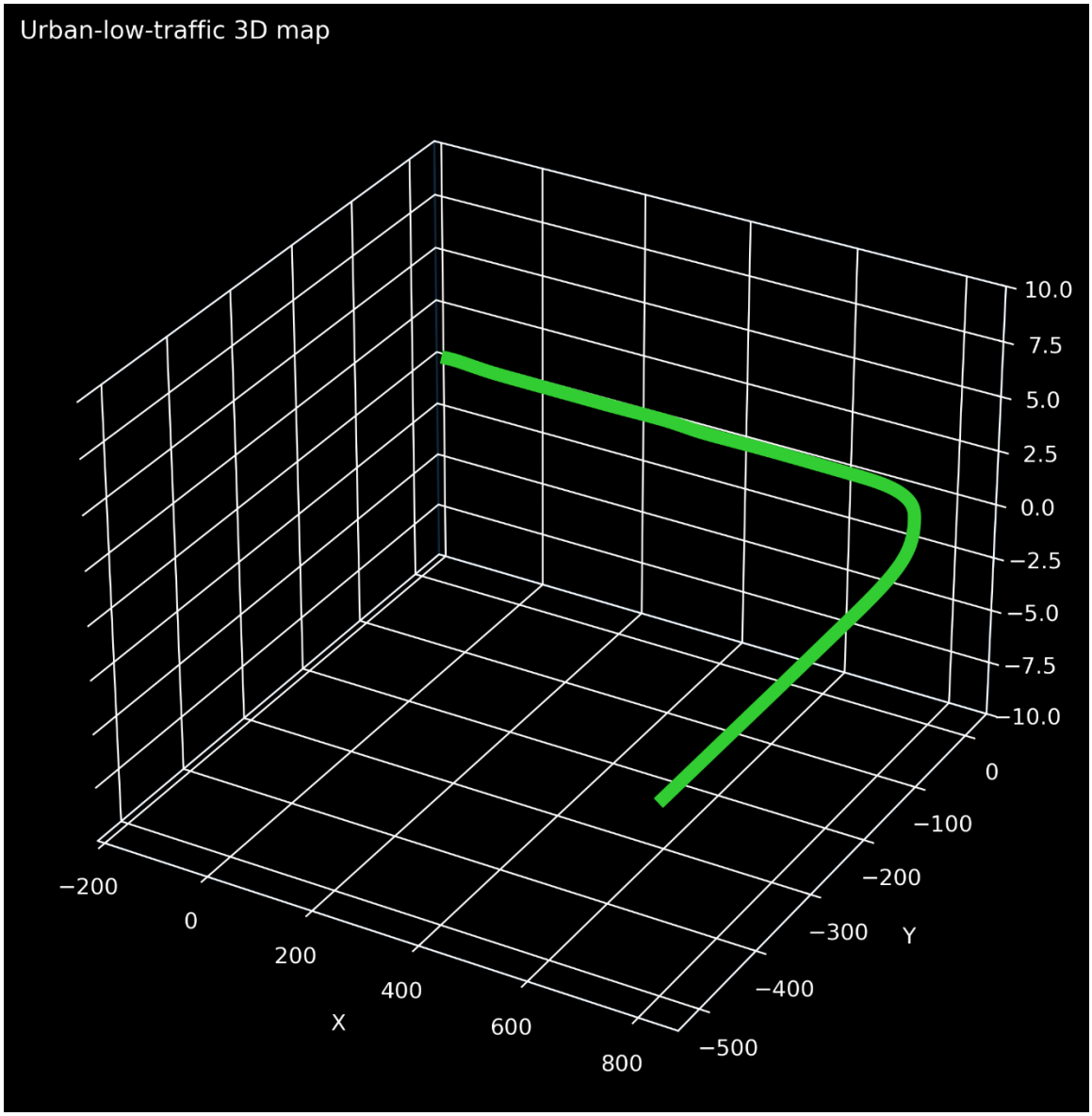


Figure 3: Urban low traffic simulation path.

Figure 3 shows the path participants took during the urban low traffic setting simulation. The urban low traffic path was created using the x and y coordinates present in the raw data.

The x and y coordinates are measurements of cardinal directions and z coordinates measure elevation.

