Machine Learning Models for Predicting Driver Behavior in Mixed Traffic Scenarios

By Dr. Michael Hitchens

Associate Professor of Cybersecurity, University of Newcastle, Australia

1. Introduction

[1] Machine learning has been widely used in recent years to predict human driver behavior. Various methods and techniques have been applied in different ch severities such as driving behavior prediction, intention prediction, and trajectory prediction, all of which are based on machine learning models. However, due to advancements in intelligent vehicles and artificial intelligence (AI) technology, the interactions in mixed traffic scenarios are becoming more complicated. Although considerable contributions have been made to enable human-like, decision-making behaviors in intelligent vehicles, e.g., improving motion planning and control, there is still a lack of investigation of adopting comprehensive machine learning models to accurately predict and model human driver behaviors. Therefore, developing machine learning models for predicting complex driver behaviors in mixed traffic scenarios with different densities is still a challenging open problem. In this work, we are targeting to build state-of-the-arts machine learning models for predicting and modeling driver behavior in mixed traffic scenarios. And our final target is to develop an intelligent vehicle that can coexist with human drivers harmoniously.[2] Automated driving systems aim at predicting the intentions or actions of human road users to plan collision-free trajectories. Scientific efforts to predict future trajectories of human drivers and pedestrians are based on models learned in a variety of ways, from simple linear models to deep learning approaches. Although such approaches generalize trajectories exhibited in data well, it is still necessary to ensure that the entire spectrum can be reproduced in prediction, even when data become scarce or absent. For instance, rare but safety-critical trajectory types like braking from an unambiguous lane change are not properly covered by these approaches. Therefore, several works have tried to enhance trajectories in various ways to improve their behavior. We refer to such an approach as a model that produces plausible trajectories1.

1.1. Background and Motivation

Improved driver behavior prediction and understanding of individual driver responses to different behaviors can offer many beneficial outcomes, including improved fuel consumption, better traffic flow, and increased safety by making human—driven and AI—operated vehicles work together more effectively [3]. Being focused on such a task of predicting how individual drivers will respond to different behaviors performed by a convoy leader, we primarily intend to make two novel contributions to the field of driver behavior prediction.

Machine learning (ML) is getting more attention in transportation, particularly for predicting and modifying driver behavior [1]. In novel transportation scenarios like mixed traffic (cooperative autonomous and human-driven vehicles), the accurate prediction of driver behavior can significantly improve autonomous car performance and safety. One of the many scenarios in this category, for example, is a moving obstacle whose future path and behavior are unpredictable to an AI driven car. Thus, what is of fundamental importance here is an AI model that predicts where a driver in a human-operated vehicle in a convoy would go and how they respond to the actions of its AI operated lead vehicle [4]. In this project, we intend to predict how individual drivers in a convoy will behave in response to a sequence of the dynamically evolving, real-time motions of its passively observed lead vehicle. In the context of this introduction, a typical type of motion we could be interested in forecasting comes from the sudden (real-time) evasive movement of a lead vehicle. This analysis attribute is generic in the sense that a similar analysis for other driving behaviors, such as overtaking, turns, and velocity changes, could be performed using the same data source.

1.2. Research Objectives

The prediction of behavior of the driver space in mixed traffic is fundamentally important for the improvement of the levels of autonomy of the vehicles as well as to generate information to other relevant stakeholders in the sector, such as vehicle manufacturers and managers and public transportation companies [5]. The proposal to contribute to the following pillars mentioned is the creation of an agent using reinforcement learning to optimize a light signal in an intersection, highlighting the versatility of the modeling of the space of behaviors in this area in the context of traffic engineering. The specific objective of the present study to contribute to the comprehension of this scenario is the determination of the main critical variables related to the driving behavior to be observed and aimed to be modeled by the approach mentioned before. Schneider and coauthors showed an advanced perception architecture capable of multi-modular perception of traffic from a visual input, enhancing the capabilities of single modular deep-learning-based approaches to enable self-supervised learning of critical vehicle-to-vehicle interactions [6]. As the literature about models for predicting driving behavior consisting in modeling the choices of paths are still narrow and as more and more manufacturers make their systems available in open source, these set of methods may be applied to models in scenarios whose real-world fluctuations are non-linear.

Modeling and predicting driver behavior in mixed traffic scenarios are key steps for the development and deployment of automated driving systems and advanced driver assistance systems. This paper aims to present the main objectives of the study, which include the identification of specific and critical variables or features of the driving behavior to be observed in driver modeling in the studied scenario [7]. For the diversity of conditions analyzed in the scenarios considered in the study, it is expected that the modeling and simulation of the driver's behavior can help in understanding the level of confidence and predictability of the vehicles and the paths used by them. Driver modeling can also help in understanding the potential risks of driving conflicts and the definition of vehicle strategies in order to reduce driving behaviors considered risky.

1.3. Scope and Limitations

-In general, the validity of our model predictions will depend on the stability of driving behavior over longer timescales compared to the prediction horizon. This assumption is always made when dealing with scarce input data. We believe it is reasonable to assume that sub-minute or minute-scale behavioral patterns are useful for driver assistance, as this timescale is relevant for the prediction of collision events. Nevertheless, we encourage more research focused on the relevance of longer-lasting patterns.

-The primary function of our machine learning models is to predict the driving behavior of vehicles on highways in San Francisco. All the models consider scenarios in which different degrees of traffic congestion occur. The prediction information generated from the ML models is intended to support driver assistance functions installed in vehicles equipped with the necessary sensors. It is important to note that all vehicles are equipped with cameras that serve as the main input data to our models. That is, our models do not take into account all the contextual information that might be available to the onboard sensors. The reasoning behind this is twofold: first, we aim for a general system whose computational demand for onboard data processing does not exceed the capabilities of currently available computer vision systems for autonomous driving. Second, we want to showcase the predictive power of our model components, irrespective of the available data. Stated differently, by using only camera information, we force ourselves to extract the relevant information for our predictions using on a very limited number of parameters.

