Computers & Industrial Engineering 198 (2024) 110667

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Artificial intelligence abnormal driving behavior detection for mitigating traffic accidents☆

Yoshiyasu Takefuji [a,*], Michiyasu Tano [b], Masaya Shigehara [b,c], Shunya Sato [c]

[a] Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan
[b] Datatec Co.,Ltd., Green Place Kamata 11th floor, 7-37-10 Nishikamata Ota-ku, Tokyo 144-0051, Japan
[c] Kanematsu Corporation, 7-2 Marunouchi 2-Chome, Chiyoda-ku, Tokyo 100-7017, Japan

ARTICLE INFO

Keywords:
Artificial intelligence
Tachograph
Ensemble machine learning
Entropy
Binary classification

ABSTRACT

Annually, a tragic toll of 1.3 million lives is lost on roads across the globe, with tens of millions more suffering injuries or disabilities. The necessity for precise detection of abnormal driving behavior is paramount in reducing traffic accidents. This paper aims to bridge the gap between normal and abnormal driving patterns, offering near-flawless detection capabilities. This paper presents a novel AI tachograph prototype, the first of its kind, that can classify driving behavior into normal and abnormal in real time with an impressive accuracy of 99.99 %. This high level of accuracy is achieved by using a bias-reduction method. The bias-reduction method focuses on minimizing biases in the dataset, such as surrounding situations, location, driver information, and car types. This approach significantly enhances the prediction accuracy of existing machine learning algorithms. The dataset used for this research is quite extensive, consisting of anomaly data collected from 10,181 commercial vehicles and 12,530 drivers in just 0.1 s. This rich dataset is crucial for building a reliable model. The effectiveness of the proposed method was validated using 10-fold cross validation on 480 k to 540 k instances with 36 determinants. The results clearly demonstrated that reducing bias leads to higher prediction accuracy. The paper also plans to compare the prediction accuracy of balanced and imbalanced datasets. The findings from this research have broader implications as the proposed method can be applied generally to machine learning to improve prediction accuracy.

1. Introduction

Every year, roadways worldwide witness the tragic loss of 1.3 million lives, while tens of millions more are injured or disabled (Kuupiel et al., 2023). Numerous researchers have endeavored to develop systems that can automatically detect dangerous driving behaviors. However, due to a scarcity of data, these efforts have yet to yield a satisfactory solution. The development of an impeccable system for the automatic detection of hazardous driving is crucial in order to reduce the number of injuries occurring on our roads. A comprehensive review of literature was carried out, focusing on the mitigation of hazardous driving in both human-operated and autonomous vehicles. Numerous strategies exist for mitigating hazardous driving in both human-operated and autonomous vehicles.

For human-operated vehicles, the indirect shared control strategy is a method that promotes safe cooperative driving, particularly on hazardous, curvy roads (Zhao, 2023). In this approach, the authority to make driving decisions is dynamically shared between the human driver and the vehicle's automated system. This allocation is based on a risk assessment conducted using a data-driven Gaussian Processes Regression (GPR) model (Kuupiel et al., 2023). On the other hand, the Reference-free approach is a framework specifically designed for the shared control system of automated vehicles (Huang et al., 2020; Hu et al., 2022). Its primary objective is to alleviate any conflicts that may arise between the human driver and the vehicle's automation functionality during their interaction.

For autonomous vehicles, four strategies play a key role in mitigating hazardous driving. First, the concept of hazard-focused training involves

☆ This manuscript has not been published and is not under consideration for publication elsewhere. The author has no conflict of interest. This research has no fund. The author has read the manuscript and has approved this submission.

* Corresponding author.
E-mail addresses: takefuji@keio.jp (Y. Takefuji), michiyasu.tano@datatec.co.jp (M. Tano), masaya.shigehara@datatec.co.jp (M. Shigehara), Shunya_Sato@kanematsu.co.jp (S. Sato).

https://doi.org/10.1016/j.cie.2024.110667
Received 30 March 2024; Received in revised form 1 October 2024; Accepted 17 October 2024
Available online 28 October 2024
0360-8352/

the collection of training data that specifically targets dangerous scenarios (Paterson and Picardi, 2023). This method provides an efficient pathway for artificial intelligence to enhance the safety measures of autonomous vehicles (Gleirscher, 2017). Second, hazard Analysis to Hazard Mitigation Planning is a strategy that equips highly automated vehicles with verified controllers (Gleirscher, 2017). These controllers are capable of reliably identifying and mitigating hazards in all conceivable operational situations. Third, safety checklists and audits refer to the process of creating a safety checklist, either from scratch or by customizing existing templates, for autonomous vehicle inspections (Lavin et al., 2022; Ko et al., 2022). This also includes scheduling fleet audits and driver assessments to identify any potential risks. Finally, cybersecurity measures are essential safeguards against cyber-attacks, which are critical for the safety of autonomous vehicles (Alsaade and Al-Adhaileh, 2023). These measures can encompass the creation of unique passwords and the implementation of other security protocols.

The suggested approach and its outcomes have potential applications in both human-driven and autonomous vehicles. By mitigating hazardous risks, they can significantly contribute to the reduction of fatalities and injuries resulting from dangerous driving.

This paper reports the world's first prototype of an artificially intelligent tachograph that can classify driving behavior into two classes, abnormal and normal, in real time, and has a prediction accuracy of 99.99 % with a bias-reduction method. Anomaly data collected in 0.1 s from 10,181 commercial vehicles and 12,530 drivers has played an important role in prototyping, building the highest accuracy dataset previously unattainable. 10-fold cross validation using 480 k to 510 k instances and 36 determinants in abnormal and normal driving behaviors justified the proposed claim. The results validated that the less bias, the higher the prediction accuracy, which had never been reached in any project in the world. The goal of the artificially intelligent tachograph is to be able to detect abnormalities in driving behavior in real time using state-of-the-art sensors in order to reduce the number of traffic accidents caused by drivers.

While traditional machine learning methods and algorithms have been extensively explored, the quality of the dataset assumes a paramount role in the industrial applications of machine learning. A superior dataset quality enhances the accuracy of detecting anomalies in driving behavior. This paper serves to underscore the validity of this proposed assertion. A thorough review of the literature on the detection of anomalies in human driving behavior was conducted, the details of which will be discussed in the section titled "Detecting Human Driving Anomalies".

Vokinger et al. (Vokinger et al., 2021) and Nazer et al. (Nazer et al., 2023) have demonstrated that diminishing bias contributes to enhanced prediction accuracy and the attainment of newly optimized outcomes. The proposed method for reducing bias involves a random selection of instances from a large pool of instances. For instance, one instance is chosen from a pool of 100 instances. Hence, a substantial pool of instances is required, a provision that conventional methods have not been able to fulfill. This paper will focus on how to generate the highest quality datasets in general, using the tachograph project as an example. The key distinction between the traditional and the proposed study lies in the emphasis placed on the quality of the dataset by reducing bias in datasets. In the existing projects on abnormal driving behavior detection, they focused on their own algorithms while this paper proposes a method to reduce a variety of data bias for achieving the highest prediction accuracy.

Our proposed method, as depicted in Fig. 1, is distinguished by its ability to capture data in real-time. We have installed our data capturing system in a total of 10,181 commercial vehicles in Japan. The system records driving behavior and transmits this information to a remote database center via a mobile network in real-time. This database, which contains data from 12,530 human drivers, forms the basis of our analysis. In other words, this paper scrutinizes the efficacy of the suggested approach in catering to the heterogeneity or diversity of various vehicles and drivers.

