Accident Analysis and Prevention 142 (2020) 105578

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Detecting lane change maneuvers using SHRP2 naturalistic driving data: A comparative study machine learning techniques

Anik Das*, Md Nasim Khan, Mohamed M. Ahmed

University of Wyoming, Department of Civil & Architectural Engineering, 1000 E University Ave, Dept. 3295, Laramie, WY, 82071, United States

ARTICLE INFO

Keywords:
Lane change detection
Naturalistic driving study
Random Forest
Support vector machine
Artificial neural network
eXtrem gradient boosting
Connected vehicle

ABSTRACT

Lane change has been recognized as a challenging driving maneuver and a significant component of traffic safety research. Developing a real-time continuous lane change detection system can assist drivers to perform and deal with complex driving tasks or provide assistance when it is needed the most. This study proposed trajectory-level lane change detection models based on features from vehicle kinematics, machine vision, roadway characteristics, and driver demographics under different weather conditions. To develop the models, the SHRP2 Naturalistic Driving Study (NDS) and Roadway Information Database (RID) datasets were utilized. Initially, descriptive statistics were utilized to investigate the lane change behavior, which revealed significant differences among different weather conditions for most of the parameters. Six data fusion categories were introduced for the first time, considering different data availability. In order to select relevant features in each category, Boruta, a wrapper-based algorithm was employed. The lane change detection models were trained, validated, and comparatively evaluated using four Machine Learning algorithms including Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and eXtrem Gradient Boosting (XGBoost). The results revealed that the highest overall detection accuracy was found to be 95.9 % using the XGBoost model when all the features were included in the model. Moreover, the highest overall detection accuracy of 81.9 % using the RF model was achieved considering only vehicle kinematics-based features, indicating that the proposed model could be utilized when other data are not available. Furthermore, the analysis of the impact of weather conditions on lane change detection suggested that incorporating weather could improve the accuracy of lane change detection. In addition, the analysis of early lane change detection indicated that the proposed algorithm could predict the lane changes within 5 s before the vehicles cross the lane line. The developed detection models could be used to monitor and control driver behavior in a Cooperative Automated Vehicle environment.

1. Introduction

Driver behavior significantly affects safety and mobility on transportation system. As a driver behavior, lane change has a substantial impact on roadway safety. A number of studies showed several effects of lane change maneuver on traffic safety. According to the USDOT motor vehicle crash report, 451,000 motor vehicle crashes occurred due to lane change maneuvers in 2015 (NHTSA, 2015). You et al. suggested that lane-changing maneuvers are responsible for around 5% of total crashes and 7% of total crash fatalities (You et al., 2015). Wang & Knipling showed that more than 60,000 people are injured each year because of lane change related crashes (Bakhit et al., 2017). Other statistics stated that 240,000–610,000 lane change crashes occur annually (Zhao et al., 2017). In addition, other studies also investigated the adverse impact of lane change maneuvers on safety issues (Mattes, 2003; Reimer et al., 2013; Wang et al., 2019; Zheng et al., 2010). In general, these studies exhibited that lane change is a crucial driving maneuver and an important element of traffic safety research.

A lane change maneuver is one of the most commonly performed maneuvers, which takes place when a driver is in the process of moving the vehicle laterally from one lane to another. Lane change has been defined in different studies based on three criteria including explicit initiation and completion point, utilized data source, and required parameters (Xi and Crisler, 2013). According to Toledo and Zohar, lane change event is defined as passing from one lane to immediate next lane. They defined the initiation point as a particular time instance when subject vehicle begins lateral movement, and the completion point is the time when the subject vehicle ends its lateral movement (Toledo and Zohar, 2007). The study of Tijerina et al. defined lane change maneuver as a separate decision and execution phase (Tijerina

* Corresponding author.
E-mail addresses: adas4@uwyo.edu (A. Das), mkhan7@uwyo.edu (M.N. Khan), mahmed@uwyo.edu (M.M. Ahmed).

https://doi.org/10.1016/j.aap.2020.105578
Received 30 October 2019; Received in revised form 17 April 2020; Accepted 28 April 2020
Available online 11 May 2020
0001-4575/

et al., 2005). In another study, Fitch et al. defined lane change as a driving maneuver that moves from one lane to another lane where both lanes have the same direction of travel (Fitch et al., 2009).

Different methods of lane change identification can be found in the literature. For instance, a study conducted by Bogard and Francher identified lane change using GPS data and used mainly path-curvature data to identify lane changes, where the process consisted of six steps (Bogard and Fancher, 1999). Miller and Srinivasan proposed a method to determine lane change maneuvers of heavy trucks using yaw rate. The study hypothesized that a lane change would produce a noisy-sine-wave-like yaw rate signal (Miller and Srinivasan, 2005). Using the Next Generation Simulation (NGSIM) trajectory data, Thiemann proposed a smoothing algorithm and studied lane change dynamics. The study identified lane change using lane index that the vehicle is currently occupying, vehicle dimension, and vehicle position (Thiemann et al., 2008). In another study, Knoop et al. examined the number of lane changes as a function of the operational characteristics of the origin and target lane. They identified lane change using vehicle passing time, lane index, vehicle speed, and length from loop detector data (Knoop et al., 2012). Koziol et al. suggested a lane change identification method using degree of curvature data. Several parameters were utilized to specify a lane change maneuver including the variation of the degree of curvature, the maximum and minimum values of the degree of curvature, the duration between the maximum and minimum degree of curvature, and the duration of the entire lane change (Koziol et al., 1999). Another study utilized yaw rate and velocity to detect lane changes, turns, and curves on different road types (Ayres et al., 2004). The lane change identification method proposed by Xuan and Coifman used vehicle lateral position acquired from the Differential Global Positioning System (DGPS) (Yiguang Xuan and Coifman, 2006). A study conducted by Papathanasopoulou and Antoniou proposed a methodology based on temporary virtual lines to identify lane change maneuvers on mixed traffic trajectory data (Papathanasopoulou and Antoniou, 2018). Moreover, several subjective methods have also been used to identify lane change maneuver. In a study, lane change maneuver was identified when the initiation and completion points of simulated multi-lane highway seemed apparent from experimenter's judgment (Salvucci and Liu, 2002). Another study conducted by Hanowski et al. used driver's pushbutton activation to identify lane changes while studying their fatigue (Hanowski et al., 1999). In general, it has been observed that most of the lane change identification methods are based on either data processing algorithms or subjective measures.

Time window selection is one of the crucial steps of lane change detection. The selection of an appropriate time window has been discussed in many literatures. The study of Hou et al. utilized 1 s time interval to identify lane change events, speed, and position of each vehicle (Hou et al., 2015). Mandalia and Salvucci analyzed various window sizes (i.e., time length) in detecting lane change and considered a moving window of 1 s to 5 s (Mandalia and Salvucci, 2005). Li et al. also adopted different length of time windows from 1 s to 5 s for extracting different features to recognize lane change maneuvers (Li et al., 2018). Similarly, the study of Bakhit et al. utilized different window sizes ranging from 1 s to 5 s. They selected a fixed moving window size with higher model accuracy for their final model (Bakhit et al., 2017). In another study, Salvucci selected a moving window of 2 s to aggregate driver data for detecting lane change maneuvers (Salvucci, 2004). However, utilizing fixed time window for segmenting time-series data to detect lane change might not capture the original lane change scenario. Since lane change maneuvers may be executed over varying lengths of time, having a fixed time window could either miss a part of the maneuver or combine successive maneuvers in the analysis. Due to the limitation of fixed time window method, this study introduced dynamic segmentation (i.e., non-uniform time window) approach for the analysis (which is further discussed in the Data acquisition and processing section).