The scope and limitations of this work are as follows:

[4] [8]

2. Literature Review

The future implementation of automated vehicles poses a major revolution for the traffic safety field, which may lead to an estimated 90% road casualties reduction [9]. Nevertheless, the appearing mixed traffic scenarios require trusting automated vehicles to make decisions in the presence of different road users in order to avoid accidents and to be properly integrated into the traffic flow. Facing this situation, different research lines have already chemically addressed the driver control function. A first group establishes speed trajectories or lane changes for automated vehicles based on traditional mathematical driving models like Sazonov/Car Following (SCF) model or Intelligent Driver Model (IDM) models, with the objective of avoiding accidents through buffer reception–time distances when fast-detecting traffic hazards. However, these models simplify or abstract the complete process of decision-making by the driver a lot, reducing the use of datasets to fine tune the models, and may not be able to correctly interpret the position of the traffic hazards in a real context [6]. On the one hand, studies have used machine learning models in order to use learning methods of control for automated vehicles, simplifying the way that they adapt to their environment in a more natural way by predicting the behavior of humans and also obtaining better performance in many scenarios by using optimal prediction times, especially for sudden accidents with stationary or pedestrian traffic. On the other hand, studies have used machine learning models in situations coupled to traditional mathematical driving models when mixing human drivers and autonomous vehicles [10]. This implies that while many of the decision-making steps are exploited by traditional or more manually codified driving models in order to obtain performance for sudden avoidance situations, machine learning models are used to adapt the vehicles to the current context, adapting the trajectory of the vehicles according to the real behavior of the road users. Once the criticality is evaluated, the selected trajectory will be based on avoiding accidents by predicting the evasive trajectories of the surrounding vehicles, calculating the maximum time required for these courses to reach critical spaces of the trajectory.

2.1. Driver Behavior Prediction

Motorists in free-inflow traffic head toward a merging bottleneck. To merge relatively efficiently, they can synthesize traffic into a zipper fashion where vehicles from one lane and the other take approximately equal turns. Or they can speed up and squeeze themselves into the lane into which they want to merge and then slow down to not become too close to the vehicle in front of them [4]. Or they can wait in the merging lane until there are no more vehicles in the right lane and then switch. To obtain data for merging strategies, the authors have observed merging on the ramps and lateral motorway sections that RC-9 crosses.

On highways, drivers who would each keep fixed following distances under free-flowing conditions appear to collaborate in regulating mutual distances near merging bottlenecks. Prediction of driver behavior near mergers is thus a crucial use case for a variety of purposes in traffic engineering, from traffic simulation to Intelligent Transportation Systems [2]. In this article we first describe a model for merging and then lay out possibilities for real-time prediction based on Bluetooth and cellular data.

2.2. Machine Learning in Traffic Scenarios

To produce realistic and fairly reasonable estimates and predictions, traffic models combining human models and AI-ML will need to be developed. The taxonomy combines the beliefs of AI/ML to define a comprehensive space to be considered within the literature on driving psychology and aggregate a broad range of driving decisions models. For accurate projections of driver behavior, end-to-end machine learning models have been increasingly gaining popularity. However, studies in traffic psychology, human behavior, and cognitive sciences show that several factors are contributing to driver behavior. These factors market drivers decision through understanding of uncertainties, higher thinking skill, and natural instincts. Model design or its introduction is often influenced by a given problem's complexity and the level of structure a data can offer.

[1] Traffic scenarios, especially those that are semi-structured, i.e. containing a mix of different types of agents, are difficult to model and predict. However, with the popularization of advanced driver assistant systems (ADAS) and level-3 and above autonomous vehicles, a driving agent or a fully autonomous vehicle has to understand the driving environment modeled by such complex systems. Thus, there is a lot of research into incorporating machine learning models to such datasets capable of handling this situation and therefore introduce lane prediction-attention to reasoning over the observed agents and traffic regulations. [7] Most driver models possessed driving and vehicular movement as the core entities, while factors such as infrastructure, surveillance, and law were given little attention. Through a comprehensive review of these models this article identifies a taxonomy of machine learning applications used in traffic behavior prediction.

3. Data Collection and Preprocessing

Given the purpose of this study, we implement two predictive models based on a logistic regression model and a random forests model. First, a descriptive analysis of the input features is conducted. The descriptive analysis provides a high-level understanding of the underlying structure of the data and thus helps inform data preprocessing [1]. Handling of NULL and infinity/NaN values, data standardization and normalization, handling categorical features, and feature selection and engineering are main steps in data preprocessing. Violin and box plots are effective visualization methods for variables that are highly correlated and show similar distribution patterns.

A critical step in this research is to use connected vehicle trajectory data to predict a driver's future behavior based on past behavior [11]. Drivers who engage in hazardous and aggressive driving behaviors, such as speeding, may contribute to the risk of a large number of accidents. To predict a driver's future aggressive speed behavior, or willingness to engage in speed-related behaviors, a machine-learning method is used to build predictive models based on driving data [6]. We define speeding-related behavior as behavior that can lead to potentially increased severity of crashes when the drivers do not follow traffic safety rules. Such behavior includes individual actions like hard acceleration and hard deceleration. Models built using fundamental speed-related behavior data have the potential to serve as predictive tools for understanding the risk and potential harm posed by aggressive drivers.

3.1. Types of Data Collected

The AD-Curator dataset was collected under different movement intentions by the preceding vehicle approaching the intersection— going straight, left/right [12]. The data collection system triggers data collection for a normal vehicle movement but also with an emphasis on safely driving and the importance of looking at potentially hazardous traffic conditions like a steadily approaching con/con turn. This is a simple and effective mechanism to simulate in-field data for collecting such con/con or con/join maneuvers safely, removing any ethical issues related to recording real-life interactions in such situations. Furthermore, we have had access to vehicle movement intention labels which is not present in the labeled Safer Driver Scheme dataset and has been unavailable for (in-group) past con/con datasets.

Leveraging the Safer Driver Scheme, [13] an explicit collection of driving data characteristic of good driving has been gathered, composed of Hazard Test incident data as well as normal driving data using Mobileye Safety Suite data, for a set of 2,764 drivers. From the Safer Driver Scheme data, data collection triggered by hazard events such as hard braking, sudden braking, dangerous tailgating, near miss from an adjacent lane, sudden lane change without indication, and speeding are used to form the PhBADT dataset. Mobileye driving data was collected at 1–2 Hz under different trigger thresholds for different Mobileye event types—Harsh Braking/Harsh Acceleration is the most frequency occurring event and triggers data collection at lower thresholds. Furthermore, the Harsh Braking and Harsh Acceleration events recorded will be biased towards more severity which is unlike what can be typically believed in the real world. Thus, the Mobileye safety event and normal driving data is leveraged to provide a much larger variation of 'good behavior' than the Safer Driver Scheme data alone.

3.2. Data Cleaning and Feature Engineering

Moreover, the profile of the 'Test Pilot' agent, including one expert driver estimator, is developed in Simulink supporting different type of agents and human behaviors. The Trancat, SqrReaction, and MarkovChain components can change the type of driver behaviour between normal, aggressive, or drowsy. The agent uses the radar and ElItalon sensors to collect environmental information while the LanesChange and ObBoundOutLiteral detect deceleration events, playing together with the StopGOSwitch to manage these situations. The simulator is managed by an ad-hoc developed Mission Manager block that orchestrates the different behaviors of the simulated system including scenario management, traffic regulation, and sending data to the EvNm Box for massive data analysis in the context of a

DLRC task. The mission manager also simulates the dynamical interaction in real-conditions with vehicle components of the DLRC proving ground.