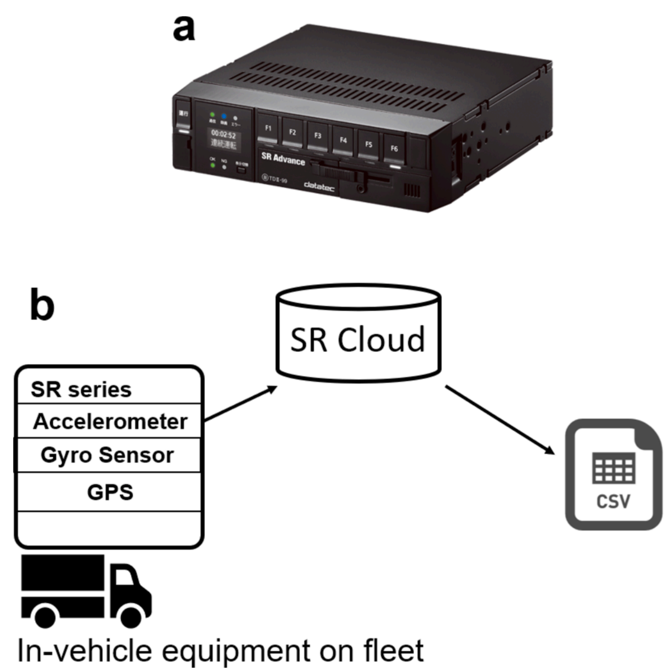

Fig. 1. Data capturing system for driving behavior by SR series. (a) Outward appearance of SR series (SR Advance, SR Connect, SR Dlite, and SR PocketII). (b) Schematic representation of driving behavior data capturing by SR series. The original driving data is captured at 10 Hz. For normal driving, the driving data for one minute is compiled in the SR itself and transferred to the cloud every minute. For abnormal driving, the data is transferred instantaneously to the cloud when the event occurs. The data transferred to the cloud is output in CSV file format and used for AI analysis.

To mitigate various data biases, we implemented a strategy of random selection from a vast pool of data instances. Our goal was to achieve a prediction accuracy of four orders of magnitude (9999). To this end, we identified 24 distinct types of abnormal driving behaviors. From the millions of instances available, we randomly selected 10,000 instances for each type of behavior. In other words, the effectiveness of random sampling with balanced and imbalanced datasets is examined.

To achieve the highest prediction accuracy of 99.99 % requires a large number of instances to significantly reduce bias in the dataset. Over the past 10 years, 40 billion instances of data have been collected for this purpose, of which 30 million instances consist of accident-induced driving data. In other words, the largest driver data in the world allows us to eliminate bothersome biases, improve data quality and achieve the highest prediction accuracy in the binary classification problem.

In our bias reduction experiments, we found that selecting data at random from billion instances reduces bias more significantly. The results validated that the less bias, the higher the prediction accuracy, which had never been reached in any project in the world. The proposed method can shed new light on machine learning in general and advance science and technology.

In this paper, we analyze accidental driving data using the world's largest dataset of driving data collected in Japan over the past 10 years. The dataset includes data from 10,181 commercial vehicles, 12,530 drivers, and 40 billion instances, including 30 million instances of accidental driving data. The data covers all of Japan. Our novel approach to detecting abnormal driving behaviors involves reducing various types of bias in the dataset, such as surrounding situations (including other cars, bikes, and pedestrians), location (including city or urban areas), driver information (including age and sex), and car types. The result will reveal that all kinds of bias with the proposed method were successfully reduced from the datasets. In other words, the proposed method can be applied to existing machine learning algorithms for improving their prediction accuracy by reducing various types of bias in the dataset.

The difference between previous researches and our study lies in that we have focused heavily on capturing real-time and enhancing the quality of data for machine learning with ensemble methods and bias reduction. In other words, no deep machine learning or GPU machines with fine-tuning are needed to achieve the highest accuracy in the proposed binary classification problem. As long as various types of bias in the dataset are reduced with the proposed bias-reduction method, the prediction accuracy of existing machine learning algorithms can be significantly improved. This represents this paper's assertion that reducing bias in the dataset is more important than improving machine learning algorithms or methods. In other words, in order to create the experimental datasets, data instances were randomly selected from 40 billion instances with conditions to reduce data bias of all kinds. The proposed bias-reduction method eliminates the need for expensive algorithms with GPU computing. Experimental results justify this claim. The method can improve existing machine learning in general.

Enhancing machine learning methodologies is crucial for improving prediction quality. However, due to the unique characteristics of datasets, there isn't a one-size-fits-all algorithm in general. Nevertheless, the proposed method for reducing bias can be effectively applied in data creation and bias reduction. This paper has demonstrated the efficacy of random sampling for detecting human driver anomalies when dealing with large, real-world datasets with diversity such as different vehicles and drivers using balanced and imbalanced datasets. The bias-reduction method is universally applicable to machine learning datasets, as it primarily focuses on minimizing bias within the datasets, irrespective of the specific machine learning algorithms employed. The training of the datasets is not conducted in real-time, but once the training is complete, it is capable of real-time detection of anomalies in human driver behavior.

In 10-fold cross-validation, the dataset is partitioned into 10 equal segments. The model undergoes training 10 times, with each iteration excluding a different segment for testing. This approach tends to yield a more reliable performance evaluation compared to 5-fold cross-validation, where the dataset is only split into 5 segments. This paper used 10-fold cross-validation for analysis of the predictive accuracy, precision, recall, and F1_score respectively.

Equilibrating a dataset simplifies the process of training a model, as it aids in avoiding the model's inclination towards a specific class. In other words, Making a dataset more balanced makes it easier to train a model because it helps to stop the model from being biased towards one class.

The Gini Index and Entropy are both methods used in decision trees to decide the best feature for splitting the data. They measure the impurity of an input set. The main difference between the Gini index and entropy is that the Gini index is a linear measure of impurity, while entropy is a logarithmic measure of impurity. This means that the Gini index is more sensitive to differences in the proportion of instances in each class, while entropy is more sensitive to the number of classes in the dataset. Hence, it is imperative that we evaluate the performance in conjunction with the proposed datasets.

1.1. Detecting human driving anomalies

A comprehensive review of literature was undertaken, focusing specifically on human driving behavior and the detection of anomalies therein. In 2013, Scientific American compiled five recipes for better driving from the perspective of human behavior (Gold, 2013). However, the proposed method can achieve better driving than human drivers. It can be applied to autonomous vehicles. With the advent of smart phones and inexpensive cameras, various systems for detecting abnormalities in driving behavior have been researched and proposed.

From 2017, machine learning approaches using datasets have been reported for detecting abnormalities in driving behavior (Russell et al., 2016; Taamneh et al., 2017). Since Russell's method only deals with steering information, it is impossible to detect all abnormalities in driving behavior (Russell et al., 2016). Taamneh's dataset consists of only 68 drivers on a driving simulator which cannot be used in practical applications (Taamneh et al., 2017).

Chuang-Wen You et al. proposed a new method for alerting drowsy and distracted drivers using dual cameras on smartphones (Chuang-Wen You et al., 2013). The proposed method did not cover all the abnormalities in driving behavior and was limited in its detection.

Rezapour et al. conducted an intensive study on driving under the influence of alcohol, fatigue, and distraction (Rezapour and Ksaibati, 2022). However, they only showed the simulation results of their models and methods. They have not verified anomaly detection for real-time driving behavior on real roads.

Baker et al. measured 26 drivers from near-infrared spectroscopy functional data, but the final dataset consisted of only 17 drivers (Baker et al., 2021). In the real world, their data is not large enough to detect anomalies in driving behavior in real time.

Di Liberto et al. studied 31 drivers using electroencephalography (EEG) signals for anticipating steering actions (Di Liberto et al., 2021). The dataset is also too small to be used to detect all abnormalities in driving behavior. In real life situations, their method is not a realistic approach using EEG signals from drivers.