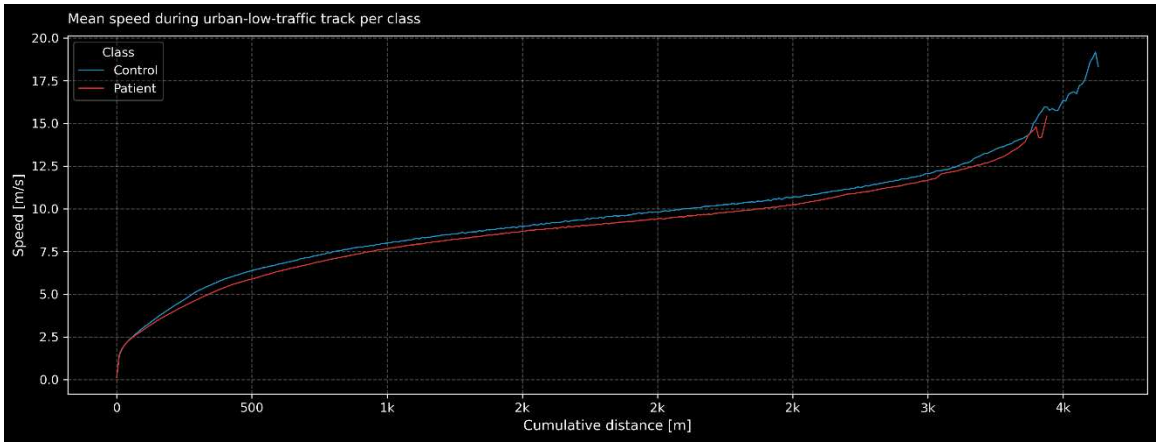


Figure 4: Mean speed during the urban low traffic simulation.

Figure 4 shows the mean speed during the urban low traffic scenario. The control class is shown in blue, and the patient class in red. The y-axis shows the speed in meters per second, and the x-axis shows the cumulative distance in meters. The speed of the control group is slightly higher than the speed of the patient group throughout the urban low traffic simulation track.

Variables	Class	
	Control	Patient
Motorway acceleration mean	0.044	0.021
Motorway acceleration std	0.861	1.031
Motorway deceleration mean	0.023	0.029
Motorway deceleration std	0.067	0.093
Motorway headway mean	0.037	0.062
Motorway headway std	0.027	0.050
Motorway lane offset mean	0.152	0.160
Motorway lane offset std	0.113	0.131
Motorway speed km/h mean	68.596	66.960
Motorway speed km/h std	5.684	6.362
Motorway steering wheel movement mean	0.025	0.037
Motorway steering wheel movement std	0.036	0.052
Urban acceleration mean	-0.009	-0.006
Urban acceleration std	1.495	1.757
Urban deceleration mean	0.063	0.093
Urban deceleration std	0.160	0.208
Urban headway mean	0.127	0.178
Urban headway std	0.161	0.177
Urban lane offset mean	0.206	0.224
Urban lane offset std	0.129	0.149
Urban orientation mean	2.783	2.695
Urban orientation std	2.627	2.785
Urban speed km/h mean	20.719	20.573
Urban speed km/h std	3.118	3.468
Urban steering wheel movement mean	-0.015	0.001
Urban steering wheel movement std	0.118	0.091

Table 1: Means and standard deviations of the data obtained across control participants and patient participants.

Headway (distance from leading vehicle) was calculated as seconds to collision to leading vehicle if the participant kept the speed they had when the leading vehicle was created in the simulation. The headway value in the absence of a leading vehicle in all scenarios was zero. Lane offset was calculated in meters.

In table 1 the mean and standard deviation of all variables for control and patient class is shown. All variables were used for the machine learning algorithm analysis.

3.3 Machine learning analysis

3.3.1 Aggregated data Decision Tree Classifier

The training model created (N=34) using the Decision Tree Classifier managed to correctly predict all the control class samples (18/18) and managed to predict correctly (15/16) patient class samples.

We obtained accuracy of 0.97 and metrics for the training confusion matrix of 0.97, 0.97 and 0.97 (precision, recall and F1-score, respectively).

The model created using the training samples was tested on the test samples. The testing model created (N=35) managed to identify most of the patient class (14/23) and was not able to identify adequately the control class (8/12). In the testing confusion matrix, we obtained accuracy of 0.63 and weighted-average precision, recall and F1-score of 0.67, 0.63 and 0.64. The modeling Root Mean Square Error metric was 0.61.

Table 2 shows the analytical metrics obtained for each class in our model in the testing of the model created. Patient participants (Class:1) were more correctly identified from the model created, compared to healthy controls (Class:0).

	Precision	Recall	F1-score
Class: 0	0.47	0.67	0.55
Class: 1	0.78	0.61	0.68
Macro average	0.62	0.64	0.62
Weighted average	0.67	0.63	0.64

0=control, 1=patient, Accuracy: 0.63

Table 2: Metrics obtained for each class in our model.

3.3.2 Aggregated data Random Forest Classifier

The training model created (N=34) using the Random Forest Classifier managed to correctly predict the control participants (18/18) and almost all the patient class (14/16) in the training sample. We obtained accuracy of 0.94 and metrics for the training confusion matrix of 0.95, 0.94 and 0.94.

The testing model created (N=35), was adequately able to identify correctly the control class (9/12) and most of the patient class (17/23). We obtained accuracy of 0.74 and weighted-average precision, recall and F1-score of 0.76, 0.74 and 0.75. The modeling Root Mean Square Error metric was 0.53 and the Support Vector Classifier had an accuracy score of 0.66.

Table 3 shows the analytical metrics obtained for each class of our Random Forest Classifier model in the testing of the model created. The patient participants (Class:1) was more correctly identified from the model created, compared to the healthy controls (Class:0).

	Precision	Recall	F1-score
Class: 0	0.60	0.75	0.67
Class: 1	0.85	0.74	0.79
Macro average	0.72	0.74	0.73
Weighted average	0.76	0.74	0.75

0=control, 1=patient, Accuracy: 0.74

Table 3: Metrics obtained for each class in our model.