For every predictive model, the first step is to prepare the required data, that is, the preprocessing step. A comprehensive literature review and an overview of different features were performed, as shown in the previous section. Here, the processed data and features used to infer several predictive models, primarily support vector machine (SVM), Ada-boost classifier, and logistic regression, are displayed . The data represent the participant number, longitudinal distances and speeds of the lead vehicles, the vehicle mass, the acceleration, the road type, the previous speed, the previous acceleration, the drowsiness estimation, and the presence of electronic devices, the wireless devices in particular (phone or music devices). Various transformations of the data have been attempted to have an intu- itive interpretation, and their impact on the final performance is also addressed. Using 2 techniques for feature selection, several models are trained. Using data im- balances, principles were followed by a synthesis of new examples, such as overparametrization, undersampling, and the SMOTE algorithm, to create synthetic data points . A3200v, a vehicle model belonging to the DILab reference database, is adopted, and it is developed in the context of an example application by the DLRC, including the control of the engine, clutch, and brakes exerting continuous dynamics actions.

4. Machine Learning Models

In this paper, we predicted the lateral motion of preceding vehicles in a naturalistic driving scenario, which is more complicated and uncertain than that of a lab scenario [12]. We have found that RNN-based models are more powerful than similar regression-based models and could generate realistic distributions from a large-scale dataset of naturalistic driving. Three prediction methods for a VRP are highlighted. For a big training dataset, the algorithm may be retrained with newer parameters and new features such that it represents the latest situation. To optimize and improve the models, driver signaling and SEM data can be used to further improve the performance of the model. A facility implementing the proposed prediction method is an FCW (forward collision warning) system that warns drivers of the potential for a rear-end collision. The data used in this study were collected into the defined dataset through SEM and simulator data as well as deep learning techniques for image processing (binarization, segmentation, and short-term and long-term motion) as well as regression and multivariate forecasting [4]. The obtained data have been used to test the designed predictors and reveal the optimal values in several studies. The optimal parameters are used to test the obtained predictors' performance on unseen data to prove their generalization over different scenarios. On the contrary, the machine learning-based-method can learn from limited data as compared to the predicted kinematic and behavioral models. Or from the wide variety of training data such as the drivers' behaviors, which will lead to more real and more general case studies. Also, the measures recommended by the projected predictors could be evaluated through simulations and real car experiments.

We use Machine Learning (ML) to predict long-term maximum speed and lateral position of the subject vehicle by inferring the speed trajectories of the Preceding Vehicles (PVs) from the Left and Right Rear (LR/LRR) Radar data [14]. Radar measurements of surrounding traffic in mixed scenarios are used to predict the long-term behavior of n (> 400) subjects over a distance of up to k (3 – 4 incidents are visible in the GPS data of the subject vehicle) kilometers. We introduce the Radar-to-Feature (R2F) framework, where we model traffic participants from their Radar data, and then remove the LR/LRR Radar measurements of each PV, replacing them with their estimated features. Random forests models use various derived probabilities as well as the PV-estimations as inputs for a network which weighs both the estimates and the contextual features .

4.1. Decision Trees

Creating informative models for businesses over various industries is a crucial need. Models are a representation of real-world phenomena built through mathematical relationships, that allow human analysts to understand complex real-world dynamics using simple mathematical laws [6]. Machine learning has become an appealing and suitable tool for building situational awareness models for future scenarios. In the traffic industry, for instance, an increasingly large number of research articles are addressing the development of machine learning models for accident prediction. Accidents present large financial costs of which approximately 3% to 5% of gross domestic product (GDP) is lost annually through accidents. Furthermore, accidents are recognized as the largest single cause of structured family tragedies every year. For these reasons and the potential applications that these predictive models enable, this is a topic of strong interest in the ambient intelligence community.

[15] Traffic safety prediction is a prompt and efficient tool for improving traffic management, mainly based on the identification of high-risk traffic accidents. In this regard, this study sets to predict the severity of crashes using data mining techniques. More specifically, machine learning techniques including support vector machine, decision tree, random forest, gradient boosting, and k-nearest neighbor were considered. The results demonstrated that machine learning approaches are comprehensive and efficient for traffic safety research. The results suggest that random forest performs better than other algorithms. Several risk factors were explored here, not only typical symptoms such as traffic conditions, weather, and accident location[16] Tachograph braking data from a fleet of emergency ambulances is used to develop a machine learning algorithm for modelling road traffic incidents. We directly address the typical challenges of traffic accident prediction due to data scarcity using a non-silent, non-emergency prescriptive driving task. The potential of different features used to describe each braking event is tested and annotated data available at the intention to stop are used to refine the data. An artificial over-sampling method is used to further counter the effects of class imbalance in our dataset. We encourage the collection of a higher fidelity driver attention data source such that the model may also warn of self-inflicted incidents in addition to environmental factors

4.2. Random Forests

To investigate the possibilities of driver signature prediction by the machine-learning algorithm, the prediction results for the three tree-based regression algorithms were compared. We conduct a set of experiments to build the prediction models with the error observed as the main interest of our study. Thus, the main objective of our study is two-fold. First of all, this study aims to analyze the potential of different tree-based regression algorithms using different features of the same network. Secondly, the prediction of the results for different features of the data is evaluated in the three tree-based algorithms under observation [17].

In this study, a Driver Signature model was proposed to simulate traffic dynamics of heterogeneous traffic flow in mixed traffic scenarios in cellular automaton (CA) framework. A driver's behavior was determined by a set of driving feature properties, which were formulating as the driver signatures [15]. A modified Intelligent Driver Model (IDM) and a machine-learning based model called Affine Driver Model were used to simulate human-driven vehicles and robot-driven vehicles, respectively [6].

4.3. Support Vector Machines

There are two different types of SVM—linear SVM and nonlinear SVM, classified by the kernel function used in each. Linear SVM is widely used in several applications, including traffic flow forecasting and transportation data collection. A primary step in handling traffic data is to perform feature extraction in reducing dimensionality. As a common feature extraction method, the principal component analysis (PCA) is used to reduce the feature dimensions into different low-dimensional feature vectors, which are named principal components. Next, the reduced features are predicted using a linear single support model. Some feedback features are used as backups with the linear single support model. If the linear prediction features are unsuccessful, the feedback features are applied in the nonlinear single support vector machine (SVM) model to improve prediction performance. After comparing the simulation results with the common prediction models, it can be clear that the proposed L-FSVM model achieves significantly better predictions than the other models, especially in highly fluctuating traffic data prediction, and the predicted value more accurately reflects real traffic flow conditions.