Saiprasert et al. proposed the first intensive abnormalities driving detection using 40 vehicles and 25 drivers (Saiprasert and Pattara-Atikom, 2012). Their data is composed of seven types: speed, position, heading, rapid acceleration, harsh braking, rapid turning, and door status. But their dataset is still too small to be used to detect all abnormalities in driving behavior for machine learning.

Chunmei Ma et al. proposed a new tool to identify three kinds of dangerous behaviors: speeding, irregular driving direction change, and abnormal speed control (Ma et al., 2017). However, their data was captured in the limited small campus so that the total length of the trace is only less than 6 km. Its dataset is limited and cannot be used to detect abnormalities in real-life driving behavior.

Naqvi et al. measured 20 drivers using near-infrared (NIR) camera sensors for driver emotion classification (Naqvi et al., 2020). However, their dataset is too small to be used for detecting all abnormalities in driving behavior.

Hu et al. surveyed and investigated the driving behavior of vehicles, but did not present a large dataset for practical use (Hu and Zheng, 2021).

Kolekar et al. proposed several models on the Driver's Risk Field (DRF), a two-dimensional field that represents the driver's belief about the probability of an event occurring (Kolekar et al., 2020). However, their method has not been practically tested to detect all abnormalities in driving behavior in real time. In other words, this model has not been validated in a real city.

Tselentis et al. summarized five manufacturers telematic recording devices of driving characteristics (Tselentis et al., 2017). However, they did not indicate the accuracy of individual devices in detecting abnormal driving behavior. Their paper reviews the most common and well implemented methodologies related to usage-based motor insurance.

Vlahogianni et al. conducted intensive driving analytics using smartphones from perspective of algorithms, comparisons and challenges (Vlahogianni and Barmpounakis, 2017). However, their experiment used three vehicles, three different drivers, and only 180 h of travel where 2145 critical events were collected and analyzed. The experimental route was also limited to a small area around the university.

Many researchers have investigated the detection of abnormal driving behavior using machine learning algorithms (Shahverdy et al., 2020; Huang et al., 2023; Hu et al., 2020). While existing studies have attempted to automatically detect abnormalities in driving behavior with their algorithms, no satisfactory system has been developed to date due to a lack of data. There is currently no dataset large enough to detect all abnormalities in driving behavior. However, the proposed method can improve the accuracy of existing algorithms by reducing bias and improving the quality of data. It is important to note that the algorithms used in existing research and the proposed method for reducing bias in datasets are distinct from one another.

Xiao et al. introduced a mixed integer linear programming model for the Green Vehicle Routing and Scheduling Problem, considering various factors like vehicle heterogeneity, traffic congestion, and load impact on emissions (Xiao, 2016). Their model, combined with a hybrid algorithm, can reduce emissions by up to 8 %.

Yue et al. explored the impact of connected and automated vehicles (CAVs) on traffic incident management in metropolitan areas (Yue et al., 2023). They proposed a traffic assignment model that considers both macroscopic traffic assignment and microscopic driving behaviors. The model, combined with dynamic signal control policies, showed that road system stability was influenced by incident severity, signal control policy, and CAVs' penetration rate and distribution. The proposed incident management policy improved the recovery rate and system stability of road networks.

Zhang et al. introduced a novel approach, C2FDA, for cross-domain object detection in traffic scenes (Zhang et al., 2022). It includes three key components: an attention-induced coarse-grained alignment module, a feature selection module, and a category-induced fine-grained alignment module. The approach outperformed state-of-the-art methods in various domain adaptation scenarios.

Yu et al. introduced an intelligent driver model, IDM-LLDS, that considers location-dependent lighting conditions in long freeway tunnels (Yu, 2023). The model describes how lighting conditions affect traffic patterns. Results suggested adaptive tunnel lighting and sunscreens at tunnel portals can improve traffic conditions, with the former enhancing efficiency and the latter reducing traffic oscillation.

Zhang et al. introduced the Multi-attention based Hybrid-convolution Spatial-temporal Recurrent Network (MHSRN) for region-based traffic flow prediction (Zhang et al., 2024). The MHSRN leverages a hybrid-convolution module and a space-aware multi-attention module to capture spatial–temporal features. It outperformed other models in terms of mean absolute error and root mean square error on three real-world datasets.

Chen et al. introduced a novel traffic flow prediction model, V-STF, that uses visual methods to quantify macroscopic traffic flow indicators and considers density features and flow feedback (Chen et al., 2023). The model improved prediction accuracy during non-periodic peak hours by considering congested road conditions. It outperformed other methods in predicting sudden traffic flow changes.

Xu et al. presented a deep learning approach for driver identification and verification using psychological behavioral data (Xu et al., 2022). The model, which includes a fully convolutional network and a squeeze-and-excitation block, achieved an identification accuracy of 99.60 % and verification accuracy of 90.91 %. The system effectively differentiates drivers and detects imposters.

Xu et al. used a driving simulator to analyze the differences between novice and experienced drivers when encountering traffic violations (Xu et al., 2022). Results showed that novice drivers struggle with vehicle positioning and emergency operations, unlike experienced drivers who effectively combine steering and braking. The study provided insights for future advanced driving assistance systems.

The comprehensive literature review as of August 12, 2024, encompassing references (Russell et al., 2016; Taamneh et al., 2017; Chuang-Wen You et al., 2013; Rezapour and Ksaibati, 2022; Baker et al., 2021; Di Liberto et al., 2021; Saiprasert and Pattara-Atikom, 2012; Ma et al., 2017; Naqvi et al., 2020; Hu and Zheng, 2021; Kolekar et al., 2020; Tselentis et al., 2017; Vlahogianni and Barmpounakis, 2017; Shahverdy et al., 2020; Huang et al., 2023; Hu et al., 2020; Brownlee, 2020; Liu et al., 2020; Pierce et al., 2020; Banerjee and Chaudhury, 2010; Foulkes et al., 2020; Xiao, 2016; Yue et al., 2023; Zhang et al., 2022; Yu, 2023; Zhang et al., 2024; Chen et al., 2023; Xu et al., 2022; Xu et al., 2022; Chengula et al., 2024; Yang et al., 2024; Verma et al., 2024; Prakash et al., 2024; Alwhbi et al., 2024; Podda et al., 2024; Ghoreishi et al., 2023), reveals that the maximum verification accuracy for abnormal driving behavior detection using existing methods is 90.91 %.

While substantial research has been conducted in driver anomaly detection, we believe there are still significant gaps that our study aims to address. Specifically, although many studies have been conducted, they all utilize small datasets and focus primarily on detection algorithms. Our research seeks to demonstrate how a large pool of data instances can generate bias-free datasets to improve prediction accuracies, as the existing datasets are not sufficient to enhance prediction accuracy in anomaly detection. The proposed bias-elimination method is a universal approach that can be applied to any combination of detection algorithms.

This is primarily due to the use of smaller datasets with inherent bias. The datasets are formulated utilizing the bias-reduction method, which involves selecting instances from a large pool. The larger this pool of instances, the more potential there is to construct datasets with reduced bias. Improved prediction of abnormal driving behaviors can lead to a reduction in the occurrence of traffic accidents. The proposed method, which effectively minimizes bias, attains an exceptional verification accuracy of 99.99 % when applied to a dataset comprising 40 billion instances for mitigating traffic accidents.

As far as the literature review is concerned, existing studies have not reached 0.999 or higher on the four ratings of predictive accuracy, precision, recall and f1_score respectively due to bias of datasets. The paper will show the best prediction accuracy in the binary classification of abnormal and normal driving behavior.

Methods.

1.2. Robustness of the proposed method and dataset

K-fold cross-validation is commonly used for evaluating classifiers. A 10-fold cross-validation using random forest binary classification with four datasets was examined in this paper which is the most commonly used error-estimation method in machine learning (Brownlee, 2020).