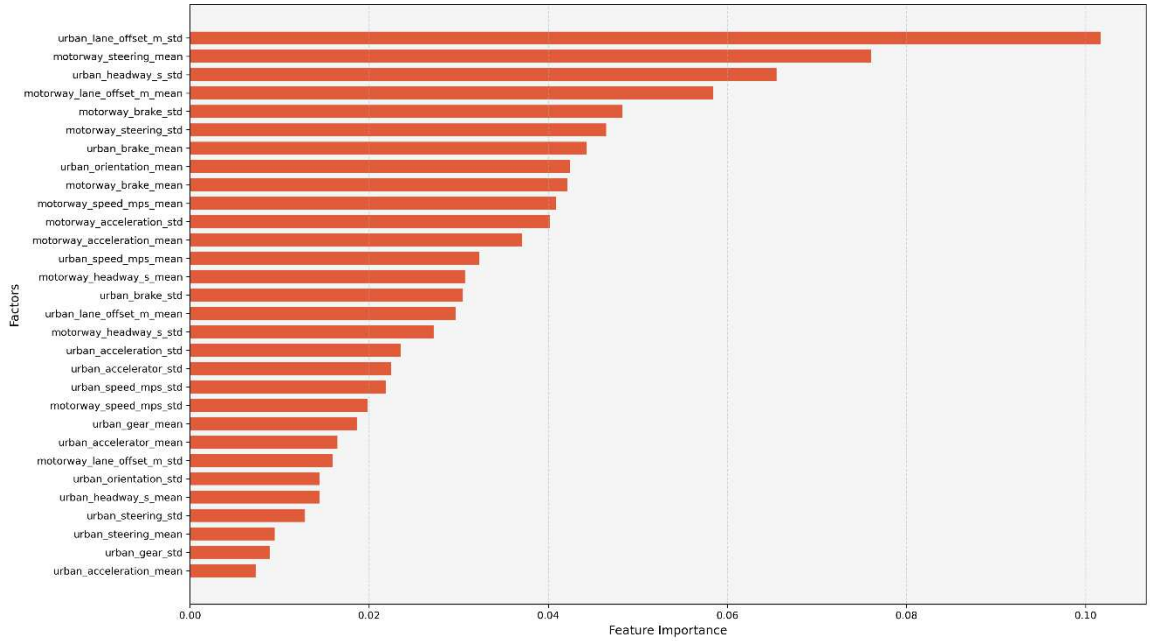


Figure 5: Feature importance Gini scores of the factors.

Figure 5 shows the results of the feature analysis of the factors present in the data collected from the driving simulators. The figure shows the Gini scores of the factors used in the Random Forest model.

3.3.3 Time series data Random Forest Classifier

We obtained accuracy of 0.87 and metrics of 0.87, 0.86 and 0.87 from the training data samples. As shown in table 4 in the test model we obtained an accuracy of 0.57 and weighted-average precision, recall and F1-score of 0.68, 0.57 and 0.60. The model created managed to correctly identify most of the patient class participants present in the scenario and failed to correctly identify the control class participants.

Table 4 shows the analytical metrics obtained for each class in our model in the testing of the model created. The patient participants (Class:1) was more correctly identified from the model created, compared to the healthy controls (Class:0) which failed to be correctly predicted.

	Precision	Recall	F1-score
Class: 0	0.30	0.55	0.39
Class: 1	0.80	0.58	0.67
Macro average	0.55	0.57	0.53
Weighted average	0.68	0.57	0.60

0=control, 1=patient, Accuracy: 0.57

Table 4: Metrics obtained for each class in our test model. (N = 16).

3.3.4 Time series data XGBoost Classifier

Booster	gbtree
learning_rate	1
max_depth	6 (15 for the feature-selection refit)
eval_metric	auc
scale_pos_weight	1
Gamma	0
n_estimators	250
subsample	1
colsample_bytree	0.8
colsample_bynode	0.8
reg_alpha	10
Objective	binary:logistic

Table 5: Parameter values of the XGBoost Classifier model.

In table 5, the parameters for the classification model of XGBoost Classifier are shown. The booster used was the 'gbtree' which uses tree-based models. Due to the classification being binary between the classes, control and patient, the objective was set as binary:logistic. The model evaluation metric used was 'error' which represents the binary classification error rate.

We obtained accuracy of 0.97 and metrics of 0.97, 0.97 and 0.97 from the training data samples. As shown in the classification report in table 6, the model created using the xgboost classifier managed to correctly predict most of the patient class subjects but showed a low ability to correctly predict control participant when tested on the test sample. Thus, for the test data we obtained an accuracy of 0.61 and weighted-average precision, recall and F1-score of 0.66, 0.61 and 0.63.

Table 6 shows the analytical metrics obtained for each class in our model in the testing of the model created. The patient participants (Class:1) was more correctly identified from the model created, compared to the healthy controls (Class:0).

	Precision	Recall	F1-score
Class: 0	0.30	0.44	0.36
Class: 1	0.78	0.66	0.72
Macro average	0.54	0.59	0.54
Weighted average	0.66	0.61	0.63

0=control, 1=patient, Accuracy: 0.61

Table 6: Metrics obtained for each class in our model. (N = 16).