In addition, several studies utilized Machine Learning approach to detect lane change maneuvers. Yang et al. adopted Random Forest (RF) model for detecting lane changing decision and found the prediction accuracy of 85 %, 91.3 %, and 88 % for no lane change, left lane change and right lane change, respectively, using the NGSIM data (Yang et al., 2017). A study focused on detecting imminent lane change maneuvers in connected vehicle environments and found 80 % lane change detection accuracy using Artificial Neural Network (ANN) (Bakhit et al., 2017). Another study used Support Vector Machine (SVM) for detecting lane change intentions using instrumented vehicle and achieved accuracy close to 98 % (Mandalia and Salvucci, 2005). A study conducted by Kumar et al. proposed SVM and Bayesian filtering algorithm for predicting lane change intention and concluded that the proposed algorithm was able to predict lane change on an average of 1.3 s before it occurs (Kumar et al., 2013). A study concentrated on predicting driver's lane change decisions using a neural network model and found the accuracy of 94.58 % and 73.33 % for left and right lane changes, respectively (Zheng et al., 2014).

Although many studies have investigated the detection of lane change maneuvers, there is a lack of studies that utilized a data fusion approach based on data availability to detect lane change maneuvers. In addition, very few studies examined the capability of Machine Learning for lane change detection. Considering the research gap, the main objective of this study was to develop an effective method to detect lane change maneuver considering different data availability. This study, for the first time, presented six logical data fusion categories depending on different data sources for real-life applications. The study adopted several Machine Learning approaches to detect driver lane change maneuvers from trajectory-level SHRP2 Naturalistic Driving Study (NDS) and Roadway Information Database (RID) datasets. More specifically, the study will compare the performance of different Machine Learning approaches in terms of detecting lane change maneuver and provide recommendations to researchers regarding the selection of optimal approach appropriate for available datasets when conducting lane change behavior research. The study tasks included; 1) Time-series data aggregation based on lane change duration from the acquired NDS trips in clear, snow, rain, and fog weather, and manual annotation of lane change using Wyoming NDS Visualization and Visibility Identification Tool and NDS video data, 2) Dynamic segmentation of the selected NDS trips for developing "no lane changes" segment database, 3) Selecting relevant features, and hyperparameter tuning, training, and validation of lane change detection models using the selected features in each data fusion category, 4) Developing several Machine Learning lane change detection models based on selected input features and evaluate the performance of the developed models, 5) Investigating the capability of the proposed algorithm to predict early lane change maneuvers.

2. Data acquisition and processing

The SHRP2 NDS study focuses on understanding of drivers' interaction with and adaptation to the vehicle, traffic, roadway characteristics, and other environmental features (Campbell, 2012). Previous studies have utilized these comprehensive datasets to investigate driver behavior and performance, develop Surrogate Measures of Safety (SMoS), and calibrate microsimulation models (Ali et al., 2019; Das et al., 2020; Ghasemzadeh and Ahmed, 2018a, 2019; Hammit et al., 2018, 2019). The NDS collected an unprecedented amount of data from more than 3400 drivers in six US states including Florida, Indiana, New York, North Carolina, Pennsylvania, and Washington between 2010 and 2013 (Hutton et al., 2014). These data include vehicle kinematics data (e.g., speed, acceleration, etc.); machine vision-based data (e.g., lane position offset); radar data (e.g., longitudinal and lateral position of the surrounding vehicles); and front and rear roadway views from four video cameras (Campbell, 2012). The RID dataset was developed by the Center for Transportation Research and Education (CTRE) of Iowa State University. The RID dataset include detailed roadway data, e.g., horizontal curvature, grade, cross-slope, shoulder type, lane information, and so on, of the six NDS states (Center for Transportation Research and Education (CTRE), 2015; Hutton et al., 2014). A subset of the large SHRP2 NDS data were acquired from Virginia Tech Transportation Institute (VTTI) and utilized in this study. The SHRP2 NDS and RID datasets were linked in this study to detect lane change maneuvers under different geographical and environmental conditions.

NDS trips that occurred in different weather conditions (i.e., clear, snow, rain, and fog) were collected using the two unique methodologies. The methodologies were developed by the research team and based on weather data from the National Climatic Data Center (NCDC) and weather-related crash locations. Both of the methods utilize a radius of five nautical miles to isolate all the extracted NDS trips that occurred in different weather conditions (Ahmed et al., 2017, 2018; Das et al., 2019a; Das and Ahmed, 2019; Ghasemzadeh et al., 2019; Khan et al., 2018; Khan and Ahmed, 2020, 2019). By using these processes, a large number of NDS trips in different weather conditions were received. From the received NDS trips, 400 trips (100 trips in each weather condition) were randomly selected and considered for this study.

The next step of data processing was to aggregate time-series data based on lane change duration. In order to capture all lane change events occurring in the randomly selected NDS trips, an algorithm was developed. A lane position offset variable was considered for identifying lane change events from the time-series data. The variable is estimated from the distance to the left or right of the center of the lane and center of the vehicle based on machine vision techniques (Insight Website., 2019). In order to develop the algorithm, the peak (P_x) of the lane position offset variable was determined. Once the peak was identified, absolute differences of lane position offset (d_1 and d_2) between the peak and the immediate after and before time stamp from the continuous time-series data were checked, as shown in Eqs. (1) and (2).

$$d_1 = |P_x - P_{x+1}| \tag{1}$$

$$d_2 = |P_x - P_{x-1}| \tag{2}$$

Where P_x represents the peak of the lane position offset variable, P_{x+1} symbolizes the peak of the immediate after time stamp, and P_{x-1} indicates the peak of the immediate before time stamp. As recommended by a previous study, a threshold of ± 100 cm lateral shift (i.e., left and right) in the position of a vehicle can be considered as a lane change maneuver (Ghasemzadeh and Ahmed, 2018b). As an example, Fig. 1 represents sample of two-lane change maneuvers with lane position offset values above and below 100 cm. As can be seen in Fig. 1(a), lane position offset value started to increase indicating that the vehicle started to move from left to right of the lane center. When the value reaches a maximum point (i.e., P_x), a jump occurred indicating that the vehicle reached the far right of the driver's adjacent lane. The gradual reductions from the P_x and P_{x+1} in the backward and forward direction, respectively, were observed in the time-series data to identify the duration of a lane change event. When the reductions were ended at point A and B, it was considered the completion of the lane change event. The time difference between A and B was defined as the lane change duration where A and B symbolizes the initiation and termination point of the lane change event, respectively. The same process occurred when driver changed lane from right to the adjacent left lane (Fig. 1(b)). More details of the algorithm can be found in (Das et al., 2019c).