1.3. Dataset preprocessing for reducing bias

Due to the frequent inclusion of data from severely unsafe drivers, the captured dataset may show strong bias or skew. Therefore, in order to control the bias, we randomly selected the data of severely unsafe drivers from large instances. Based on the valid rule of thumb on bias suppression, we captured 24 types of dangerous and cautious driving behavior, and randomly selected 10,000 instances from more than millions of instances of each. Abnormal data is composed of from 240,000 instances to 300,000 instances with four datasets. To achieve an accuracy of 4 orders of magnitude 9999, instances were randomly selected from a large instance population of more than 2 orders of magnitude. This is because the bias-free datasets were created from 40 billion instances of data.

As shown in Fig. 1, the developed data capturing system is capable of collecting abnormality data from 10,181 commercial vehicles and 12,530 drivers in real time in 0.1 s increments, 24 h a day, 365 days a year. The method we propose is notable for its real-time data capturing capabilities as shown in Fig. 1. A total of 10,181 commercial vehicles have been equipped with our data capturing system. The driving behavior is recorded and transmitted via a mobile network to a remote database center in real-time. This database, which comprises data from a total of 12,530 human drivers, serves as the foundation for our analysis.

We have recently added a new normal data capturing function to the data capturing system for machine learning. Normal data can be captured and collected every minute from 10,181 commercial vehicles and 12,530 drivers. The proposed real-time data capturing system is the first large-scale system of its kind in the world for practical applications such as detection of abnormalities in driving behavior. The contribution of this paper lies in that bias-free datasets were used and validated with the prediction accuracy, the precision, the recall and the f1_score respectively.

Fig. 1 shows the image of the SR series capturing driving behavior data. The original driving data is captured at 10 Hz. For normal driving, the driving data for one minute is compiled in the SR itself and transferred to the cloud every minute. For abnormal driving, the data is transferred instantaneously to the cloud when the event occurs. The data transferred to the cloud is output in CSV file format and used for AI analysis.

There are 35 types of dangerous and cautious driving behaviors, and we extracted 24 types of unsafe driving behaviors from each over 10,000 instances for machine learning. The remaining 11 types are extremely very rare with only 0.168 % of the total, and there is a lack of data available for machine learning. However, we may be able to add 11 more types in the future.

Table 1a and Table 1b show 35 types and 24 types of dangerous and cautious driving behaviors respectively.

In order to capture normal data from the data capture system, two conditions need to be met following the "easy-to-drive roads" designated in Japan:

Sqrt(ax*ax + ay*ay) < 0.2G where ax is front-back acceleration and ay is lateral acceleration.
vehicle speed over 10Km/h

In addition to above conditions, a normal data instance should not fit into any of the categories shown in Table 1b.

In this paper, the total abnormal data will consist of 240,000 instances while the normal data will consist of 240,000 instances, 250,000 instances, 270,000 instances, and 300,000 instances respectively. Therefore, we have experimented four datasets for examining the consistency of data and cross validation algorithm.

36 determinants in the datasets such as 24_24rand.csv, 24_25rand.csv, 24_27rand.csv and 24_30rand.csv file are as follows: Normal, 1Kj, 2Kj, Car ID, Employe Cod, Event Id, GX_MAX, GX_MIN, GY_MAX, GY_MIN, H, Latitude, Longitude, M, Min, SPEED_MAX, Sec, YR_MAX, YR_MIN, Y, accumulated wind direction X, accumulated wind direction Y, ccw, cw, d, direction of wv max, humidity, pressure, rain fall, temperature, visibility, weekday, wind run, wind velocity max, wind velocity min, and car type. Details of the 36 determinants can be provided upon request of the reader.

1Kj and 2Kj represent solar radiation (kJ/m2). 'Car ID' is used for car identification, while 'Employee Code' is used for driver identification. 'Event ID' is used for event identification. 'GX_MAX' and 'GX_MIN' denote the maximum and minimum values of front and rear acceleration, respectively, recorded 30 s before and after the event occurrence. Similarly, 'GY_MAX' and 'GY_MIN' represent the maximum and minimum values of left and right acceleration, respectively, recorded 30 s before and after the event occurrence. 'SPEED_MAX' refers to the maximum speed at the location. 'YR_MAX' and 'YR_MIN' denote the maximum and minimum angular speed, respectively, recorded 30 s before and after the event occurrence. 'Y' stands for year, 'M' for month, 'H' for hour, 'Min' for minute, and 'Sec' for second.

In random forest classification, the default criterion is Gini impurity. However, in the proposed random forest binary classification, entropy criterion and Gini impurity were compared for investigating the criterion effect.

Based on many experiments, we found a valid rule of thumb for improving the dataset: the more randomly we extracted instances from a population with a larger number of instances, the less biased the dataset would be. Therefore, in order to improve the dataset by reducing bias, we captured 24 types of dangerous and cautious driving behavior, and randomly selected 10,000 instances from more than millions of instances of each. Therefore, abnormal data is composed of 240,000 instances. Similarly, in each dataset, normal data of 240,000 instances, 250,000 instances, 270,000 instances and 300,000 instances were also randomly selected from large instances.

Table 1a
List for 35 types of dangerous and cautious driving behavior.

Type	Explanation	Event ID	Classification
Incident	Incident or almost incident	9998	–
Dangerous driving	Highly incident-inducing behavior	1	Double lane change and back
		2	Frequent lane change
		3	Lane change after hard braking
		4	Hard breaking after lane change
		5	Sudden lane change
		6	Steep turning during starting
		7	Sudden steering during turing
		8	Rapid acceleration during turing
		9	Steep turing during braking
		10	Hard braking during turning
		11	Steep turing with over angular velocity
		12	Curve approch with overspeed
		13	Hard breaking immediately after acceleration
		14	Hard braking immediately after starting
		15	Hard braking during braking
		16	Hard braking
		17	Straight backward after turing
Cautious driving	Incident-cautious behavior	101	Double lane change and back
		102	Frequent lane change
		103	Lane change after hard braking
		104	Hard breaking after lane change
		105	Sudden lane change
		106	Steep turning during starting
		107	Sudden steering during turing
		108	Rapid acceleration during turing
		109	Steep turing during braking
		110	Hard braking during turning
		111	Steep turing with over angular velocity
		112	Curve approch with overspeed
		113	Hard breaking immediately after acceleration
		114	Hard braking immediately after starting
		115	Hard braking during braking
		116	Hard braking
		117	Straight backward after turing

We would like to examine how the imbalance in the dataset affects the quality of the prediction accuracy. Therefore, we prepare for four experimental datasets such 24_24rand.csv, 24_25rand.csv, 24_27rand.csv and 24_30rand.csv respectively. For example, 24_27rand.csv represents 240,000 abnormal instances and 270,000 normal instances. A large number of instances are needed to reduce data bias of the dataset to improve the prediction accuracy.

Table 1b
List for 24 types of dangerous and cautious driving behavior.