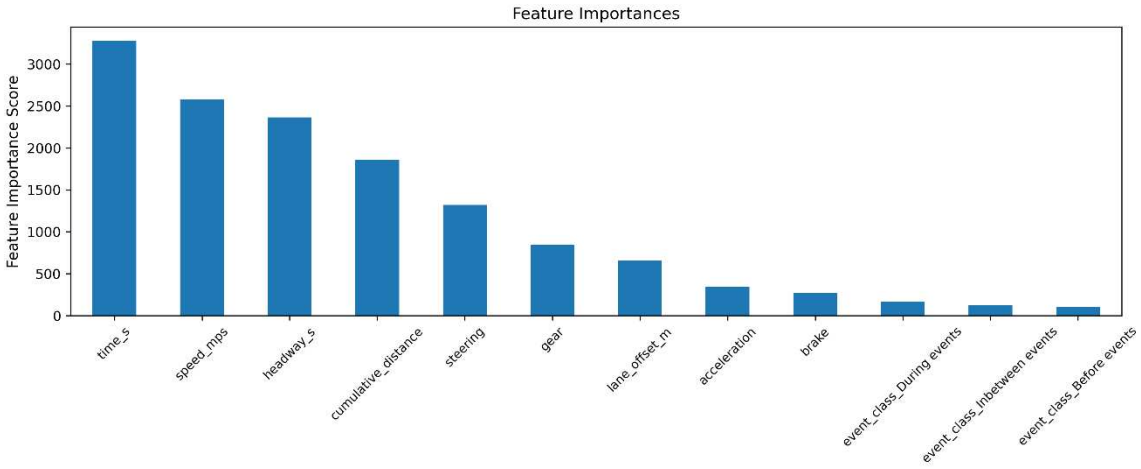


Figure 6: Feature importance Gini scores of the factors.

Figure 6 shows the results of the feature analysis of the factors present in the data collected from the driving simulators. The figure shows the F-scores (number of times a feature is used to split the data across all trees) of the 12 factors which were used for our model.

On the x axis we see the names of the factors and on the y axis we see the times those factors were used for splitting of the nodes in the model decision making.

The parameter of time_s is the time it took the participant to complete the driving simulation, headway_s (see table 1) is calculated in seconds, speed_mps is the speed in meters per second, cumulative_distance is the distance traveled plus the distance traveled after the end of the trial given the speed of the car and a linear deceleration to a full stop. Steering is the total steering wheel movements, gear is the gearbox number, lane_offset_m (see section 2.4) calculated in meters, acceleration (see section 2.4) is the acceleration of the car. event_class_During events is if an event was taking place, event_class_Inbetween events, is if an event has happened and another event is upcoming, event_class_Before events is if no event has happened yet.

Factors ordered	Gini - Threshold	Number of factors used	Accuracy
Time_s	0.152	1	48.96%
Headway_s	0.130	2	59.15%
Speed_mps	0.119	3	56.22%
Cumulative_distance	0.100	4	63.30%
Steering wheel movements	0.093	5	65.27%
Gear	0.085	6	64.15%
Lane_offset_m	0.085	7	74.85%
Acceleration	0.080	8	77.48%
Brake	0.060	9	79.10%
Event_class_During_events	0.045	10	78.67%
Event_class_Inbetween_events	0.030	11	78.84%
Event_class_Before_events	0.020	12	77.83%

Table 7: Changes of our model using the Gini threshold of the factors.

Table 7 shows the number of factors used and their consequent threshold value and the obtained model accuracy. The training and testing models were transformed based on the factor importance values. The threshold values generated are the normalized values for the gradient boosted algorithm tree splitting method, the Gini AUC coefficient values. The threshold value increases as the number of factors increase and the accuracy decreases. This means that using more factors than necessary, our model starts losing accuracy

For example, if we use 9 factors, our models' thresholds will be 0.060 and accuracy 79.1%. The factors used would be Time_s, Headway_s, Speed_mps, Cumulative_distance, Steering, Gear, Lane_offset_m, acceleration and Brake.

4. Discussion

In a previous report on the same dataset (Tsoutsi et al. 2023) results based on mean values analyses showed that, while severity of insomnia symptoms and of somnolence in patients were positively associated with lateral position standard deviation (denoting difficulty in maintaining a stable position), patients with depression especially if somnolence was present kept a longer safety distance from the leading vehicle, probably as a safety measure to counteract any impairment in driving performance. Understanding the effect of depression on driving ability is of major importance given the prevalence of depression and its effect on executive functions, which in turn can compromise driving ability (Aduen et al. 2015; Alavi et al. 2017; Brunnauer et al. 2008; Buckley et al. 2015, 2017; Bulmash et al. 2006; Ferrari et al. 2012; Liang and Yang 2022; Perrier et al. 2014; Shen et al. 2009; Tsoutsi et al. 2019, 2023; WHO 2017, 2025; Wickens et al. 2014; Wingen et al. 2006).