Support vector machines (SVM) are a famous machine learning decision support tool nowadays, which are increasingly used in many research fields, including machine learning, computer vision, time series forecasting, medical and biological fields, and traffic research. SVM advances research in traffic flow forecasting, network security scenarios, behavior identification, and other fields. It works by calculating appropriate support vectors and separating the hyperplane space to the maximum extent. As a general statistical learning strategy, SVM gains strong generalization capability and efficiently solves large-scale complex small-sample learning problems [1,2]. In crowdsourcing-based transportation applications, SVM is implemented for performing well in busy transportation systems, including nonlinear short-term traffic flow forecasting and complex traffic data classification problems.

4.4. Neural Networks

The architecture-based model enhances the core of our model network through the highway structure [18]. The traditional LSTM network neglects spatial and temporal influence. In contrast, the highway network model carries out information transformation by considering both path continuity and discontinuity. This is conducive to the spatial and temporal influence of sequences in the training process. Although the traditional capsule neural network has a good performance, existing research still has some problems, such as poor generalization ability and insufficient recognition of traffic congestion. Reinforcement learning provides a way to address this problem in traffic flow detection [34, 35]. It can learn a model through continuous interaction with the surrounding environment, effectively binding traffic scene semantics to a specific model structure. Meanwhile, a combination of capsule neural network and reinforcement learning is also able to improve the generalization ability of the network and further increase the recognition rate of traffic jams.

The long short-term memory neural network (LSTM) was proposed in 1997 by Hochreiter and Schmidhuber and has become widely used in the field of time series prediction, such as financial transactions, traffic speed, and energy consumption, as well as natural language processing [19]. Gers and Schmidhuber further improved the LSTM model to the forget LSTM by introducing an adaptive control mechanism. According to the commonly used criterion in the model algorithm design of field neural networks, the adaptive control mechanism considers the context information of the model network and has a significant influence on the overall performance of the model. Compared with the traditional LSTM, the forget LSTM model adds increased error feedback to better consider the impact of different modules in the future and adjust the optimization strategy of the model.

5. Evaluation Metrics

To evaluate our predictor performance, we can use different evaluation metrics, observing driver reaction type and its timing as the evaluation criteria as would provide valuable insights to the ADS control. The current literature reviews and uses eight different evaluation metrics in two categories; (1) the predictions are fine-tuned to increase precision with the merit that the margin of error does not fall within 3 s (i.e., high precision predictions at the expense of being less useful for proactive decision making) as discussed in the following section, and (2) the predictions are fine-tuned to minimize the early warnings such that there are no FPs predicted in (<3 s) duration which can result in giving alerts to late to a decision assist system and real-time control frameworks. [20].

Driver modeling in mixed traffic scenarios (MTS) is essential for the safety, accuracy, and efficiency of autonomous vehicles. Researchers in vehicle perception & behavior study the interactions of autonomous vehicles and human-driving vehicles (HDVs) on numerous aspects. Nevertheless, no integration of the developed machine learning models, obtained for supervised learning of the features of driver behavior, have been attempted for MTS scenarios. In this paper, a unified formalized mathematical flow has been developed to obtain the proposed predictors [2].

5.1. Accuracy

We identify a set of diversified and yet hopefully relevant driving models and evaluate their prediction performance inside real complex traffic in urban environments. By evaluating high-fidelity spatiotemporal rules, trajectory prediction, and various learned approaches, we show how the human decisions could be modeled and compared. Finally, the impact of using real instead of random numbers or learned actions for the future of today's driver behavior prediction is discussed. We demonstrate how various interesting statistics we provide can be included in the evaluation process.

We hope to gain the attention of researchers from different disciplines aiming to study driver behavior in mixed traffic. To facilitate statistical comparison of different models, we hope to provide a valuable tool for benchmarking purposes to the scientific community. Indeed, driver behaviors are complex spatiotemporal phenomena inside the traffic and there are many variables that are not easily predicted [2].

Predicting future driver behavior is an important prerequisite for enhancing the functional safety of autonomous vehicles. The development of models for predicting future driver behavior and their evaluation require a potentially enormous quantity of realistic driving scenarios to be handled in mixed traffic. This paper presents a comprehensive benchmark of state-of-the-art driver behavior models in a gap acceptance scenario involving a leading car and oncoming traffic in mixed traffic situations. The models are tested in two freely available, large-scale datasets for mixed urban traffic, enabling researchers to foresee their driver model's capability of predicting realistic human driver behavior within urban mixed traffic [21].

5.2. Precision and Recall

The recall metric indicates the security of the system, with recall=1 being a result where all the drivers are correctly classified. This balanced measure, the F1-score, is used to balance non-inconvenience and model security. When recall and precision conflict, e.g., reducing the false positive reduces the true positive, the F1-score rates the increase of high precision through robbing relatively less true positive. Furthermore, as a robust statistic to measure agreement between evaluators, e.g. two persons, Cohen Kappa can be used in machine learning to assess how well each instance is labeled. Thus, we also include a comparison between the Cohen Kappa characterizing our machine learning classification model and a revealed label by two persons on the same dataset. In this experiment, we have Big Data in labelling comparison between 2 evaluators, with more than 8 million labels pairs.

[22] Assessing the performance of a machine learning model is essential to determining the optimality of the model. Models may perform well on only one measure at the expense of another, e.g., maximum precision in classification at the expense of low recall. [23]In other words, operational measures optimize performance towards one task of the total problem, avoiding in most cases the task of the other end of the total problem. Precision captures how well the model correctly classifies the positive cases, and recall captures how many of the actual positive cases the model found. In this work, we apply precision and recall to evaluate our model performance in classifying passenger and driver on-board reading behavior, which is one kind of behaviors that human monitoring at work on machines labeled of interfering and not interfering.

5.3. F1 Score

It is with the use of such data driven approach from past, that it has been proposed to estimate driver's behavior and predict vehicle's collision risk on real highways by leveraging prior knowledge from future infrastructure terrain. As detailed, it was observed that the vehicle-to-infrastructure communication demonstrated superior prediction for driver's complete stop in near-crash scenarios in comparison to the driver's rapid deceleration. The experimental setup also assisted in learning this classification invariants from the observed near-crash events while they were unfolded one after the other utilizing the recorded multi-modal driving dataset using video, lidar, and RADAR and presented as driver's action probability and estimated associated risk probability template. It nodes it performance by evaluating the next driver's actions predicted through decelerating vehicle-to-infrastructure communication and infrastructure-to-driver action learning observed through the collected dataset of near-crash risk driving experiments [24].