The algorithm was repeated for all considered NDS trips in the MATLAB environment and relevant parameters, including mean, maximum, minimum, and standard deviation of speed, acceleration, yaw rate, and lane position offset were computed for each lane change event. It is worth mentioning that accurate lane change annotation is essential for appropriate training of Machine Learning models. To verify the accuracy of lane change annotation, all the lane change events in different weather conditions were manually confirmed utilizing the developed Wyoming NDS Visualization and Visibility Identification Tool. (Ahmed et al., 2018, 2015; Das et al., 2020, 2019b; Das and Ahmed, 2019; Ghasemzadeh et al., 2019). A sample of 1200 "lane changes" (i.e., 300 lane changes from each weather conditions) were randomly selected for this study. Afterward, all the selected trips were dynamically segmented based on the distribution of lane change durations ranging from 1 s to 16 s. Once the dynamic segmentation was completed, lane change events occurring in any segment were removed for all trips to create only dynamic "no lane changes" segment database in a MATLAB environment. The dynamic segmentation of "no lane changes" were created due to the fact that lane change maneuvers could be executed over different lengths of time. It is worth mentioning that previous studies utilized higher non-matched samples to some extent than the matched samples in developing Machine Learning models considering the fact that the occurrence of "lane changes" are typically lower than "no lane changes" (Bakhit et al., 2017; Hou et al., 2015). Therefore, this study also utilized higher "no lane changes" samples compared to "lane changes" samples. To be specific, a sample of 2400 "no lane changes" were selected in such a way that the distribution of the duration of no lane change segments matches the distribution of lane change duration. The reason behind this matching/similar distribution is to match upon a potential confounding variable, which is lane change duration, in order to remove the confounding effect (i.e., to reduce bias) on lane change detection (Collett, 2003; Vandenbroucke et al., 2007). The "lane changes" and "no lane changes" datasets were

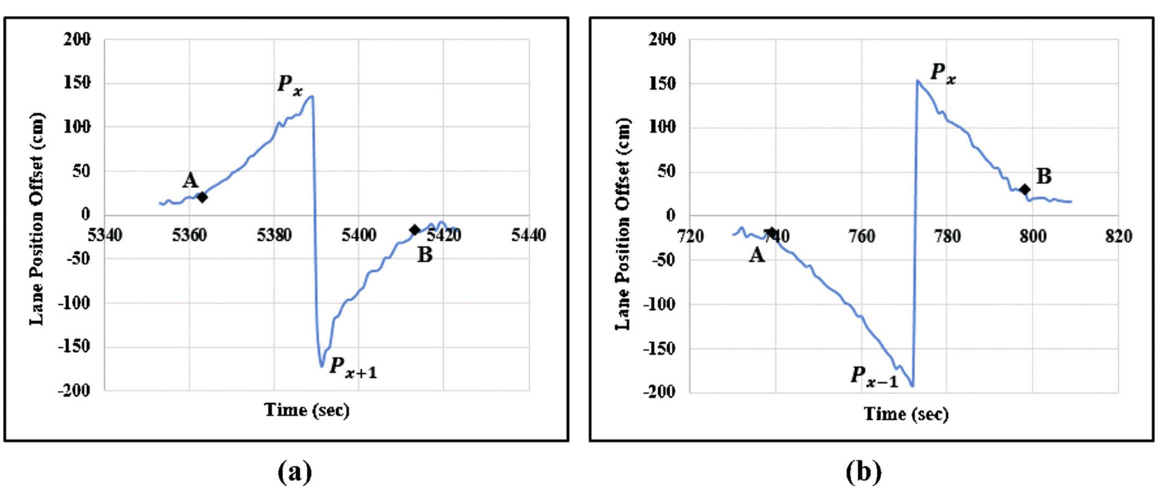


Fig. 1. Illustration of Lane Change Maneuver Using Lane Position Offset. (a) Lane Change to the Right, (b) Lane Change to the Left. A and B Represent the Start and End of Lane Change Maneuver.

then combined. Subsequently, roadway characteristics provided in the RID database and driver demographics provided in the SHRP2 administrated survey questionnaires were linked to create the final dataset. The final dataset corresponded to 145 drivers with ages ranged between 16–89 years with a significant number of drivers in the age group 20–24 years. It is worth mentioning that the sampled data resampled the original distribution of the original SHRP2 NDS data.

3. Methodology

As mentioned earlier, Machine Learning techniques were utilized to detect lane change maneuvers considering six different categories of features from vehicle kinematics, machine vision, roadway characteristics, and driver demographics. First, relevant features were selected and then trained and validated using different classic Machine Learning classifiers, which included Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN). However, eXtreme Gradient Boosting (XGBoost), a cutting-edge classifier, was also exploited in addition to the abovementioned classifiers. Then, the accuracy of the trained models was tested using a new dataset. It is worth mentioning that 80 % of data were utilized for training and validation, and the remaining 20 % of data were used for testing (Peng et al., 2004; Zhou, 2018).

3.1. Feature description

Four time-series parameters including vehicle speed, longitudinal acceleration, lateral acceleration, and yaw rate were extracted and utilized as measures of vehicle kinematics. The mean, maximum, minimum, and standard deviation of each vehicle kinematics parameter were considered as features. In addition to vehicle kinematics, lane position offset, which is based on machine vision, was also utilized. Moreover, additional parameters from roadway characteristics and driver demographics were considered, as these features can influence lane change maneuver. Table 1 shows a summary of the selected features used in different models.

It is worth mentioning that weather conditions could affect the quality of data collection and have a huge effect on lane change behavior. In order to investigate the lane change behavior in different weather conditions, descriptive statistics were conducted. It was observed that there were significant differences among different weather conditions for most of the parameters, such as speed, longitudinal acceleration and deceleration, lateral deceleration, and yaw rate during lane changes.

3.2. Classification algorithm

As stated earlier, the study utilized four widely used classification algorithms including Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and eXtreme Gradient Boosting (XGBoost) to detect the lane change maneuvers. The description of the four classification algorithms is provided in the following section.

3.2.1. Random Forest (RF)

Random Forest (RF) is a supervised classification algorithm that makes a set of Decision Trees from a randomly selected subset of training dataset and aggregates the prediction from each tree using voting (e.g., the prediction provided by each tree) to decide the final prediction. This technique usually builds forest with many classification trees. Typically, large number of trees in RF models provide more accurate predictions. The splitting decision of RF for classification purposes depends on Gini index (i.e., a measure of node purity) or entropy (i.e., a measure of node impurity). Some basic parameters of RF classifier include-the total number of trees to grow, number of randomly selected variables at a node split, and maximum tree depth. One of the advantages of RF is that it prevents overfitting problem by building trees on random subsets of the data. In addition, it can handle the missing values and provide relative feature importance that helps to select the most significant contributing features from the training dataset (Saraswat, 2019).

3.2.2. Support Vector Machine (SVM)

SVM is a discriminative classifier that is based on finding an optimum hyperplane in a high or infinite-dimensional space. The hyperplane is a line that separates and classifies data into two classes. SVM uses different kernel functions to transform data that can be used to form the hyperplane. The distance between the hyperplane and the nearest data point from either side of the hyperplane is known as the margin. The goal of a SVM classifier is to choose an optimum hyperplane with the maximum possible margin between two classes, which provides a greater chance to classify new data correctly. The points that lie closest to the hyperplane are the most difficult to classify. They are known as “Support Vector” (Bambrick, 2016; Hastie et al., 2001). In SVM, better classification results can be obtained by maximizing the margin between the hyperplane and the data points through a regularization function. The two main parameters of SVM-cost and gamma that can be tuned during the training. The cost parameter is related to the amount of allowable misclassification error in the training process during SVM optimization. If the cost is large, the optimization will select a narrow margin hyperplane that has the capability of classifying all the training points correctly. In contrast, for a smaller cost value, the optimizer will seek a large margin hyperplane; nevertheless, the hyperplane creates additional misclassification points. The gamma parameter represents the amount of influence area of a single training. The hyperplane is formed by only the closest points of the probable line (i.e., decision boundary) with a larger value of gamma, whereas the points further from the probable line are considered with a smaller gamma (Suthaharan, 2016). It is worth mentioning that different dimensionality reduction methods (e.g., fixed-slope regression, principal component analysis, etc.) can be used with different kernel functions in order to reduce the number of computations within the SVM kernel (Mazibuko and Mashao, 2006).