Type	Explanation	Event ID	Classification
Incident	Incident or almost incident	9998	–
Dangerous driving	Highly incident-inducing behavior	5	Sudden lane change
		12	Curve approch with overspeed
		14	Hard braking immediately after starting
		15	Hard braking during braking
		16	Hard braking
		17	Straight backward after turing
Cautious driving	Incident-cautious behavior	101	Double lane change and back
		102	Frequent lane change
		103	Lane change after hard braking
		104	Hard breaking after lane change
		105	Sudden lane change
		106	Steep turning during starting
		107	Sudden steering during turing
		108	Rapid acceleration during turing
		109	Steep turing during braking
		110	Hard braking during turning
		111	Steep turing with over angular velocity
		112	Curve approch with overspeed
		113	Hard breaking immediately after acceleration
		114	Hard braking immediately after starting
		115	Hard braking during braking
		116	Hard braking
		117	Straight backward after turing

Fig. 1 illustrates the data flow process, capturing instances from 10,181 commercial vehicles and 12,530 drivers. Within these vehicles, original driving data is recorded at a rate of 10 Hz. This data, compiled over one-minute intervals, is then transferred to the cloud every minute. In the case of abnormal driving events, the relevant data is immediately transferred to the cloud as the event occurs. The proposed method and results can be used not only for abnormal driving detection but also for training the artificial intelligence of autonomous vehicles. Finally, the paper's assertion that reducing bias in the dataset is more important than improving machine learning methods will be verified.

This is one of the largest and the long-term experiments in the world for improving driver's behavior with abnormal driving behavior detection to mitigate traffic accidents.

Despite the challenges posed by the lack of public access to the dataset for independent validation, this paper presents a comprehensive methodology to address this issue. It details the algorithms, Python programs, parameters, and statistical methods used to capture driver behavior values, which are crucial for engineers to replicate this study. More accurate predictions not only reduce traffic accidents but also decrease mortality rates. Essentially, this paper proposes a novel method for generating bias-free datasets from a large pool of data instances that enable users to optimally predict driving anomalies. Consequently, those with access to similar data can replicate the study, thereby validating its findings.

2. Results

The datasets utilized in previous studies for detecting human driving anomalies were considerably smaller than the ones we used in our research. In essence, there is no established benchmark dataset for the analysis of detecting human driving anomalies. The predictive accuracy, precision, recall, and F1_score results from these studies are significantly lower than our findings. This observation does not imply that the existing algorithms are less effective; rather, their performance may vary due to the distinct characteristics of the datasets used. Furthermore, it is not scientifically sound to compare their algorithms using different datasets. The algorithm introduced in this paper is designed to mitigate bias, and it operates on both balanced and imbalanced datasets. The instances within these datasets are selected randomly. Furthermore, we have conducted an in-depth investigation into the efficacy of random sampling.

We have experimented four datasets with one balanced dataset and three imbalanced datasets such as 24_24rand.csv, 24_25rand.csv, 24_27rand.csv and 24_30rand.csv respectively. Individual datasets are composed of two classes: abnormal driving behavior and normal behavior instances. In order to observe the accuracy difference between different ratios of two classes in machine learning, four datasets were used. Remember that both class instances were randomly selected from 40 billion instances to reduce data bias.

For example, 24_27rand.csv file is a dataset that is composed of total of 510,000 instances with 36 determinants. For example, the file name of "24_27" represents the size of abnormal and normal instances: "24" for 240,000 and "27" for 270000. The test size in this experiment is 9 %, or 45,900 instances. The remaining 91 % of the dataset is used for training. There are no leaks between test data and training data. In this paper, 10-fold cross-validation was performed. Random forest binary classification with entropy criterion and Gini impurity was studied to investigate the effect of optimization criterion.

Python source program is based on random forest classification cross validation, rfcv.py as shown in Fig. 2. rfcv.py is a random forest program with entropy in decision trees. rfcvgini.py is a program with Gini in decision trees. The difference lies in the criterion option using "entropy" or "gini" in the single shaded source line that is shown in Supplementary materials.

The binary classification of random forest is to classify normal and abnormal states. The dataset is based on 24 types of abnormal driving behavior and the normal driving behavior as shown in Table 1b. There are four metrics to evaluate the performance of the proposed classification on abnormal driving behavior detection: accuracy, precision, recall and f1_score. Instead of showing the 10-fold cross validation result composed of 10 numbers per accuracy, precision, recall and f1_score, the average results will be given.

3. Entropy criterion

Criteria such as Gini impurity and entropy in random forest classification were compared based on the prediction accuracy.

The 10-fold cross validation result with balanced dataset, 24_24rand.csv for accuracy, precision, recall and f1_score was obtained by random forest with the entropy criterion as follows:

test_accuracy mean: 0.999944
test_precision mean: 0.999990
test_recall mean: 0.999898
test_f1_score mean: 0.999944

The 10-fold cross validation result with 24_25rand.csv for accuracy, precision, recall and f1_score is as follows:

test_accuracy mean: 0.999922
test_precision mean: 0.999986

```
# -*- coding: utf-8 -*-
# rfcv.py
import sys,codecs
import numpy as np
import pandas as pd
import sys
filename='24_27rand.csv'
trees=493
taco=pd.read_csv(filename)
taco.fillna(0,inplace=True)
crossv=10
print('filename:',filename)
print("data instances and parameters:",taco.shape)
print("cross-validation:",crossv)
print("trees:",trees)
y = taco['Normal']
X=taco.drop(['Normal','Event Id'],axis=1)
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score
from sklearn.metrics import *
from sklearn.model_selection import ShuffleSplit
cv = ShuffleSplit(n_splits=crossv, test_size=0.09, random_state=54)
clf=RandomForestClassifier(criterion='entropy',n_estimators=trees,
max_depth=None,min_samples_split=2, random_state=54,n_jobs=-1)
scores = cross_val_score(clf, X, y, cv=cv)
print(scores,scores.mean(),round(scores.std(),6))
```

```
# -*- coding: utf-8 -*-
# rfcvgini.py
import sys,codecs
import numpy as np
import pandas as pd
import sys
filename='24_27rand.csv'
trees=493
taco=pd.read_csv(filename)
taco.fillna(0,inplace=True)
crossv=10
print('filename:',filename)
print("data instances and parameters:",taco.shape)
print("cross-validation:",crossv)
print("trees:",trees)
y = taco['Normal']
X=taco.drop(['Normal','Event Id'],axis=1)
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score
from sklearn.metrics import *
from sklearn.model_selection import ShuffleSplit
cv = ShuffleSplit(n_splits=crossv, test_size=0.09, random_state=54)
clf=RandomForestClassifier(criterion='gini',n_estimators=trees,     max_depth=None,min_samples_split=2,
random_state=54,n_jobs=-1)
scores = cross_val_score(clf, X, y, cv=cv)
print(scores,scores.mean(),round(scores.std(),7))
```

Fig. 2. Python code for 10-fold cross validation: rfcv.py and rfcvgini.py.

test_recall mean: 0.999862
test_f1_score mean: 0.999924

The 10-fold cross validation result with 24_27rand.csv for accuracy, precision, recall and f1_score is as follows.

test_accuracy mean: 0.999915
test_precision mean: 0.999971
test_recall mean: 0.999868
test_f1_score mean: 0.999919

The 10-fold cross validation result with most imbalanced dataset, 24_30rand.csv for accuracy, precision, recall and f1_score is as follows.

test_accuracy mean: 0.999876
test_precision mean: 0.999944
test_recall mean: 0.999833
test_f1_score mean: 0.999888

4. Gini criterion

The result of the same validation with Gini impurity was calculated with balanced dataset, 24_24rand.csv:

test_accuracy mean: 0.999951
test_precision mean: 0.999995
test_recall mean: 0.999907
test_f1_score mean: 0.999951

The result with Gini impurity was calculated with 24_25rand.csv.

test_accuracy mean: 0.999913
test_precision mean: 0.99999558
test_recall mean: 0.999835
test_f1_score mean: 0.999915

The result with Gini impurity was calculated with 24_27rand.csv.

test_accuracy mean: 0.999893
test_precision mean: 0.999975
test_recall mean: 0.999822
test_f1_score mean: 0.999899

Finally, the result with Gini impurity was calculated with most imbalanced dataset, 24_30rand.csv.

test_accuracy mean: 0.999864
test_precision mean: 0.999962
test_recall mean: 0.999792
test_f1_score mean: 0.999877

The eight results show that there is no significant difference between the entropy criterion and Gini impurity. However, the carefully comparing eight results shows that the mean prediction accuracy of Gini impurity with balanced dataset of 24_24rand.csv was the best among eight results, calculated 0.999951 slightly better than the entropy criterion with the prediction accuracy to 4 digits 9999 of 0.999944.