We aim to better understand the driving behavior during noncritical events that could lead to an imminent traffic crash. To achieve this, we developed a label set for immediate driver actions (important for CR prediction) and distant driver actions (important for near-crash scenario prediction). As a part of this, artificial intelligence helped us identify discerning features that capture spatial-temporal dependencies. Our F1 score based result outline the stark difference in near-crash category classification which alternatively vanished in case of immediate driver actions. This implies that the latter were easier to predict as compared to the near-crash scenarios [25].

6. Experimental Setup

The following section describes the experimental setup in a Traffic Engineering and will show by using of a Simulation the development of an EKF for driver behavior prediction. Therefore, the traffic/classic EKF based on the driving scenarios "lane change", "overtaking" and "driving straight" will be determined. While structured on this basis, a forecast of the driving characteristics and also a prediction of the lane change behavior is made. Finally, the process develops mathematical connections between radar measurements and vehicle model speed and acceleration and verifies that the EKF can be parameterized on this basis to predict driver behavior.

A highway scenario was chosen for the experimental implementation of the generic Extended Kalman Filter (EKF) as the basis for the driving behavior prediction. While the prediction of the surrounding traffic participants would normally also require a localization system, this was not necessary due to the already existing sensor data in the simulation environment VTD and the simple adjustment of the position data from the ego vehicle. The approach was proven to be powerful as the prediction of traffic participants' future behavior was possible with relatively high accuracy, predicted up to approximately 7 seconds or within up to 70 vehicle lengths. The present work pointed out that this approach has the potential to be extended to any mixed traffic scenario.

6.1. Dataset Description

[14],[26].

The dataset contains five subsets, namely S1, S2, S3, S4 and S5. The (X, Y) values drawn in red and blue are the original data points. Curve fits corresponding to different hypotheses learnt by five different Climbing RBF 2 based on different training data are also shown. It is observed that for the Climbing RBF 2 learnt with S3 as training data, the curve fit closely follows the envelope curve consisting of all the training samples in S5 while at the same time maintaining juniors low values of the objective function for the training data. Hence S5 can be used as test data to evaluate the performance of the Climbing RBF 2 learnt with S3 as training data. Similar tackles be done for the other training data. The training data for the Climbing RBF 1 comprises some sample points from a sixth subset S6 that contains the samples from S1, S2, S3, S4 and S5. In Fig. 4 is shown the performance chart of the Climbing RBF 1, obtained with different subset training data, for predicting on subset S6. It was observed that the curve fit corresponding to hypothesis learnt with any training input is pretty close to the envelope curve of all the samples from S6. It was concluded from the experiment that the Climbing RBF learning on S6 as training data delivers an accurate juniors fit for predicting the quantity for an independent test set that could be S6 or any subset of the dataset. The inherent property of Climbing RBF 1 that also it like Climbing RBF 2 but embracing composite hidden feeble neurons of radius generated by a Gaussian evolution operator like RBF 2 showed performance that encapsulated climb ability was observed where many limitations which were present in the original RBF while learning the principles of generalization. This is due to the climbing attribute, observed in the Climbing RBF which was more functional while learning the lower cut-off radius of an excluded beadle. The aforementioned facts about Climbing RBF 2 were vindicated with the help of projections obtained via reduction of clutter by the SOM networks and information about regions of non-linear separability obtained by the fuzzy c-means clustering are part of our future research. Clustering subgeneric data to climb the incomplete portraits of the patterns requires to climb the performance of the classification model of graph based machine learning model necessitates to decrease the complexity of the net bass for storage capabilities, comic collecting noise and increasing the performance saliency. The complete body of the learning samples exist in the feature space of the training environment of the CLA of Climbing RBF 2 . This is beneficial in scaling learning machine learning algorithms during training forecast elimination and t draft. The experimentations were conducted utilizing the dataset of the drivers. The generation of the dataset for the research work was according to the concession taken from the shodhan trader

6.2. Feature Selection

We tested four state-of-the-art machine learning models for this purpose [27]. The results showed that XGBoost combined with cost sensitive learning for compensating for the class imbalance improves the performance of the model. As a consequence, the machine learning ensemble Gycler, which builds on frequentistic data augmentation techniques to balance the classes, performed less well in the present prediction task. For the intention recognition of pedestrians and cyclists, there is limited literature and in particular small traffic conflict datasets cannot be generated for the validation of the intention recognition models. Comparing mean absolute error for the output estimates G2F is learned with innovation potentials of several forgetting schemes. Different to interactive behaviour, individual processing of maps of intention clusters increases prediction accuracy for pedestrians. Random forgetting with exponential decay and a fixed learning rate work best. Several of the maps are of the size shorter than the actual trajectory of the road user and can give valuable insights in state of the art intention recognition [28].

The traffic accident data are imbalanced, with most of the traffic accidents being less severe overall. To address the imbalanced class issue, many researchers in the field of traffic safety have used a resampling technique to deal with the imbalanced class problem when predicting traffic accident severity using machine learning techniques. This approach to mitigate the shortcomings associated with imbalanced class in traffic safety data is not without criticism, as shortcomings of the resampling technique are also discussed. Moreover, this study showed that a resampling technique does not improve the predictive performance of our gradient boosting machine-learning model. For researchers and professionals involved in traffic safety, the present study suggests that models predicting the severity of traffic accidents need not to be developed as oversampling may not be an effective remedy of the issue at hand. Lastly, it is also beneficial to apply and compare various state-of-the-art machine learning techniques rather than to simply apply the techniques underlying ensemble learning methods as gradient boosting.

Traffic safety researchers and professionals are often interested in assessing how predictor variables may be related to the severity of traffic accidents. Many researchers and professionals have over the last decade used various machine learning techniques to better understand how predictor variables in traffic safety data are related to traffic accident severity. In 2018, a study by Way [2] examined how predictor variables in traffic safety data might be related to traffic accident severity using the extreme gradient boosting machine learning algorithm. The results showed that the gradient boosting machine learning algorithm performed well, and that being a driver, a young driver, a female driver, being involved in a roll severity accident type, being involved in a wet road condition, having sustained black ice on the roadway, being involved in a chain reaction type of traffic accident, running off the road, falling asleep or nodding off, and entering a skid are predictive of sustaining fatal injuries in traffic accidents.

6.3. Model Training and Testing

In our work on proposing and validating a set of sensory features that are capable of representing driving interactions, we concentrate on estimating the responsiveness of surrounding vehicles under different situations. These features enable on-line recognition of criticality, which is considered as a measure of responsiveness in our study [29]. Having these inputs, a model will predict whether a surrounding vehicle will yield to the ego-vehicle or has imperturbable behavior. Such models provide more explicit understanding of how vehicles respond to one another and how to represent the interactions in mixed traffic scenarios. Therefore, these models provide valuable information to designing cooperative autonomous vehicles, especially if the vehicle is equipped with different prediction models, and it needs to estimate the likely behavior of different road users under different situations.