3.2.3. Artificial Neural Network (ANN)

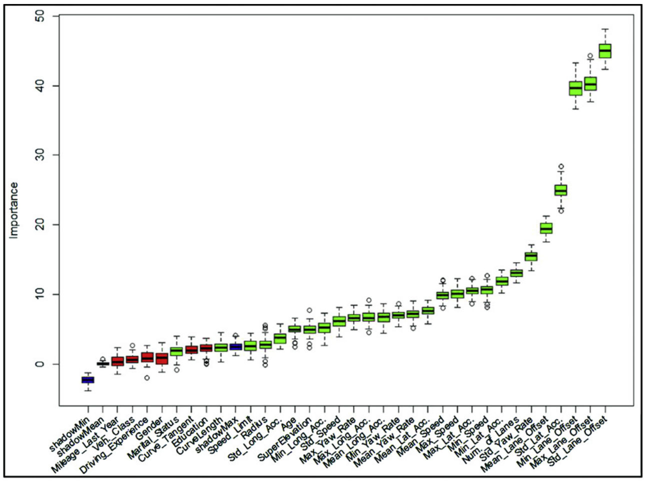
Artificial Neural Network (ANN) is information-processing systems based on the neural structure of the brain. An ANN system consists of a large number of interconnected neurons that are arranged in layers (i.e., input, hidden, and output layers) and works in a unit to solve a particular problem. The number of hidden units are important parameters of an ANN. In a neural network, nodes in one layer are interconnected to all nodes in the neighboring layers. Typically, the total number of weights in the network are controlled by the number of hidden nodes. In order to control the values of the weights among each neuron, forward and backward propagation are utilized during training. The complexity of a neural network during training process can be adjusted by limiting the number of hidden units and size of the weights, also known as weight decay. The weight decay method helps to prevent the network from using unnecessary weights (i.e., decaying the weights by some factors) and improve the model performance through backward propagation (Castrejón, 2016; Lek and Park, 2008; Valous and Sun, 2012).

3.2.4. eXtrem Gradient Boosting (XGBoost)

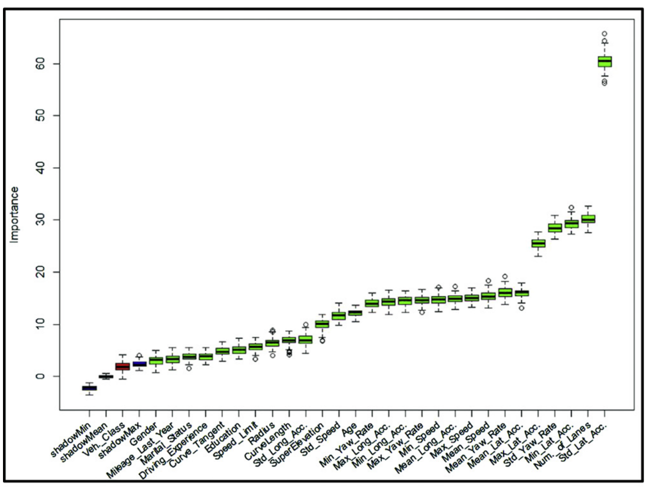
eXtrem Gradient Boosting (XGBoost) is an efficient implementation of Gradient Boosting method. Gradient Boosting is an ensemble learning method, which builds a prediction model based on a series of individual models (i.e., weak learners). In each individual model, the weights are adjusted in a sequential way based on information from the previous the model. The individual models in Gradient Boosting are not produced based on random selection of data, instead the models are formed sequentially. During each sequential iteration, the weights are

Table 1
Description of the Selected Features.

Features	Description	Type	Levels	Source
Vehicle Kinematics Features				
Speed (Mean, Maximum, Minimum, Standard Deviation) (mph)	Speed during lane change	Continuous	–	Vehicle Kinematics of NDS Time-Series Data
Longitudinal Acceleration/Deceleration (Mean, Maximum, Minimum, Standard Deviation) (g)	Acceleration/Deceleration in the longitudinal direction versus time during lane change	Continuous	–	Vehicle Kinematics of NDS Time-Series Data
Lateral Acceleration/Deceleration (Mean, Maximum, Minimum, Standard Deviation) (g)	Acceleration/Deceleration in the lateral direction versus time during lane change	Continuous	–	Vehicle Kinematics of NDS Time-Series Data
Yaw Rate (Mean, Maximum, Minimum, Standard Deviation) (deg/sec)	Angular velocity of vehicle around the vertical axis during lane change	Continuous	–	Vehicle Kinematics of NDS Time-Series Data
Machine Vision-based Feature				
Lane Position Offset (Mean, Maximum, Minimum, Standard Deviation) (cm)	Distance to the left or right of the center of the lane and center of the vehicle during lane change	Continuous	–	Machine Vision-based Data of NDS
Roadway Characteristics Features				
Presence of Curve	Whether the participants drove curve or tangent during lane change	Categorical	Curve Tangent	Roadway Information Database (RID)
Radius (Feet)	Curve radius in which participants drove during lane change	Continuous	–	Roadway Information Database (RID)
Superelvation (Percent)	Average Cross Slope of the segment during lane change	Continuous	–	Roadway Information Database (RID)
Curve Length (Feet)	Curve length in which participants drove during lane change	Continuous	–	Roadway Information Database (RID)
Number of Lanes	Number of lanes during lane change	Continuous	–	Roadway Information Database (RID)
Speed Limit (mph)	Speed limit during lane change	Categorical	≤ 60 mph > 60 mph	Roadway Information Database (RID)
Driver Demographics Features				
Gender	The participant's gender	Categorical	Male Female	SHRP2 Administrated Survey Questionnaires
Age	The participant's age	Categorical	Young: < 25 years Middle: 25−55 years Old: > 55 years	SHRP2 Administrated Survey Questionnaires
Education	The participant's highest completed level of education	Categorical	Low: High school diploma or G.E.D. Medium: Some education beyond high school but no degree and College degree High: Some graduate or professional school, but no advanced degree (e.g., J.D.S., M.S. or Ph.D.) and Advanced degree (e.g., J.D.S.,M.S. or Ph.D.)	SHRP2 Administrated Survey Questionnaires
Marital Status	The participant's marital status	Categorical	Single Married Other (Divorced, Widow(er), Unmarried Partners)	SHRP2 Administrated Survey Questionnaires
Vehicle Type	The participant's vehicle type	Categorical	Passenger Car/SUV Minivan/Pick-up	SHRP2 Administrated Survey Questionnaires
Driver Mileage Last year Details	The approximate number of miles the participant drove last year	Categorical	< 10,000 miles 10,000−20,000 miles > 20,000 miles	SHRP2 Administrated Survey Questionnaires
Driving Experience	The participant's number of driving years	Categorical	≤ 10 years > 10 years	SHRP2 Administrated Survey Questionnaires



Category 1



Category 2

Fig. 2. Relevant Feature Selection of Different Data Fusion Categories using Wrapper (Category 1 and 2).

adjusted by providing more weights on observation that are difficult to predict correctly. To minimize the loss function (a measure of prediction capability), gradient descent is used at each iteration in Gradient Boosting method. XGBoost utilizes more accurate estimation than the Gradient Boosting to find the best prediction model. XGBoost calculates second-order gradient of the loss function that provides additional information regarding gradient direction and getting minimum loss function. In addition, XGBoost utilizes regularized boosting technique that can reduce overfitting; therefore, better model performance is attained. The parameters of the XGBoost are the shrinkage (i.e., reduction of the learning rate), boosting iterations, minimum loss reduction, and other parameters related to decision tree (James et al., 2013; Nishida, 2017).