On the other hand, the use of machine learning methods for the analysis of micro-driving data, such as those obtained from driving simulators, remains at an early stage (Koohestani et al. 2018; Yoshizawa et al. 2016). Although recent studies have begun to explore the application of machine learning to driving behavior in individuals with depression (Katrakazas et al. 2020; Chen et al. 2025), its use in high-frequency driving simulator data from individuals with depression remains limited.

The present study applies machine learning methods to develop predictive models capable of distinguishing between healthy controls and individuals with depression using both aggregated and time-series driving simulator data.

4.1 Aggregated data

The models created were analyzed and scored using accuracy, recall, precision and F1-score. The metrics obtained were compared to the accuracy of the model using a Support Vector Classifier penalizing algorithm due to the unbalanced class distribution of the samples. The classifier was used as a threshold for the performance of the models used. The data from both the simulators, Virage Driving Simulator “Virage VS500M”and the Foerst Driving Simulator “FPF F10P”, were used to create the aggregated dataset, using mean and standard deviation of the data. In the aggregated data approach, it was found that the best predictive model was the Random Forest Classifier.

The Decision Tree Classification model was outperformed by the Random Forest Classification model. We obtained accuracy scores of 0.63 (see section 3.3.1), 0.74 (section 3.3.2) for the Decision Tree Classifier and Random Forest Classifier, respectively. The Random Forest Classifier also outperformed the Support Vector Classifier, which serves as a measurement of classification baseline. The Root Mean Square Error, the square root of error variance, observed for the Decision Tree Classifier was 0.61, which demonstrates the weakness of the Decision tree model. In the Random Forest Classifier, the Root Mean Square Error was 0.53 which suggests that the ability of the Random Forest Classifier in error variance is better than the Decision Tree Classifier.

The Random Forest model created is better at predicting the classification of the participants classes than randomly. Both classifiers were better at predicting the classification of patient participants than control participants as seen in tables 2 and 3. From **Fig. 5** we can see the factors of importance of the Random Forest Classifier (factors of importance with Gini value higher than 0.05) were urban lane offset standard deviation, motorway steering wheel movement mean and urban headway standard deviation. Those 3 factors were the most important in creating the splits of the model of the Random Forest Classifier, and contribute as new parameters in differentiating between controls and patients to those already identified through the analysis of mean values on the same dataset (Tsoutsi et al. 2019, 2023).

4.2 Time series

The time series models created were scored using the same metrics. The Random Forest Classifier used on the time series data obtained an accuracy of 0.57 (and metrics of 0.68, 0.67 and 0.60) in the testing sample (see section 3.3.3). The balanced subsample approach using SMOTE-ENN in the Random Forest Classifier performed worse than expected. The SMOTE-ENN subsample undersampled the patient class, but from the classification report in table 4 we observe that the ability to classify controls failed. The model obtained metrics of 0.30, 0.55 and 0.39 for the control class and 0.80, 0.58 and 0.67 for the patient class respectively.

The XGBoost classifier obtained an accuracy of 0.61 and metrics of 0.66, 0.61 and 0.63 on the testing sample (see section 3.3.4). In table 6 we can see that the model obtained AUC metrics of 0.30, 0.44, 0.36 for the control class and 0.78, 0.66 and 0.72 for the patient class. From the feature of importance obtained in **Fig 6** the model was refitted using those threshold values and managed to obtain an accuracy of 0.79. From table 7, through different feature selection runs, the difference between n=9 and n=10 accuracy values is ± 0.5%. This means that the factors Event_class_Before_events and Event_class_Inbetween_events are not useful for our model, meaning that the driving events were of little importance to the classification of the controls and patients.

5. Limitations

One of the main limitations of this study was the heterogeneity of the data provided by the simulators. This led to data pre-processing and feature creation needed to achieve homogeneity between simulators. Another limitation is that the average of the present healthy control participants cannot be considered a true representation of the possible average driving ability of the population, due to the relatively small number of participants. Both aforementioned limitations could be remedied by acquiring a larger participant group, where the hypothesis of a control subset and patient subset can be established with the backing of sufficient participants. However, acquiring and testing a large number of subjects on driving simulators is not practical. In addition, the training and splitting of the data in the model creation phases further diminishes the volume of participants

in each set, making cross validation an inconclusive task. The confidence levels of the actual predictive abilities of our models are not high, and through the use of different sampling methods, our models had high predictive ability on predicting only the patient class.

6. Conclusion

Depression is one of the most frequent psychiatric disorders. In our study, we are seeking to identify depression from driving characteristics by using machine learning methods, while dealing with heavily unbalanced data between healthy control participants and depressed patient participants, as well as with a significant level of noise in the data and lack of homogeneity between the two simulators used.

As for the vector point aggregated data approach, using the Random Forest Classifier, we managed to correctly identify control drivers and patient drivers with an accuracy of 0.74. The most important factors for the classification of this approach were the urban lane offset standard deviation, motorway steering wheel movement mean and urban headway standard deviation. These findings are important since they add to the results regarding differences between control and patient groups regarding driving data. More specifically, in the study of the same dataset, results based on mean values analysis showed that differences between patients and controls regarded lateral position standard deviation and mean headway values. The findings of the present analysis add to the above steering wheel movement and headway standard deviation, providing two additional parameters of differentiation.