In our work on proposing and validating a set of sensory features that are capable of representing driving interactions, we concentrate on estimating the responsiveness of surrounding vehicles under different situations [12]. These features enable on-line recognition of criticality, which is considered as a measure of responsiveness in our study. In this section, we address the process of training and testing our predictive models.

7. Results and Discussion

In recent years, the topic of vehicle-vehicle interaction has attracted significant research effort, and while much remains unknown, studies on this topic have been fairly coherent. The relevance of the classic kinetic-based CCC was often the focus. At the same time, some of the emerging driver-centered behavioral-zone-based approaches have proven effective for predicting motion behaviors. At the same time, specifically in terms of driver-vehicle interaction, relational spatial models have proven valid for various task environments. The focus on the interoperation between vehicle models and drivers is promising but still lacks indepth study.[30] Whether and how car drivers comply with traffic rules at road intersections and possibly overlook motorcycles as vulnerable road users illustrates the importance of carefully considering driver behaviors in transportation science and engineering. The flow of vehicles actions has been shown to be determined by drivers' value over time functions and their value assessment processes. However, the contribution of drivers' value assessment functions to mixed traffic flow in urban roads, considering different vehicle types using acceleration/deceleration concerning drivers' characteristics as key factors has not been comprehensively investigated yet. According to the study all streams tend to spend more time obeying traffic rules with increase in deceleration. For the same acceleration with new drivers, MCs are decoded by positive rewards length and time faster than three-wheelers, but when acceleration increases, drivers drive longer with positive rewards for three-wheelers than for MCs. The VISSIM model was calibrated so that it was relative to the behavioural characteristics of different driver types. ViSSim determines the maximum deceleration and acceleration in the program model. Then, the analysis will benefit from various valuable information about drivers and be able to simulate different vehicles for different traffic. To investigate the influence of various driving dynamics on road behavior changes, the moment and length of MC traffic fragmentation were evaluated and compared with those of different driver models, considering stream interactions at intersection entrances and exits. The results show that better vehicles or better neutral items in queues significantly increase the chance of conflict related to concerns._conflict_weight value at the conflict point The three-wheelers and driver models with better driving dynamics are slightly less compliant with {yield}.

[10] Automated and autonomous vehicles and advanced driver assistance systems have the potential to deliver ground-breaking change to how traffic is managed in future Intelligent Transportation Systems (ITS). They also bring new challenges; those systems need state-of-the-art traffic prediction algorithms delivering good predictive performance because the behavior of susceptible vehicles has a major impact when making driving decisions. Critical incident prediction or traffic state scenario prediction have been two highly practical research areas in this regard. Researchers largely focus on the typical classification error metrics, false positive or negative percentages, true positive or negative rates, precision, recall, specificity, etc.. Predictor comparison can typically be made by receiver operating characteristic (ROC) curve performance comparisons, using area under the ROC curve (AUC) scores. Partial AUC score, that is, area under the ROC curve for the bottom part, avoids being affected by the

unbalance of the negative and positive instances and therefore can be used when focussing unnecessarily on the dominant negative instances creates ethically sensitive bias.[6] Machine-learning models were found to, in part, predict more accurately than a simplified rule-based model, using only parameters regarding one's distance to the pedestrian. Furthermore, it was found that considering additional contextual parameters relevant to the scene—such as velocity-like features and the presence of other pedestrians—mitigates much of the additional inaccuracy observed for the rule-based model. However, contrary to expectations, behavioral-type models did not generally perform worse than the machinelearners' predictions. Classifiers towards the combined target "large aversion" generally performed worse than the unrealistic rule, but those in the "small aversion" class often suffered similarly to or more than the human classifiers. Applying the most consistent scan-path and visual-feature-recognition models did not produce more accurate behavior predictions, which deviates from common assumptions that they provide more human-like results, but is in line with the decreased predictive power of those modalities.

7.1. Performance Comparison

The paper incorporates three different datasets from different countries. The first is JHopkins (represented by the Highway) dataset, the second dataset is the NGSIM dataset (represented by the US-101 dataset) [31]. From the transmitted results, models trained on the COMMA dataset provides the better results for predictive time stamps than the iLyft, that too when the trajectory data of the car is considered for prediction. The Highway could be utilized for real-scenario analysis, whereas US-101 has different lanes capturing different traffic conditions. CONCLUSION: The results show the comparative analysis and clearly states the efficiency of the models under different machine learning research. The comparison represents how well the model is accurate under multi lane and higher density traffic conditions.

Machine-learning-based vehicle trajectory prediction has become an essential model in mixed-traffic scenarios. However, the basic evaluations of these models usually focused on metrics of prediction precision [10]. Unfortunately, these evaluation criteria do not necessarily correspond to better performance from the perspective of real-life traffic flow. Therefore, a comparison of different machine learning models for vehicle trajectory prediction in real life mixed traffic scenarios is required. The study aims to compare vehicle trajectory prediction models based on different machine learning paradigms over two different traffic conditions representing various mixed traffic scenarios using two different datasets using two evaluation criteria over two scenarios, predicting trajectory data 0.5, 1, and 3 seconds in advance.

7.2. Interpretation of Results

After developing interpretation maps and explicating evaluation landscape models that respect the main tasks, several recommendations result, e.g. concerning the number of Public Road Safety Awareness Campaign Participants without a Driving License on the total number of all participants; concerning correct signaling and respect of right of way, and slowing down the vehicle enough to ensure it yields to other road users who have priority; the fines for motorcyclists who are caught riding without wearing helmets; the level of knowledge and respect of the road rules by the drivers; reckless driving; proceeding beyond speed limitations; avoiding driving under the impact of alcohol or drugs; wearing seat belts and using child seats; the improvement of highway infrastructure in the vicinity of urban areas; periodic road safety campaigns with special focus on drunk driving, enforcement including breath and saliva tests.

Despite the potential of machine learning (ML) and computational models in the analysis of road users' behavior in both traditional [32] and future [1] scenarios, recent research and investigations deal with predictive modeling and scenario analyses; known as what-if simulations, using models that can relate traffic and road users' behavior and attitudes attributes with corresponding performance measures and targets related to safety, environment, and traffic efficiency. Accordingly, the aim of this chapter is to (i) derive rules and patterns to predict the occurrence of behavioral changes and risky interactions in drivers that may subsequently lead to vehicle crashes [33], (ii) to quantify the characteristics of significant determinants, (iii) and to contribute to the development and evaluation of machine learning models that predict the likelihood of emerging faulty and risky maneuvers of passengers/private transport users in different real mixed traffic scenarios and coping with potential changes in the EU national traffic and urban-area characteristics and (iv) policy.