3.3. Data fusion categories

Once the features and classification algorithms were selected, the analysis was performed in six steps. The first category included combined features based on vehicle kinematics, machine vision (i.e., lane position offset), roadway characteristics, and driver demographics (Category 1). This category investigated the performance of lane change detection models when all the data are available. Although lane position offset illustrates a significant pattern during the lane change process, it might not be available or might provide erroneous value during harsh weather conditions where lane markings are not visible. It is worth noting that during the data reduction process, extreme harsh weather was excluded from the Category 1 dataset, which contained machine vision-based data.

The subsequent category considered all the features except lane position offset, which was collected using machine vision algorithms (Category 2). Note that this step considered all the weather conditions, including extreme harsh weather. One of the major limitations of machine vision algorithm is that it might not work properly during extreme adverse weather including heavy snow. In addition, lane position offset parameter might not be readily available. Therefore, the second step investigated the performance of the models in the absence of machine vision-based feature.

Category 3 utilized features from vehicle kinematics and roadway characteristics, and Category 4 considered features from vehicle kinematics and driver demographics. Roadway characteristics data are currently provided by many DOTs, therefore, these data could be available. Although data from driver demographics could be an issue for potential Personally Identifiable Information (PII), Category 4 might be possible if the PII issue could be compromised. The next category was based on the features of machine vision and roadway characteristics (Category 5), which could be available in newly emerging Autonomous Vehicles (AVs) technology.

The final category considered only vehicle kinematics-based features, which are readily available (Category 6). It is worth mentioning that myriad of similar data to NDS will become available with the advent of Connected and Automated vehicle deployment. Hence, the main purpose of using only vehicle kinematics-based features was to check whether the trained models could detect the lane change maneuver with acceptable accuracy when other data are not available including roadway characteristics, driver demographics, and machine vision. This splitting of the data in six different steps will provide guidance to transportation researchers on what data should be collected for lane change detection models.

3.4. Relevant feature selections

In order to discard the effect from noisy features and select relevant features in each step, Boruta, an all relevant feature selection wrapper algorithm, was employed using the "*Boruta*" package in R® (Kumar and Shaikh, 2017; Shaheen and Iqbal, 2018). The reason for choosing this technique is that it utilizes "all-relevant" feature selection method that allows to select all features (i.e., both strongly and weakly relevant) related to the outcome feature. On the contrary, most of the traditional feature selection methods aim to find a minimum possible subset of features for a chosen classifier to provide the best classification accuracy. Consequently, the algorithm reduces the random selection of the subsets of the features to find their correlation with the outcome feature. Moreover, the method can be dispersed over hundreds of cores, given that RF algorithm is utilized parallelly (Kursa and Rudnicki, 2010; Poona and Ismail, 2014; Sarkar et al., 2020). The method utilizes RF to compare the importance of features with the importance of shadow features iteratively. At every iteration, the algorithm examines whether a specific feature has significantly higher importance than the shadow features and removes the features that have highly less importance compared to shadow features. The algorithm ends when all the features are confirmed, dropped, or when the algorithm reaches a particular limit of RF runs (Kursa, 2018; Rudnicki et al., 2015). Figs. 2–4 display the final confirmed and rejected features based on Boruta. The blue boxplots indicate minimum, mean, and maximum Z score of a shadow feature. The green and red boxplots represent Z scores of confirmed and rejected features, respectively. The selected noise free features based on wrapper algorithm were utilized as the input for training and validating Machine Learning models.

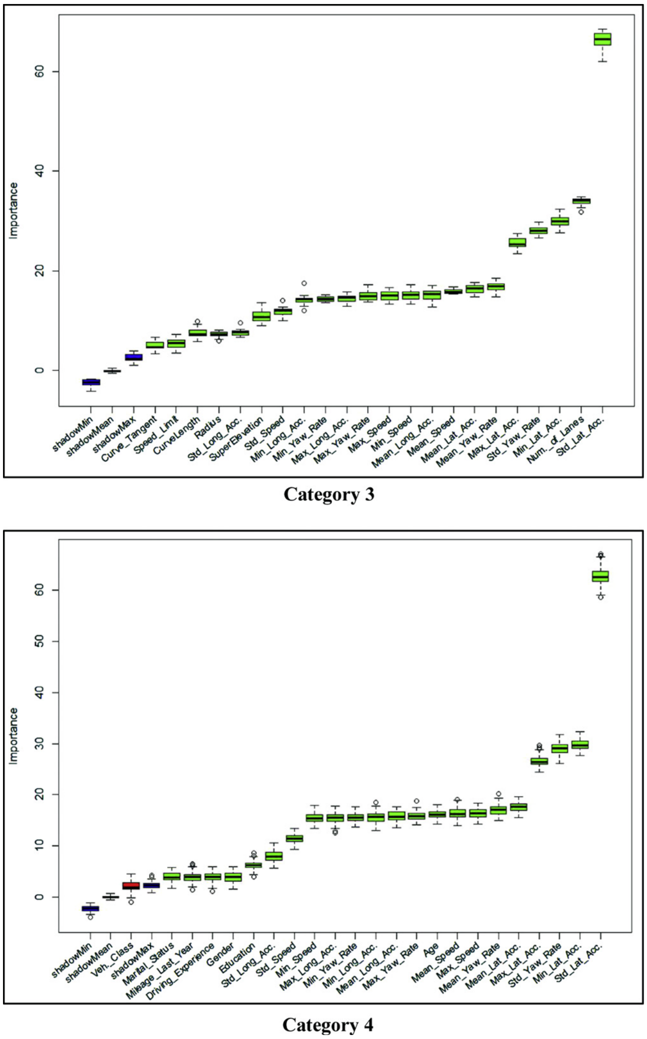


Fig. 3. Relevant Feature Selection of Different Data Fusion Categories using Wrapper (Category 3 and 4).