The time series approach failed to properly predict the control participant class as seen in the classification reports in table 4 and table 6. Through the use of the F-scores obtained from the XGBoost classifier (table 7) we optimized our model and obtained a model accuracy of 0.79. Unfortunately, the results still favor the patient class compared to the control class, skewing the model towards favoring the patient participants.

The aggregated Random Forest Classifier is a proof of concept of the ability of machine learning models to distinguish between patient and control participants. Furthermore, from the classification reports of both aggregated model approaches and the time series approach we observed that the patient participant class is a highly homogeneous group compared to the control participant class, as shown from the Decision Tree models and Random Forest models which are based on the outcome variable variance. In conclusion, this study provides insight into the driving behavior in different groups of drivers with a focus on drivers suffering from depression. The homogeneity observed in drivers suffering from depression should be further studied to obtain a better understanding of the relationship between driving and depression.

7. Future Work

The current study is preliminary research on how machine learning can be utilized in predicting psychiatric disorders from driver analytics. Deep learning, micro-driving data simulator data and inclusion of other psychiatric disorders can enhance the findings of the current investigation. Moreover, in order for our approach to become more specific to participants, more samples of the same participant in the same driving simulator environment should be used. The classified

aggregated and time series driving data models created should include a larger population sample to better learn to correctly classify both healthy controls and patients with depression. Future studies should also improve the interpretability of the machine learning models by employing feature importance techniques such as SHAP (SHapley Additive exPlanations) (Lundberg and Lee 2017; Sekadakis et al. 2025), thereby facilitating a better understanding of the contribution of individual driving variables to model predictions.

References

1. Afghari A, Papadimitriou E, Pilkington-Cheney F, Filtness A, Brijs T, Brijs K, Cuenen A, De Vos B, Dirix H, Ross V, et al. (2022) Investigating the effects of sleepiness in truck drivers on their headway: An instrumental variable model with grouped random parameters and heterogeneity in their means. Anal. Methods Accid. Res. 36:100241. https://doi.org/10.1016/j.amar.2022.100241

2. Aduen P, Kofler M, Cox D, Sarver DE, Lunsford E (2015) Motor vehicle driving in high incidence psychiatric disability: Comparison of drivers with ADHD, depression, and no known psychopathology. J. Psychiatr. Res. 64:59–66. https://doi.org/10.1016/j.jpsychires.2015.03.009

3. Alavi S, Mohammadi M, Souri H, Kalhori S, Jannatifard F, Sepahbodi G (2017) Personality, driving behavior and mental disorders factors as predictors of road traffic accidents based on logistic regression. Iran J. Med. Sci. 42:24–31. https://pubmed.ncbi.nlm.nih.gov/28293047/

4. Batista G, Prati R, Monard M (2004) A Study of the Behavior of Several Methods for Balancing machine Learning Training Data. SIGKDD Explorations. 6:20-29. https://doi.org/10.1145/1007730.1007735

5. Birchler C, Khatiri S, Bosshard B, et al. (2023) Machine learning-based test selection for simulation-based testing of self-driving cars software. Empir. Software Eng. 28:71 https://doi.org/10.1007/s10664-023-10286-y

6. Brunnauer A, Laux G, David I, Fric M, Hermisson I, Möller H (2008) The impact of reboxetine and mirtazapine on driving simulator performance and psychomotor function in depressed patients. J. Clin. Psychiatry 69:1880–1886. https://doi.org/10.4088/jcp.v69n1205

7. Brunnauer A, Buschert V, Fric M, Distler G, Sander K, Segmiller F, Zwanzger P, Laux G (2015) Driving performance and psychomotor function in depressed patients treated with Agomelatine or Venlafaxine. Pharmacopsychiatry 48:65-71. https://doi.org/10.1055/s-0034-1398509

8. Buckley S, Robinson K, Stapleton T (2015) Medical fitness to drive and mental health: Are you aware of the guidelines? Ir. J. Occup. Ther. 2015 43:35–39.
9. Buckley S, Robinson K, Stapleton T (2017) Driving and depression: Health professional's perspectives in Ireland. J. Transp. Health 7:235–246. https://doi.org/10.1016/j.jth.2017.09.003

10. Bulmash E, Moller H, Kayumov L, Shen J, Wang X, Shapiro C (2006) Psychomotor disturbance in depression: Assessment using a driving simulator paradigm. J. Affect Disord. 93:213–218. https://doi.org/10.1016/j.jad.2006.01.015

11. Chawla N, Bowyer K, Hall L, Kegelmeyer W (2002) SMOTE: Synthetic Minority Over-sampling Technique. J. Artif. Intell. Res. 16:321-357. https://doi.org/10.1613/jair.953

12. Chen C, Brown DC, Al-Hammadi N, Bayat S, Dickerson A, Vrkljan B, Blake M, Zhu Y, Trani JF, Lenze EJ, Carr DB, Babulal GM (2025) Identifying major depressive disorder in older adults through naturalistic driving behaviors and machine learning. NPJ. Digit. Med. 8(1):102. https://doi.org/10.1038/s41746-025-01500-w

13. Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). 785-794. https://doi.org/10.1145/2939672.2939785

14. Elassad Z, Mousannif H, Al Moatassime H, Karkouch A (2020) The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. Engineering Applications of Artificial Intelligence 87:103312. https://doi.org/10.1016/j.engappai.2019.103312

15. Fawcett T (2006) Introduction to ROC analysis. Pattern Recognit. Lett. 27: 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

16. Ferrari A, Somerville A, Baxter A, Norman R, Patten S, Vos T, Whiteford H (2012) Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. Psychol Med. 43:471–481. https://doi.org/10.1017/S0033291712001511

17. Fugiglando U, Massaro E, Santi P, Milardo S, Abida K, Stahlmann R, Netter F, Ratti C (2019) Driving Behavior Analysis through CAN Bus Data in an Uncontrolled Environment. IEEE Trans. Intell. Transp. Syst. 20.2:737-748. https://doi.org/10.1109/TITS.2018.2836308

18. Katrakazas C, Antoniou C, Yannis G(2020) Identification of driving simulator sessions of depressed drivers: A comparison between aggregated and time-series classification. Transp. Res. Part F Traffic Psychol. Behav. 75:16-25. https://doi.org/10.1016/j.trf.2020.09.015

19. Koohestani A, Kebria P, Khosravi A, Nahavandi S (2018) Drivers Performance Evaluation using Physiological Measurement in a Driving Simulator. Digital Image Computing: Techniques and Applications (DICTA) 2018:1-6. https://doi.org/10.1109/DICTA.2018.8615763

20. Lemaître G, Nogueira F, Aridas C (2016) Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. J. Mach. Learn. Res. 18: 17:1-17:5. https://doi.org/10.48550/arXiv.1609.06570

21. Liang O, Yang C (2022) Mental health conditions and unsafe driving behaviors: A naturalistic driving study on ADHD and depression. J Safety Res. 82:233–240. https://doi.org/10.1016/j.jsr.2022.05.014

22. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30. https://doi.org/10.48550/arXiv.1705.07874

23. National Highway Traffic Safety Administration (2008) National Motor Vehicle Crash Causation Survey: Report to Congress. DOT HS 811 059. U.S. Department of Transportation, Washington, DC. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811059 Accessed 15 June 2025

24. Nishat M, Faisal F, Ratul I, Al-Monsur A, Ar-Rafi A, Nasrullah S, Reza T, Khan R (2022) A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset. Sci. Program. 3649406:1-17. https://doi.org/10.1155/2022/3649406

25. Perrier J, Bertran F, Marie S, Couque C, Bulla J, Denise P, Bocca M (2014) Impaired driving performance associated with effect of time duration in patients with primary insomnia. Sleep 37:1565–1573. https://doi.org/10.5665/sleep.4012

26. Philip P, Sagaspe P, Taillard J, Valtat C, Moore N, Akerstedt T, Charles A, Bioulac B (2005) Fatigue, sleepiness, and performance in simulated versus real driving conditions. Sleep 28:1511–1516. https://doi.org/10.1093/sleep/28.12.1511

27. Rudin-Brown C, Williamson A, Lenné M (2009) Can Driving simulation be used to predict changes in real-world crash risk? Report #299, Monash University, Accident Research Center, Melbourne, Australia. https://www.monash.edu/__data/assets/pdf_file/0010/216487/Can-driving-simulation-be-used-to-predict-changes-in-real-world-crash-risk.pdf. Accessed 10 June 2024

28. Schreier D, Banks C, Mathis J (2018) Driving simulators in the clinical assessment of fitness to drive in sleepy individuals: A systematic review. Sleep Med. Rev. 38:86–100. https://doi.org/10.1016/j.smrv.2017.04.004

29. Sekadakis M, et al. (2025) Analyzing SHAP values of XGBoost algorithms to understand driving features affecting take-over time from vehicle alert to driver action. *Displays* 103263.https://doi.org/10.1016/j.displa.2025.103263

30. Shen J, Moller H, Wang X, Chung S, Shapiro G, Li X, Shapiro C (2009) Mirtazapine, a sedating antidepressant, and improved driving safety in patients with major depressive disorder: A prospective, randomized trial of 28 patients. J. Clin. Psychiatry 70:370–377. https://doi.org/10.4088/jcp.08m04234

31. Tsoutsi V, Dikeos D, Basta M, Papadakaki M (2019) Driving behaviour in depression: Findings from a driving simulator study. Safety 5:70. https://doi.org/10.3390/safety5040070

32. Tsoutsi V, Papadakaki M, Yannis G, Pavlou D, Basta M, Chliaoutakis J, Dikeos D (2023) Driving behaviour in depression based on subjective evaluation and Data from a Driving Simulator. *Int. J. Environ. Res. Public Health 20*:5609. https://doi.org/10.3390/ijerph20085609

33. Van der Sluiszen N, Wingen M, Vermeeren A, Vinckenbosch F, Jongen S, Ramaekers J (2017) Driving performance of depressed patients who are untreated or receive long-term antidepressant (SSRI/SNRI) treatment. Pharmacopsychiatry 50:182–188. https://doi.org/10.1055/s-0043-111600