The more imbalanced the dataset, the lower the prediction accuracy on a Gini or entropy criterion.

In other words, from a human perspective, the slight difference in prediction accuracy from 4-digit 9999 to 3-digit 999 makes a huge difference in the real world of reducing traffic accidents.

5. Discussion

Despite the challenges posed by the lack of public access to the dataset for independent validation, this paper presents a comprehensive methodology to address this issue. It details the algorithms, Python programs, parameters, and statistical methods used to capture driver behavior values, which are crucial for engineers to replicate the study. More accurate predictions not only reduce traffic accidents but also decrease mortality rates. Essentially, this paper proposes a novel, universal method for generating bias-free datasets from a large pool of data instances, enabling users to optimally predict driving anomalies. Consequently, those with access to similar data can replicate the study, thereby validating its findings.

At the inception of the project, the number of female drivers participating in road freight transport in Japan was quite limited. In other words, the current datasets used in this paper have limited determinants. However, for future iterations of the project, it is recommended to enrich the datasets with additional demographic details such as the age and gender of the drivers.

Equilibrating a dataset simplifies the process of training a model, as it aids in avoiding the model's inclination towards a specific class. In other words, making a dataset more balanced makes it more accurate to train a model because it helps to stop the model from being biased towards one class. The empirical evidence from experiments conducted on both balanced and imbalanced datasets corroborated this claim.

Random sampling from a large pool of instances and signal averaging, a technique used in GPS systems and other scientific applications, are both methods used to reduce noise and improve the accuracy of data. Random sampling is a statistical technique where a subset of data is chosen from a larger dataset (Pierce et al., 2020; Banerjee and Chaudhury, 2010). Each data point in the larger set has an equal opportunity to be included in the sample, ensuring that the sample accurately mirrors the larger population it's drawn from. In other words, the larger the population, the less bias (Banerjee and Chaudhury, 2010). While these two methods are used in different contexts (random sampling for large datasets and signal averaging for signal processing), they both aim to reduce noise and improve the accuracy of the data or signal. Random sampling from a large pool of instances is equivalent to signal averaging in their goals, even though they are applied in different domains. On the contrary, non-random sampling often results in estimates that are skewed (Foulkes et al., 2020).

A false positive implies that abnormal driving behavior is incorrectly classified as normal, whereas a false negative suggests that normal driving behavior is mistakenly identified as abnormal. In essence, a false negative poses a greater risk to road safety compared to a false positive.

5.1. Decision trees criteria: Gini impurity and entropy

This paper identifies the scope and aims, presenting a method for improving driver anomaly detection not through a detection algorithm, but by eliminating bias from datasets. While existing methods focus on detection algorithms, our bias-elimination method is a universal and independent algorithm that can be applied to diverse areas and applications with a large pool of data instances. The comparison of the two algorithms (Gini impurity and entropy) demonstrates that there is no significant difference between the two methods.

Gini impurity and entropy are used in decision trees for classification algorithms in random forest, each of which is a measure of the likelihood of an incorrect classification of new instances of a random variable. In general, entropy is of better quality than Gini impurity, but entropy calculations take more time than Gini impurity. However, in this paper, random forest with Gini criterion was the best classification in accuracy, precision, recall and f1_score respectively.

Gini was named after an Italian statistician. In the following, we explain how to calculate the Gini impurity and entropy, respectively.

The Gini impurity is calculated by the following Eq. (1):

$$\mathrm{GiniIndex} = 1 - \sum_{\mathrm{i}} \mathrm{p}_{\mathrm{i}}^2 \tag{1}$$

where p_{i} is the probability of class i.

Entropy is a measure of information that indicates the disorder of the features with the target. The entropy is calculated by the following Eq. (2):

$$\mathrm{Entropy} = -\sum_{\mathrm{i}} \mathrm{p}_{\mathrm{i}} \log_2(\mathrm{p}_{\mathrm{i}}) \tag{2}$$

where p_{i} is the probability of class i. The interval of the Entropy is [0, 1] while Gini Index has values inside the interval [0, 0.5].

In the proposed machine learning, random forest classification cross validation with entropy criteria and Gini impurity was used and compared for investigating the prediction accuracy, precision, recall and f1_score respectively. This paper showed that Gini impurity is slightly better than entropy criterion. The similar result was reported in other classification (Ma et al., 2017).

5.2. Reducing dataset bias

A universal bias reduction method is one that can be applied across various fields and types of data to minimize bias. One such method is randomization, which forms the basis of the proposed bias-elimination method. However, a significant disadvantage is that it requires a large pool of data instances. In our study, we focus on the randomization-based bias-elimination method due to its broad applicability and robustness. While other methods like stratification and matching are effective in specific contexts, they may not be universally applicable. Our proposed method, on the other hand, can be used across different datasets and detection algorithms, making it a versatile tool for bias reduction.

The proposed method is to reduce various types of bias in the datasets, such as surrounding situations (including other cars, bikes, and pedestrians), location (including city or urban areas), driver information (including age and sex), and car types. To improve prediction accuracy, it is necessary to eliminate as much as possible any bias in the datasets. Based on many experiments, we found a valid rule of thumb for improving the dataset: the more randomly we extracted instances from a population with a larger number of instances, the less biased the dataset would be. To achieve a prediction accuracy of 4 orders of magnitude 9999, 24 types of abnormal driving behavior, and randomly selected 10,000 instances from more than millions of instances of each. Similarly, normal driving behavior data is also randomly selected from more than 100 million instances to 240,000 and to 300,000 instances with four datasets. Bias-free datasets were successfully created from 40 billion instances of data over the past 10 years.

5.3. Effects of imbalanced datasets

Eight experiment results showed that balanced dataset of 24_24rand.csv with both entropy and Gini has the best accuracy. The more imbalanced the dataset, the lower the prediction accuracy on a Gini or entropy criterion. In other words, the most imbalanced dataset, 24_30rand.csv has the worst prediction accuracy with both entropy and Gini. Our novel approach demonstrated that it was successful in detecting abnormal driving behaviors with reducing various types of bias in the dataset, such as surrounding situations (including other cars, bikes, and pedestrians), location (including city or urban areas), driver information (including age and sex), and car types.

We have examined several sets of imbalanced datasets, consistently finding that the greater the imbalance, the lower the accuracy.

We recognize the necessity of demonstrating how our methodology can be adapted or validated using alternative datasets or simulated data. To address this, we propose the following steps. First, we provide detailed information on the parameter settings and procedures used in our experiments to ensure that other researchers can replicate our model training process and achieve similar results. By applying our methodology to alternative datasets, we can demonstrate its generalizability and robustness. We also offer a comprehensive description of the data preprocessing steps, including data sampling and creation techniques, to reduce biases and create datasets from a pool of data. This will enable other researchers to replicate our preprocessing pipeline accurately. Additionally, to further enhance replicability, we have made our code publicly available on a repository such as GitHub, allowing other researchers to access and run our code, thereby facilitating the replication of our experiments. This approach allows us to validate our methodology in a controlled environment. We describe the process of data simulation, including the assumptions and parameters used, to ensure transparency and reproducibility.

To optimize research time, we initially focused on small datasets to evaluate the effectiveness of random sampling in reducing biases. Our current conclusion is that creating a single instance in the dataset requires more than 100 instances to sufficiently reduce biases through random sampling. Moving forward, it is crucial to verify whether this threshold is adequate for reducing biases across different domains and datasets. This future work will help ensure the robustness and generalizability of our methodology.