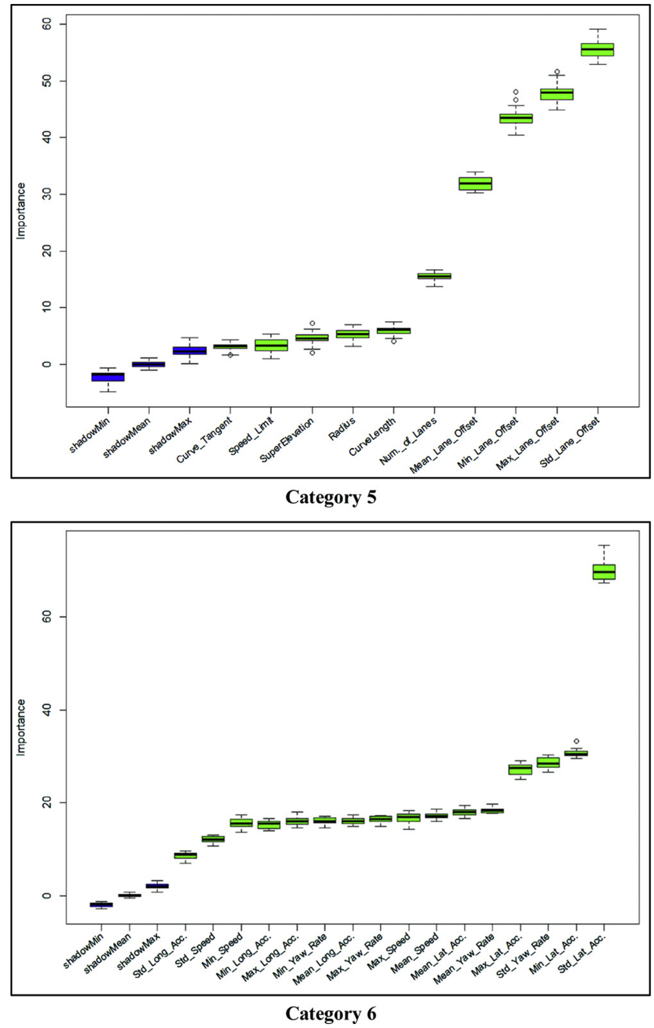


Fig. 4. Relevant Feature Selection of Different Data Fusion Categories using Wrapper (Category 5 and 6).

3.5. Hyperparameter tuning of classification algorithms

The study developed the detection models using the "*caret*" and "*h2o*" packages in R®. After splitting of data into six categories, Machine Learning algorithms were applied to train and validate the lane change detection models at every step. A 5-fold cross-validation technique was utilized to train and validate RF, SVM, ANN, and XGBoost models. Subsequently, a new test dataset was also utilized to test and evaluate the performance of the lane change detection models. It is worth noting that the testing dataset had never been used during training and validation of the Machine Learning models. The parameters of all the models were tuned in order to achieve the best performance. It is worth mentioning that this study utilized one of the most commonly used method named grid-search for tuning hyperparameters of the lane change detection models (Chicco, 2017; Goh and Ubeynarayana, 2017). For the RF model, three parameters, i.e., the number of trees to grow (ntrees), the number of variables randomly sampled at each tree node (mtries), and maximum depth of each tree in the forest (max_depth) were tuned. For instance, it was found that ntrees equals to 150, mtries equals to 5, and max_depth equals to 40 provided the best performance for the RF model used in Category 1. Similarly, the best SVM model after parameter tuning in Category 1 was based on radial kernel function, a cost and gamma value of 1 and 0.03, respectively. Different kernel functions including linear, polynomial, radial, and sigmoidal were utilized; however, radial kernel function provided the best performance results. In addition, the feed-forward neural network was utilized to train and validate the ANN model. Based on parameter tuning, a single hidden unit and a weight decay value of 0.1 were used for the ANN model in Category 1. Furthermore, several parameters including number of boosting iterations (nrounds), maximum tree depth (max_depth), shrinkage (eta), minimum loss reduction (gamma), subsample ratio of columns (colsample_bytree), minimum sum of instance weight (min_child_weight), and subsample percentage (subsample percentage) were tuned during the training of XGBoost model. The best performance was found for an nrounds of 50, a max_depth of 3, an eta of 0.4, a gamma of 0, a colsample_bytree of 0.6, a min_child_weight of 1, and a subsample percentage of 1 in Category 1. The hyperparameters of all the lane change detection models were tuned in a similar manner for other data fusion categories. Table 2 shows the parameter tuning results of the lane change detection models. Fig. 5 demonstrates the overall methodology of this study.

4. Results and discussions

4.1. Classification results of data fusion categories

The following sub-sections present a discussion of the results obtained for each of the data fusion categories.

Table 2
Parameter Tuning of the Lane Change Detection Models for Six Data Fusion Categories.

Models	Parameters	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
RF	Number of Randomly Selected Variables	5	15	6	10	4	5
	Number of Trees	150	150	200	200	200	150
	Maximum Tree Depth	40	20	30	50	40	10
SVM	Cost	1	1	1	1	1	1
	Gamma	0.03	0.02	0.05	0.03	0.13	0.11
ANN	Number of Hidden Units	1	5	5	5	5	3
	Weight Decay	0.1	0	0	0	0	0
XGBoost	Number of Boosting Iterations	50	150	150	150	50	100
	Maximum Tree Depth	3	2	2	3	1	2
	Shrinkage	0.4	0.3	0.3	0.4	0.3	0.4
	Minimum Loss Reduction	0	0	0	0	0	0
	Subsample Ratio of Columns	0.6	0.6	0.8	0.6	0.8	0.6
	Minimum Sum of Instance Weight	1	1	1	1	1	1
	Subsample Percentage	1	1	0.75	1	1	0.75

4.1.1. Category 1 – fusing vehicle kinematics, machine vision, roadway characteristics, and driver demographics

The detection summary of the RF, SVM, ANN, and XGBoost models using features based on vehicle kinematics, machine vision, roadway characteristics, and driver demographics is shown in Fig. 6 in the form of a confusion matrix. The confusion matrix represents the percentages of the correctly and incorrectly classified lane change events. The highest overall detection accuracy of 95.9 % was found for the XGBoost model during validation and 95 % for the XGBoost and RF models during testing. As can be seen in Fig. 6, the true positive rate of the XGBoost model was about 96 % during validation meaning that 96 % of lane changes have been detected correctly. However, the ANN model provided the highest true positive rate during validation (98.2 %). On the contrary, SVM model had the lowest true positive rate (95.1 %) during validation. Moreover, it was found that ANN model had the lowest true negative rate (89.4 %) and the highest false positive rate (10.6 %) during testing, which indicated that about 11 % of no lane changes have been detected as lane changes. Similar to the validation, the highest true positive rate and the lowest false negative rate were found for the ANN model for test dataset.

To measure the classification performance of the Machine Learning models, the Receiver Operating Characteristic (ROC) curve was utilized. More specifically, Area Under the Curve (AUC) of a ROC plot can be used to evaluate the performance. An AUC value over 0.9 indicates high accuracy, in between 0.7 to 0.9 represents moderate accuracy, and values between 0.5 to 0.7 denote poor accuracy (McDowell, 2006). Fig. 6 shows the ROC curves of the RF, SVM, ANN, and XGBoost models for Category 1 features during validation. All the models had AUC