34. Wickens C, Smart R, Mann R (2014) The Impact of depression on driver performance. Int. J. Ment. Health Addict. 12:524–537. https://doi.org/10.1007/s11469-014-9487-0

35. Wingen M, Ramaekers J, Schmitt J (2006) Driving impairment in depressed patients receiving long-term antidepressant treatment. Psychopharmacology 188:84–91. https://doi.org/10.1007/s00213-006-0471-7

36. World Health Organization (2009) Global status report on road safety. World Health Organization Eastern Mediterranean region. https://www.emro.who.int/press-releases/2009/global-status-report-on-road-safety.html Accessed 15 June 2025

37. World Health Organization (2017) Depression and other common mental disorders: global health estimates. World Health Organization. https://www.who.int/publications/i/item/depression-global-health-estimates Accessed 15 June 2025

38. World Health Organization (2025) Depressive disorder (depression). World Health Organization. https://www.who.int/news-room/fact-sheets/detail/depression Accessed 1 September 2025

39. Yoshizawa A, Nishiyama H, Iwasaki H, Mizoguchi F (2016) Machine-learning approach to analysis of driving simulation data. IEEE 15th International Conference on

Cognitive Informatics & Cognitive Computing (ICCI*CC) 2016:398-402. https://doi.org/10.1109/ICCI-CC.2016.7862067

Declarations

Funding
This research received no external funding. The study has been conducted in the frame of a PhD Degree (D.B.), Medical School, National and Kapodistrian University of Athens, Athens, Greece. The authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement
The data that support the findings of this study are available on request from the corresponding author, but restrictions apply to the availability of these data, which were used under license from the First Department of Psychiatry, Eginition Hospital, National and Kapodistrian University of Athens and the collaborating centres for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission from the First Department of Psychiatry, Eginition Hospital, National and Kapodistrian University of Athens and the collaborating centres.

Conflicts of Interest
The authors declare no conflict of interest regarding this work.

Acknowledgement
The authors appreciate the support of the First Department of Psychiatry, Medical School, National and Kapodistrian University of Athens, Athens Greece, the Laboratory of Health and Road Safety (LaHeRS), Department of Social Work, Hellenic Mediterranean University and the Traffic Engineering Laboratory at the National Technical University of Athens.

Institutional Review Board Statement
The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Ethics Committee of the First Department of Psychiatry, Eginition Hospital, National and Kapodistrian University of Athens (464/5-7-2019) and by the Ethics Committee of University General Hospital of Heraklion (798/20/25-7-2018). All personal data were collected, processed, and stored in accordance with the General Data Protection Regulation (GDPR; Regulation (EU) 2016/679). Written informed consent was obtained from all participants prior to participation.

Code availability
The code used has been made available on GitHub (‘https://github.com/DBobos/D.Bobos-V.Tsoutsi ‘) for interpretation and reproducibility.

Author Contributions

Conceptualization, V.T., D.B.; Data curation, V.T. and D.D.; Formal analysis, V.T.,D.B., D.D., P.K. and G.M.; Investigation, V.T., D.B.; Methodology, V.T., D.B. D.D., P.K., G.M. and G.Y.; Project administration, V.T. and D.D., D.V.; Resources, G.Y., D.D., V.M. and E.G.; Supervision, D.D.; Validation, V.T., D.D., G.Y., V.M., E.G. and D.V.; Visualization, V.T. and D.B.; Writing, V.T., D.B., P.K., G.M. and D.D.; Original Draft Preparation, V.T., D.B.; Writing Review and Editing, D.B, V.T., G.M., P.K., E.G., V.M., G.Y., D.D., D.V. All authors have read and agreed to the published version of the manuscript.

Author information

Dimitris Bobos

Laboratory of Biology, Medical School of Athens, National and Kapodistrian University of Athens, Greece

dbobos@uoa.gr
https://orcid.org/**0009-0003-5570-3133**

Vagioula Tsoutsi

Assistant Professor

First Department of Psychiatry, Eginition Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece

Faculty of Social Work, University of Iceland, Iceland

vagia.ts@gmail.com

https://orcid.org/0000-0002-6303-2558

Panagiotis Kazanis

Data Analyst**,** Circana Athens, Greece

pankaza4@gmail.com

Giannis Manesis

Data & Analytics Leader, Cosmos Sport SA, Heraklion, Greece

gimanesis@gmail.com

Evangelos Grinakis

Head of Heraklion Dual Diagnosis Unit, KETHEA, University Hospital, Heraklion, Greece

egrinakis@gmail.com

Vasilios G. Masdrakis

Professor of Psychiatry, National and Kapodistrian University of Athens, Medical School, Greece

vmasdrakis@med.uoa.gr

George Yannis

Professor of Traffic and Safety Engineering, National Technical University of Athens, Greece

geyannis@central.ntua.gr

https://orcid.org/0000-0002-2196-2335

Dimitrios Vogiatzis

NCSR “Demokritos”, Athens, Greece

Professor of Data Science Deree - The American College of Greece, Athens

dimitrv@acg.edu

Dimitris Dikeos

First Department of Psychiatry, Eginition Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece

ddikeos@med.uoa.gr