8. Practical Applications

With the increasing use of AI in automotive applications such as advanced driver assistance systems (ADAS) and automated vehicles, prediction of the motion of road users in mixed traffic is becoming increasingly important. Predictive algorithms are used to estimate future behaviors of vehicle parameters (speed, position, etc.), and are important for the forecasting of future positions of road users as well as for safe and robust motion planning in these scenarios. State-of-the-art work reduces the accuracy of these methods in unpredictable traffic scenarios, such as changes in acceleration patterns and speed profiles. Proposed FNN-based motion prediction model has the potential to accurately and safely predict the motion of both LTCs and Bicycles in mixed traffic driving scenarios. For all the presented scenario types, LSTM appears to outperform the model accuracy of the normal FNN, while the normal FNN outperforms LSTM in terms of the structural simplicity and the inference time of the model .

A machine learning (ML) model works by fitting a curve to a given dataset using a specific learning algorithm. Over time, the model learns this curve. However, if multiple, or even a large number of curves are obtained using the same model, the variance of such a model is considered to be large. Even though individual curves generate better local fits, their summation often makes the global fit poor and prevents the model from generalizing well to new, unseen data . Consequently, model uncertainty (or the variance of model functional covariance) can hamper the utility of trained ML model in real-world applications. Model uncertainty arises from various sources. Each component of an ML model is probabilistic in nature and contributes to the resulting prediction uncertainty. In some applications, such as safety-critical autonomous vehicles and robotics, it is crucial for an ML model to learn input-output relationship as well as to capture model uncertainty so that it can abstain from making critical decisions when the model is uncertain .

8.1. Autonomous Vehicles

Mixed traffic broadly means having vehicles in traffic system with a mix of conventional human-driven vehicles and modern AVs. The main attribute of considering mixed traffic as a system under investigation is the fluctuating change between human and autonomous driving behavior that demands continuous interaction between them to achieve smoother and efficient journey experiences. The paradigm can be stated as autonomous driving models learn human and vice versa to give rise to socially adaptable autonomous driving, in that case it remains important to reference driving behavior models related works. For reference, different approaches have been summarized into two main categories. Approaches that are according to developed for human drivers' description are discussed first. Afterwards AVs' models and their connection to mixed traffic are discussed [34]. On AVs' models, the section of mixed traffic human-AV junction also refers to some of the research that were intended to model human-drivers' interaction with AVs. A straight forward direction to utilize these models thereafter can be in relation to behavior change intentions that could arise due to the presence of different kind of junctions.

This chapter provides a concise review of different modeling methodologies that have been introduced to describe the driving behavior in now traditional traffic and in vision of upcoming AVs [35]. Safety is one of the most crucial topics in the field of traffic science. One main task for the safety of human beings in traffic is the prediction, detection and classification of dangerous and non-dangerous driving behavior. To use these paradigms while designing new real-time traffic systems as well as AVs, the use of emerging driving behavior models could shed light on the kinematic and dynamic decision-making strategies for interactively operating such entities within a mixed traffic environment.

8.2. Driver Assistance Systems

However, before implementation, the potential to improve the safety in realworld driving scenarios, as well as the potential to influence driver behavior in a desired way, needs to be well understood. The results presented in this paper contribute to a better understanding of these aspects. Also, research combining several sensors and establishing ground truth for different systems might further the development of more advanced models. Also, due to the limited range, resolution, and refresh rate of on-board sensor systems, hardly any data has been captured in any of the reviewed studies for the large margin of the time each vehicle drives on highways as oppose to inner city scenarios.

Driver Assistance Systems (DAS) are of growing importance in modern passenger cars. Today, most manufacturers offer at least some form of DAS as an option [12]. Based on a review of recent work and findings in this paper, it appears that DAS have strong potential to enhance the safety of driving maneuvers. For instance, eco-assistant systems may help drivers to adopt fuel-saving driving styles. However, intelligent driver assistance systems based on machine learning models with sensor-fusion, have demonstrated potential to significantly reduce the impact of traffic jams and improve overall driver satisfaction [10].

9. Challenges and Future Directions

In many robotics and autonomous vehicle applications, practitioners prefer to directly specify behavioral or task objectives while training learning agents. [35] Defining rewards for complex autonomous systems is hard, but a better approach is to have a user click on representative states in a state-action space to specify a desirable behavior. Depending on the strengths and limitations of the reward functions, these IRS methods mostly aim to use them in two ways: (1) as guidance to establish stable imitation codes for steering a model of a human driver or (2) as optimization cost functions to discover human-like behavior from data. Many of these approaches learn task objectives by observing human demonstrations instead of traditional task-based programming or reinforcement learning.

Graph-based deep learning techniques have been gaining much attention in the broader scientific community due to their effectiveness on graph data such as meshes, point clouds, text (graphs of words in sentences), and communications (graphs of connected devices). [1] Graph-based deep learning methods have been used to formulate complex and dynamic interactions between driving agents (e.g., vehicles, pedestrians, and cyclists) in mixed traffic scenarios, making them promising candidates for modeling heterogeneous human driver behaviors, especially at an intersection. On the other hand, joint training of downstream tasks--e.g., trajectory prediction in autonomous driving scenarios-- can be made more robust by oversampling these data points in the training set (to encourage heterogeneous representation in the feature space). Understanding and predicting drivers' behavior is crucial for the development and deployment of safe and efficient autonomous vehicle systems. [8] In most of the real engineering problems, a system that needs to be controlled is disturbed by stochastic, unknown, and unmeasurable parameters. Under these conditions, it is heavily challenging to develop classical control strategies. Model-free control strategies offer an alternative solution to this problem. Using the model-free strategies, a learning agent collects input-output data of the system and tries to minimize a loss function using this data.

9.1. Data Privacy and Ethics

Recent ethical discussions related to data, predictive analytics, and machine learning have generated considerable debate, although our approach has a general relevance too. In contrast to passive data sources, such as criminal records or credit files, the collection and generation of personal data by means of technologies that draw on enormous web forces lead to complex relationships. Large-scale analyses of behavioral data generate new administrative and technical challenges to data collection and methods of study. Far from being superficial, the collection of these data involves profound ethical issues, such as guarantees of good consent from relevant individuals, safeguarding the security of researchers and research assistants, and maintaining an eye toward bias towards societal and institutional institutional advantages.