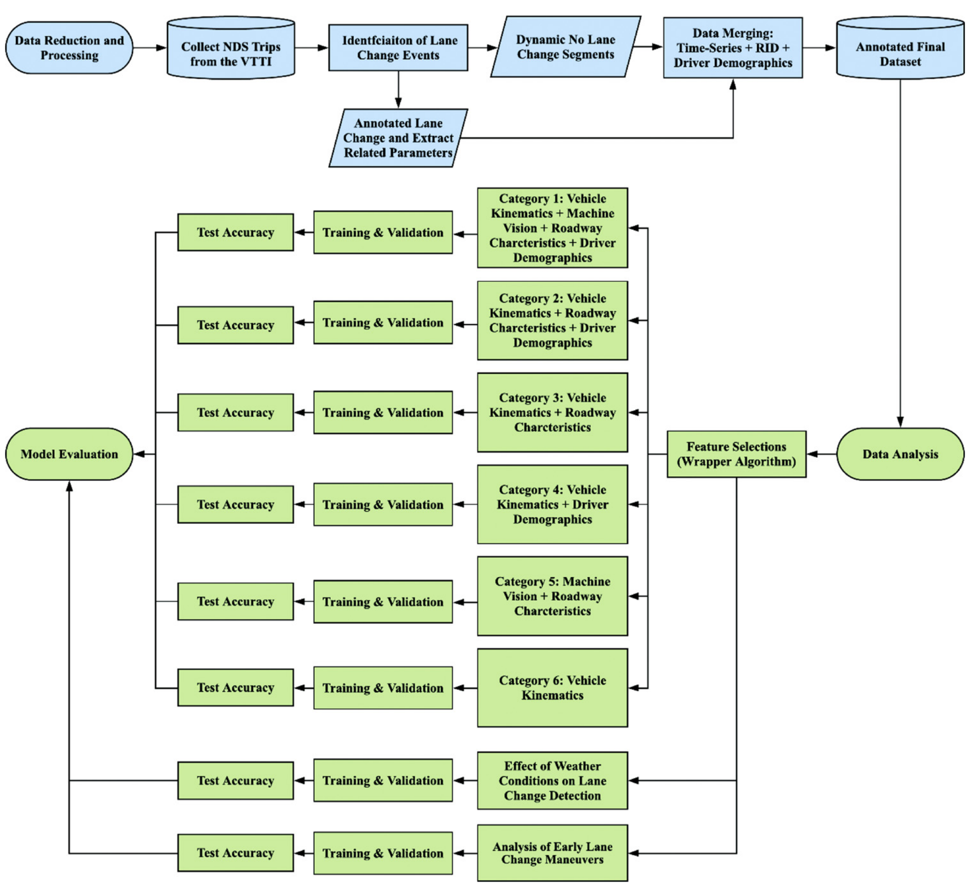


Fig. 5. Overall Methodology of Lane Change Detection Considered in this Study.

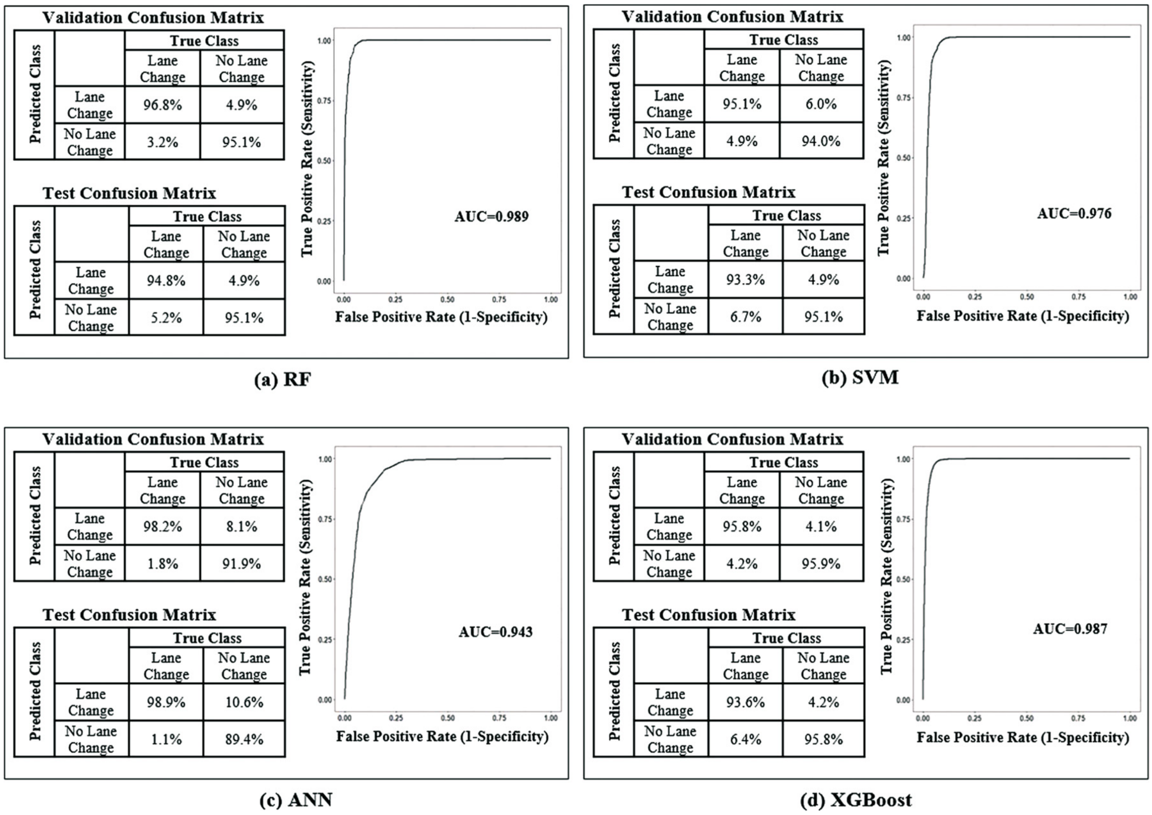


Fig. 6. Detection Summary of the Four Machine Learning Models Using Features Based on Category 1.

values greater than 0.9, which indicates all the models detected lane change maneuver with a pretty high accuracy. Considering the model evaluations, XGBoost model is the recommended model to be used for the Category 1 features.

4.1.2. Category 2 – fusing vehicle kinematics, roadway characteristics, and driver demographics

The overall detection accuracy of the RF, SVM, ANN, and XGBoost models was found to be comparatively lower but still impressive using features based on vehicle kinematics from CAN-bus and external sensors, roadway characteristics, and driver demographics. Fig. 7 shows the detection summary of the four models using confusion matrix. It was found that the RF classifier outperformed all other classifiers with a detection accuracy of 83 % and 81.3 % during validation and testing, respectively. In contrast, the lowest detection accuracy was found for the ANN model (75.9 % during validation and 74 % during testing). The highest true positive rate (68.7 %) and the lowest false negative rate (31.3 %) were found in the XGBoost model during validation, which indicates that the model has accurately detected around 69 % of lane changes. It is worth noting that the highest true positive rate (66.7 %) and the lowest false negative rate (33.3 %) were also found in the XGBoost model during testing suggesting that around 33 % of lane changes have been misclassified by the model. According to Fig. 7, the values of AUC of the trained models were found to be greater than 0.7, which indicates moderate accuracy. The finding pointed out that excluding the machine vision-based feature (i.e., lane position offset) did not significantly reduce the detection accuracy. Based on the evaluation results, the study suggested that the RF model would provide better lane change detection in presence of vehicle kinematics, roadway characteristics, and driver demographics.

4.1.3. Category 3 – fusing vehicle kinematics and roadway characteristics

The overall detection accuracy of the Machine Learning models was found to be similar to Category 2 considering the features from vehicle kinematics and roadway characteristics. The highest overall detection accuracy of 83 % and 81.7 % were found for the XGBoost classifier during validation and testing, respectively. Similar to Category 2, the lowest detection accuracy of 76.9 % and 73.8 % were found for the ANN model during validation and testing, respectively. The highest true negative rate (92.4 %) and the lowest false positive rate (7.6 %) during validation were found for the SVM model indicating that around 92 % of no lane changes have been correctly classified by the model (Fig. 8). Note that, SVM had similar results during testing. In contrast, the lowest true negative rate (88.7 %) and highest false positive rate (11.3 %) were found in ANN model during validation. The AUC of all the models were found to be greater than 0.7, as can be seen in Fig. 8. The findings indicated that the detection accuracy did not change after excluding driver demographics in addition to the machine vision-based feature. The model evaluations indicated that the XGBoost model should be used when considering features from vehicle kinematics and roadway characteristics.