Our paper emphasizes the creation of accurate and bias-free datasets, independent of the algorithms employed. In other words, the effectiveness of our data creation method remains unaffected by the choice of algorithms. This is why we opted to use the traditional random forest algorithm in this study. However, we recognize the importance of comparing our methodology with other algorithms to demonstrate its robustness and versatility. In future work, we plan to incorporate comparisons with a variety of algorithms, including but not limited to gradient boosting machines, support vector machines, and neural networks. By doing so, we aim to provide a comprehensive evaluation of our data creation method across different algorithmic approaches.

6. Conclusion

The results of 10-fold cross-validation showed that the proposed random forest binary classification with Gini impurity and 493 trees achieved the highest prediction accuracy of 0.999951 on average with the standard deviation of 0.000012 in detecting abnormal driving behavior with 240,000 abnormal instances and 240,000 normal instances. The robustness of the proposed method and dataset was justified by 10-fold cross validation with the prediction accuracy, precision, recall and f1_score respectively. Removal of bias can dramatically improve prediction accuracy. However, removing bias requires large data instances to generate a bias-free dataset. Experiments proved that the less bias, the higher the prediction accuracy. The proposed bias reduction method can be applied to any machine learning and, in general, can significantly improve prediction accuracy. The paper's assertion that reducing bias in the dataset is more important as well as improving machine learning methods was justified. However, to create unbiased instances, a large population of instances, more than two orders of magnitude, is required. Finally, the effects of imbalanced datasets were investigated. The experimented result showed that the more imbalanced the dataset, the lower the prediction accuracy on a Gini or entropy criterion. The proposed bias-reduction method can significantly reduce various types of bias in the datasets which has achieved the highest prediction accuracy. The proposed method can be applied to the existing machine learning in general.

This paper highlighted the exceptional accuracy in detecting anomalies in human driving behavior. Our future plans include implementing real-time alerts for drivers and providing them with safe-driving training to help mitigate potential risks. It may be necessary to devise intervention strategies to curtail driving risks.

Supplementary materials

https://github.com/y-takefuji/driving_behavior/blob/main/rfcv.py.

https://github.com/y-takefuji/driving_behavior/blob/main/rfcvgini.py.

CRediT authorship contribution statement

Yoshiyasu Takefuji: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Michiyasu Tano:** Resources, Methodology, Data curation, Conceptualization. **Masaya Shigehara:** Investigation, Data curation. **Shunya Sato:** Supervision, Data curation.

Acknowledgment

Not applicable.

Data availability

The data that has been used is confidential.

References

Alsaade, F. W., & Al-Adhaileh, M. H. (2023). Cyber attack detection for self-driving vehicle networks using deep autoencoder algorithms. *Sensors (Basel), 23*(8), 4086. https://doi.org/10.3390/s23084086. PMID: 37112429; PMCID: PMC10142967

Alwhbi, I. A., Zou, C. C., & Alharbi, R. N. (2024). Encrypted network traffic analysis and classification utilizing machine learning. *Sensors (Basel), 24*(11), 3509. https://doi.org/10.3390/s24113509. Published 2024 May 29.

Baker, J. M., Bruno, J. L., Piccirilli, A., et al. (2021). Evaluation of smartphone interactions on drivers' brain function and vehicle control in an immersive simulated environment. *Scientific Reports, 11*, 1998. https://doi.org/10.1038/s41598-021-81208-5

Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: Populations and samples. *Indian Journal of Psychiatry, 19*(1), 60–65. https://doi.org/10.4103/0972-6748.77642. PMID: 21694795; PMCID: PMC3105563

Jason Brownlee, (2020). How to Fix k-Fold Cross-Validation for Imbalanced Classification, https://machinelearningmastery.com/cross-validation-for-imbalanced-classification/.

Chen, J., Xu, M., Xu, W., Li, D., Peng, W., & Xu, H. (2023). A Flow feedback traffic prediction based on visual quantified features. *IEEE Transactions on Intelligent Transportation Systems, 24*(9), 10067–10075. https://doi.org/10.1109/TITS.2023.3269794

Chengula, T. J., Mwakalonge, J., Comert, G., Sulle, M., Siuhi, S., & Osei, E. (2024). Enhancing advanced driver assistance systems through explainable artificial intelligence for driver anomaly detection. *Machine Learning with Applications., 17*, Article 100580. https://doi.org/10.1016/j.mlwa.2024.100580

Chuang-Wen You et al., CarSafe app: alerting drowsy and distracted drivers using dual cameras on smartphones, MobiSys '13: Proc. of the 11th annual int conf on Mobile systems, applications, and services, 2013, pp 461–462 https://doi.org/10.1145/2462456.2466711.

Di Liberto, G. M., Barsotti, M., Vecchiato, G., et al. (2021). Robust anticipation of continuous steering actions from electroencephalographic data during simulated driving. *Scientific Reports, 11*, 23383. https://doi.org/10.1038/s41598-021-02750-w

Foulkes, A. S., Balasubramanian, R., Qian, J., & Reilly, M. P. (2020). Non-random sampling leads to biased estimates of transcriptome association. *Scientific reports, 10* (1), 6193. https://doi.org/10.1038/s41598-020-62575-x. PMID: 32277087; PMCID: PMC7148323

Ghoreishi SGA, Moshfeghi S, Jan MT, et al. (2023). Anomalous behavior detection in trajectory data of older drivers. In: *2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET)* (pp. 146-151). IEEE. 10.1109/HONET59747.2023.10374878.

Gleirscher, M., et al. (2017). From hazard analysis to hazard mitigation planning: The automated driving case. *NASA Formal Methods, 10227*. ISBN: 978-3-319-57287-1.

Gold, S. S. (2013). How to Be a Better Driver. *Scientific American Mind, 24*(1), 15. http://www.jstor.org/stable/24942333.

Hu, Z., Zhang, Y., Li, Q., & Lv, C. (2022). Human-machine telecollaboration accelerates the safe deployment of large-scale autonomous robots during the COVID-19 pandemic. *Frontiers Robotics Artificial Intelligence, 13*(9), Article 853828. https://doi.org/10.3389/frobt.2022.853828. PMID: 35494540; PMCID: PMC9043527

Hu, J., Zhang, X., & Maybank, S. (2020). Abnormal Driving Detection With Normalized Driving Behavior Data: A Deep Learning Approach. *IEEE Transactions on Vehicular Technology, 69*(7), 6943–6951. https://doi.org/10.1109/TVT.2020.2993247

Hu, X., & Zheng, M. (2021). Research progress and prospects of vehicle driving behavior prediction. *World Electric Vehicle Journal, 12*, 88. https://doi.org/10.3390/wevj12020088

Huang, C., Lv, C., Naghdy, F., & Du, H. (2020). Reference-free approach for mitigating human-machine conflicts in shared control of automated vehicles. *IET Control Theory & Applications, 14*(18), 2752–2763. https://doi.org/10.1049/iet-cta.2020.0289

Huang, X., Yun, P.o., Shuhui, W., & Zhongfeng, H. (2023). Abnormal driving behavior detection based on an improved ant colony algorithm. *Applied Artificial Intelligence, 37*, 1. https://doi.org/10.1080/08839514.2023.2216060

Ko, W., Park, S., Park, S., Jeong, H., & Yun, I. (2022). Development of freeway-based test scenarios for applying new car assessment program to automated vehicles. *PLoS One, 17*(7). https://doi.org/10.1371/journal.pone.0271532. PMID: 35862304; PMCID: PMC9302810 e0271532.