Unbiased driving response prediction has both ethical and practical value. It can provide a better insight into what can be interpreted and highlighted as beneficial driving behavior. [36] From an ethical perspective, identifying risky driving behavior based on anomalies might discriminate against instances of safe driving. For this reason, it is crucial to balance both types of data. [5] In terms of research focus, an approach based on imbalanced learning is required to handle a massive influx of risky driving over safe driving. There is a strong imbalance among the distributions of different driving behaviors depending on the definitions or characteristics of the behaviors considered. The imbalance problem risks understating the results of our system, especially if only the features of risky driving are learned, disadvantaging instances of safe driving. Such a learning bias could have consequences which would not provide a balanced classification between risky and safe driving for the behaviors.

9.2. Real-time Implementation Challenges

The control process parameters change according to the real-time infrastructure and also there are lots of issues like unexpected behaviors from the diverse vehicles in the different scenarios. Cars, as entity of taxies, swill in and out the scenario, pedestrians or animals [37]. We have additional changes in the traffic polices that change the behaviors drivers in the different geographies. The forecasts of driver model control process should be able to distinguish different geo-scenarios, perform vary well independent on the infrastructures and be sustainable for traffic policies changes.

Real-time systems have been traditionally implemented in different smart vehicles and are not unique for autonomous vehicles. Most of the systems have been developed for driver-assistance systems. These real-time control processes are essential for demands of autonomous driving systems and the needed actuator´s implementation [38]. In real autonomous vehicles, these control processes also might be relevant for implementation of new actuator control or actuarial layers more adaptive to mixed-traffic scenarios [39].

10. Conclusion

[4] In this paper, an RNN-based traffic model is introduced capable of learning and following a car-like mobility pattern while obeying vehicle dynamics constraints. We observe that the results are as good as the previously presented solutions while having a much lower degrees of freedom. Furthermore, the mathematical essence of this model aligns very closely with the approaches that are responsible for the successful developments of other DNN structures.[1] As a result, RNNs deserve to be recognized as a very convenient DNN architecture for learning new traffic patterns not only for the sake of the internationally envisaged industrial developments but also for the sake of theoretical interpretation of human mobility. Future research can focus on traffic and motion prediction for mixed scenarios consisting of both humans and vehicles. Such studies are anticipated to reveal further behavioral attributes of the respective entities that is different from the traditional approaches. For instance, individual humans tend to exhibit more upper level learning while following vehicles mostly rely on predefined paths.еще a lot of artical → technologylock-com secnews lotartical.

10.1. Summary of Findings

In this chapter, the main findings of research project in the frame of Autostrade Tech S.p.A. (AT) and on the use of Machine Learning achieved using a Mobile App installed on a smartphone in order to collect data about driver behavior in terms of lane-changing maneuvers will be exposed. It has been used anywhere on the Italian motorway network. "Big Data" have been registered. Not only we have an acceleration index, we have also speed data, pitch and roll of vehicle at the moment of the lane changing maneuvers occurred. And finally, a learning approach has been tested through cross-validation to predict driver behavior in the case of missing event and in the case of different scenario with different phone from the one used for online prediction [33].

Automatic driver assistance systems have been significantly improved thanks to Deep Learning technologies in several categories such as traffic modeling and congestion prediction, collision avoidance, and driver behavior modeling [4]. These software have the potential to enhance the safety of car driving and to help drivers in critical road situations. However, to keep it useful, it is essential to consider the vast complexity of a common scenario, where often autonomous and human-driven vehicles coexist. Under some circumstances and in some areas, fully autonomous vehicles could encounter dangerous behaviour of other drivers. For this reason, understanding and simulating manually driven cars' (MDC) behavior will be a key factor for the success of these new technologies.

10.2. Contributions and Implications

To clarify the impact of the training data, the metrics for the test data in Case A are calculated as follows: a minimum value of 0.01 and an average value of 0.10 are % positive (0.8~1) driver behavior accuracy rate; 0.24 (minimum value) and % 0.39 (average value) negative (−1~0.2) driver behavior accuracy rate; 0.28 (minimum value) and % 0.34 (average value) driver misclassification accuracy rate. The smart risk manager, resulting from the decision support system is responsible for immediately decelerating when there is unsafe driving behind and changing the lane since the passengers are often aware of more risky driving than the autonomous vehicle during the driving process. Consequently, driver dangerous following behavior can also be interpreted as a trigger that triggers changing lanes in addition to decelerating. Note that, being located in the AV occasionally requires following an unsafe driver. If we consider the approach as the perception of humans driving the analyzed data in training data construction, we thus eliminate the possibility of making situational decisions in human characteristics. The number of time steps in the dataset obtained is 8000.

In the scope of this research, two different study cases were defined for the analysis and modeling of driver behavior: urban intersection and highway driving scenarios. Starting with highway scenarios (Case A), driver following behavior is dealt with here. The Harbor Freeway dataset was used for training and testing the Machine Learning-based models. The dataset represents 1.66GB, which covers 880.79km of travel by 1094 different drivers capturing subjective lane choice and driver behavior on the road. A time horizon of 10 seconds for 8000 driving scenarios was considered, resulting in 18,412,681 data points collected from the dataset. The dataset is taken completely from the General Supervised Autonomous Driving scenario, which means the dynamics of other nearby vehicles are fully observed. Additionally, we did not consider the dynamics of 4-wheel vehicles while building the data according. This is due to the fact that the vast majority of commercially available autonomous vehicles and softwares are being optimized and trained for general traffic rather than just 4-wheel vehicles. The traffic density and road users' safety conditions in urban driving scenarios may be more uncertain—that is, even 4-wheel vehicles may behave non-safely. Therefore, if we train a machine learning model that partially includes 4-wheel vehicle trajectories, we may not get effective decision support.

Reference:

1. Tatineni, S., and A. Katari. "Advanced AI-Driven Techniques for Integrating DevOps and MLOps: Enhancing Continuous Integration, Deployment, and Monitoring in Machine Learning Projects". *Journal of Science & Technology*, vol. 2, no. 2, July 2021, pp. 68-98, https://thesciencebrigade.com/jst/article/view/243.
2. Prabhod, Kummaragunta Joel. "Advanced Techniques in Reinforcement Learning and Deep Learning for Autonomous Vehicle Navigation: Integrating Large Language Models for Real-Time Decision Making." *Journal of AI-Assisted Scientific Discovery* 3.1 (2023): 1-20.
3. Tatineni, Sumanth, and Sandeep Chinamanagonda. "Leveraging Artificial Intelligence for Predictive Analytics in DevOps: Enhancing Continuous Integration and Continuous Deployment Pipelines for Optimal Performance". Journal of Artificial Intelligence Research and Applications, vol. 1, no. 1, Feb. 2021, pp. 103-38, https://aimlstudies.co.uk/index.php/jaira/article/view/104.