Kolekar, S., de Winter, J., & Abbink, D. (2020). Human-like driving behaviour emerges from a risk-based driver model. *Nature Communications, 11*, 4850. https://doi.org/10.1038/s41467-020-18353-4

Kuupiel, D., Jessani, N. S., Boffa, J., Naude, C., De Buck, E., Vandekerckhove, P., & McCaul, M. (2023). Prehospital clinical practice guidelines for unintentional injuries: A scoping review and prioritisation process. *BMC Emergency Medicine, 23* (1), 27. https://doi.org/10.1186/s12873-023-00794-x. PMID: 36915034; PMCID: PMC10010958

Lavin, A., Gilligan-Lee, C. M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Baydin, A. G., Sharma, A., Gibson, A., Zheng, S., Xing, E. P., Mattmann, C., Parr, J., & Gal, Y. (2022). Technology readiness levels for machine learning systems. *Nature Communications, 13*(1), 6039. https://doi.org/10.1038/s41467-022-33128-9. PMID: 36266298; PMCID: PMC9585100.

Liu, B., Li, C., Li, Z., et al. (2020). An entropy-based metric for assessing the purity of single cell populations. *Nature Communications, 11*, 3155. https://doi.org/10.1038/s41467-020-16904-3

Ma, C., Dai, X., Zhu, J., Liu, N., Sun, H., & Liu, M. (2017). DrivingSense: Dangerous driving behavior identification based on smartphone autocalibration. *Mobile Information Systems, 2017*, 15. https://doi.org/10.1155/2017/9075653. Article ID 9075653.

Naqvi, R. A., Arsalan, M., Rehman, A., Rehman, A. U., Loh, W.-K., & Paul, A. (2020). Deep learning-based drivers emotion classification system in time series data for remote applications. *Remote Sens., 12*, 587. https://doi.org/10.3390/rs12030587

Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digit Health, 2*(6), e0000278.

Paterson, C., & Picardi, C. (2023). Hazards help autonomous cars to drive safely. *Nature, 615*(7953), 594–595. https://doi.org/10.1038/d41586-023-00798-4. PMID: 36949332.

Pierce, M., McManus, S., Jessop, C., John, A., Hotopf, M., Ford, T., Hatch, S., Wessely, S., & Abel, K. M. (2020). Says who? The significance of sampling in mental health surveys during COVID-19. *Lancet Psychiatry, 7*(7), 567–568. https://doi.org/10.1016/S2215-0366(20)30237-6. Epub 2020 Jun 2. PMID: 32502467; PMCID: PMC7266586.

Podda, A. S., Carta, S., & Barra, S. (2024). Artificial intelligence methods for smart cities. *Sensors (Basel), 24*(8), 2615. https://doi.org/10.3390/s24082615. Published 2024 Apr 19.

Prakash, J., Murali, L., Manikandan, N., Nagaprasad, N., & Ramaswamy, K. (2024). A vehicular network based intelligent transport system for smart cities using machine learning algorithms. *Scientific reports, 14*(1), 468. https://doi.org/10.1038/s41598-023-50906-7. Published 2024 Jan 3.

Rezapour, M., & Ksaibati, K. (2022). Identification of factors associated with various types of impaired driving. *Humanities & Social Sciences Communications, 9*, 26. https://doi.org/10.1057/s41599-022-01041-7

Russell, H. E. B., Harbott, L. K., Nisky, I., Pan, S., Okamura, A. M., & Gerdes, J. C. (2016). Motor learning affects car-to-driver handover in automated vehicles. *Science Robotic, 1*(1). https://doi.org/10.1126/scirobotics.aah5682. eaah5682.

Saiprasert, C., & Pattara-Atikom, W. (2012). Computation of driver safety rating using in-vehicle data recorder: Case study of bangkok public transportation. In *2012 15th International IEEE Conference on Intelligent Transportation Systems* (pp. 1121–1126). IEEE. https://doi.org/10.1109/ITSC.2012.6338776.

Shahverdy, M., Fathy, M., Berangi, R., & Sabokrou, M. (2020). Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications, 149*, Article 113240.

Taamneh, S., Tsiamyrtzis, P., Dcosta, M., et al. (2017). A multimodal dataset for various forms of distracted driving. *Sci Data, 4*, Article 170110. https://doi.org/10.1038/sdata.2017.110

Tselentis, D. I., Yannis, G., & Vlahogianni, E. I. (2017). Innovative motor insurance schemes: A review of current practices and emerging challenges. *Accident; Analysis and Prevention, 98*, 139–148. https://doi.org/10.1016/j.aap.2016.10.006

Verma, M. E., Bridges, R. A., Iannacone, M. D., et al. (2024). A comprehensive guide to CAN IDS data and introduction of the ROAD dataset. *PLoS One, 19*(1). https://doi.org/10.1371/journal.pone.0296879. e0296879. Published 2024 Jan 22.

Vlahogianni, E. I., & Barmpounakis, E. N. (2017). Driving analytics using smartphones: Algorithms, comparisons and challenges. *Transportation Research Part C-emerging Technologies, 79*, 196–206. https://doi.org/10.1016/j.trc.2017.03.014

Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S. (2021). Mitigating bias in machine learning for medicine. *Communication Medical, 1*, 25. https://doi.org/10.1038/s43856-021-00028-w

Xiao, Y., et al. (2016). The heterogeneous green vehicle routing and scheduling problem with time-varying traffic congestion. *Transportation Research Part E: Logistics and Transportation Review, 88*, 146–166. https://doi.org/10.1016/j.tre.2016.01.011

Xu, J., Guo, K., & Sun, P. Z. H. (2022). Driving performance under violations of traffic rules: Novice vs. Experienced drivers. *IEEE Transactions on Intelligent Vehicles, 7*(4), 908–917. https://doi.org/10.1109/TIV.2022.3200592

Xu, J., Pan, S., Sun, P. Z., Park, S. H., & Guo, K. (2022). Human-factors-in-driving-loop: Driver identification and verification via a deep learning approach using psychological behavioral data. *IEEE Transactions on Intelligent Transportation Systems, 24*(3), 3383–3394. https://doi.org/10.1109/TITS.2022.3225782

Yang, G., Ridgeway, C., Miller, A., & Sarkar, A. (2024). Comprehensive assessment of artificial intelligence tools for driver monitoring and analyzing safety critical events in vehicles. *Sensors (Basel), 24*(8), 2478. https://doi.org/10.3390/s24082478. Published 2024 Apr 12.

Yu, S., et al. (2023). Understanding traffic bottlenecks of long freeway tunnels based on a novel location-dependent lighting-related car-following model. *Tunnelling and Underground Space Technology, 136*, Article 105098. https://doi.org/10.1016/j.tust.2023.105098

Yue, W., Li, C., Wang, S., Xue, N., & Wu, J. (2023). Cooperative incident management in mixed traffic of CAVs and human-driven vehicles In: *IEEE Transactions on Intelligent Transportation Systems*. doi: 10.1109/TITS.2023.3289983.

Zhang, H., Luo, G., Li, J., & Wang, F.-Y. (2022). C2FDA: Coarse-to-fine domain adaptation for traffic object detection. *IEEE Transactions on Intelligent Transportation Systems, 23*(8), 12633–12647. https://doi.org/10.1109/TITS.2021.3115823

Zhang, X., Wen, S., Yan, L., Feng, J., & Xia, Y. (2024). A hybrid-convolution spatial–temporal recurrent network for traffic flow prediction. *The Computer Journal, 67*(1), 236–252. https://doi.org/10.1093/comjnl/bxac171. bxac171.

Zhao, X., et al. (2023). Indirect shared control strategy for human-machine cooperative driving on hazardous curvy roads. *IEEE Transactions on Intelligent Vehicles, 8*(3), 2257–2270. https://doi.org/10.1109/TIV.2023.3242762