4.1.4. Category 4 – fusing vehicle kinematics and driver demographics

The detection summary of the Machine Learning models using the features from vehicle kinematics and driver demographics is provided in Fig. 9. The highest overall detection accuracy was found for the RF and XGBoost models (82 % during validation), and XGBoost model (80.3 % during testing). It is worth noting that the highest overall detection accuracy during validation was found to be marginally lower compared to Category 2 and 3. The true positive rate of the XGBoost model was 66.2 % during validation, meaning that 66 % of lane changes have been detected correctly by the model. In fact, the XGBoost model provided the highest true positive rate during validation. However, the highest true negative rate (95.2 %) was found in RF model during validation, where 95 % of no lane changes were classified correctly. According to the AUC, similar moderate accuracy was found compared to Category 2 and Category 3 for all the trained Machine Learning models. RF model is the suggested model to be used for the

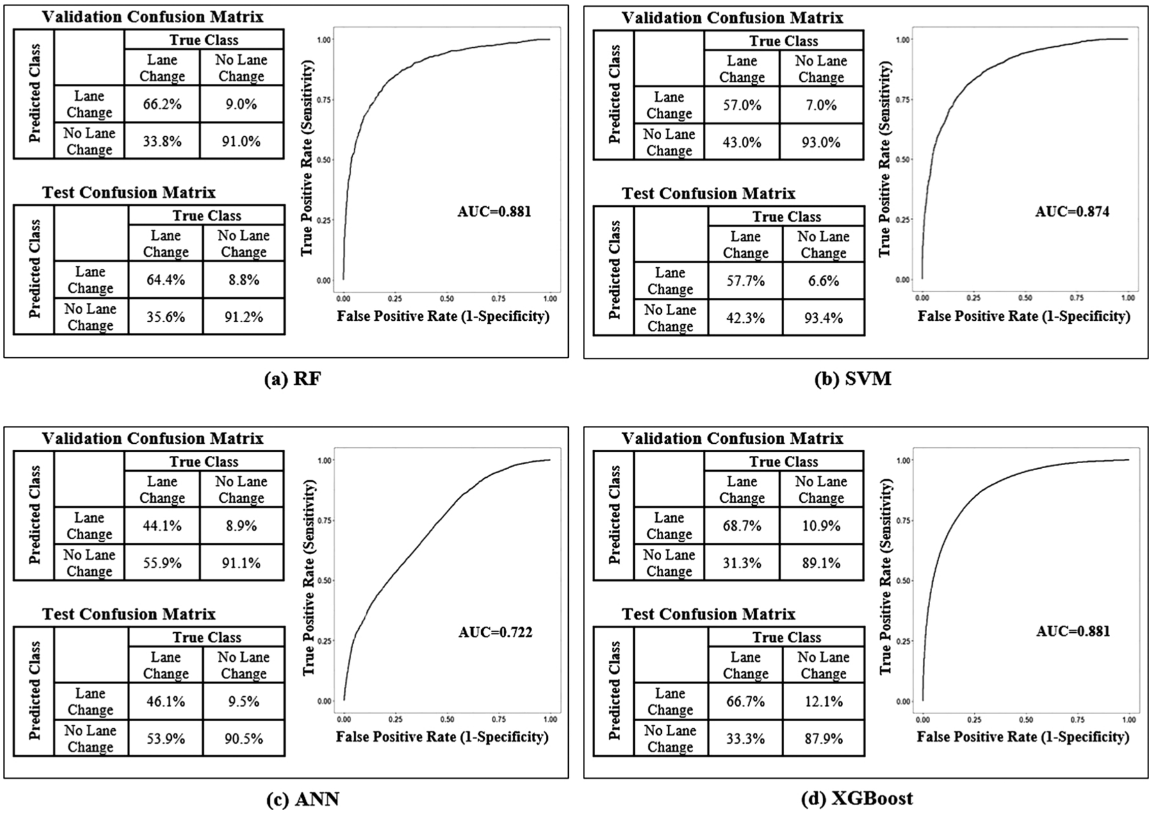


Fig. 7. Detection Summary of the Four Machine Learning Models Using Features Based on Category 2.

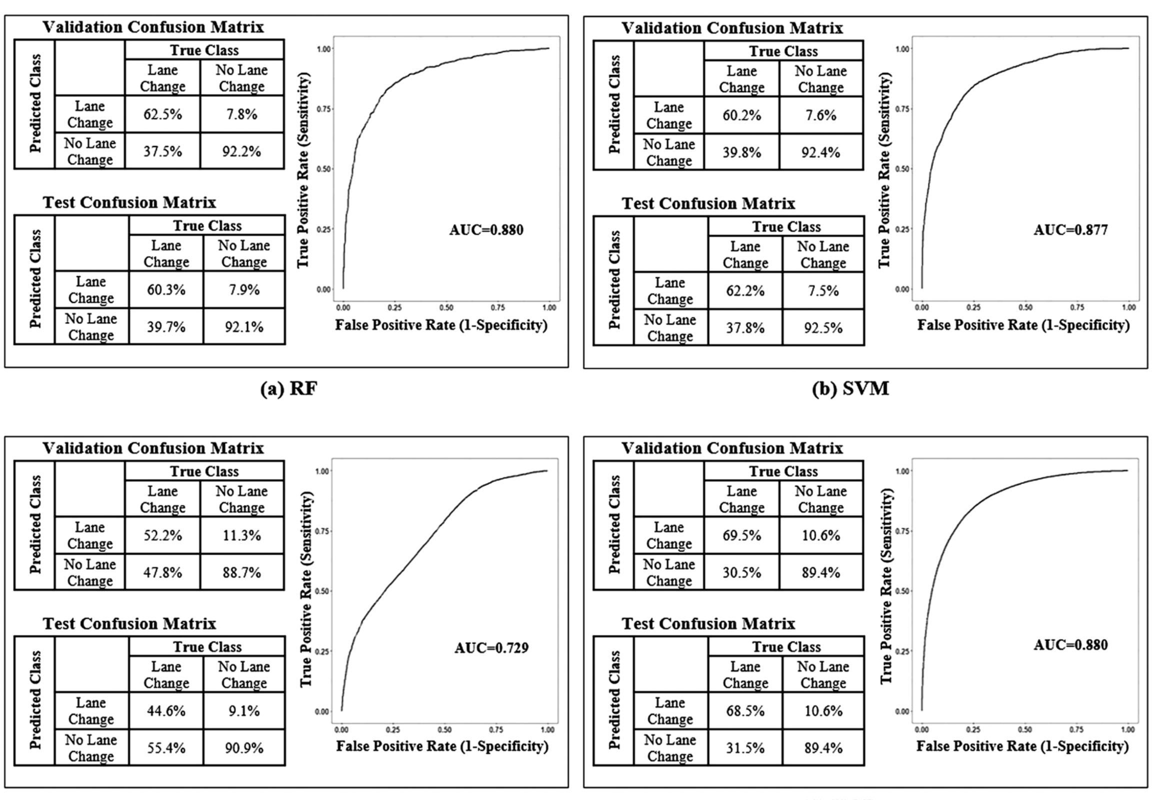


Fig. 8. Detection Summary of the Four Machine Learning Models Using Features Based on Category 3.

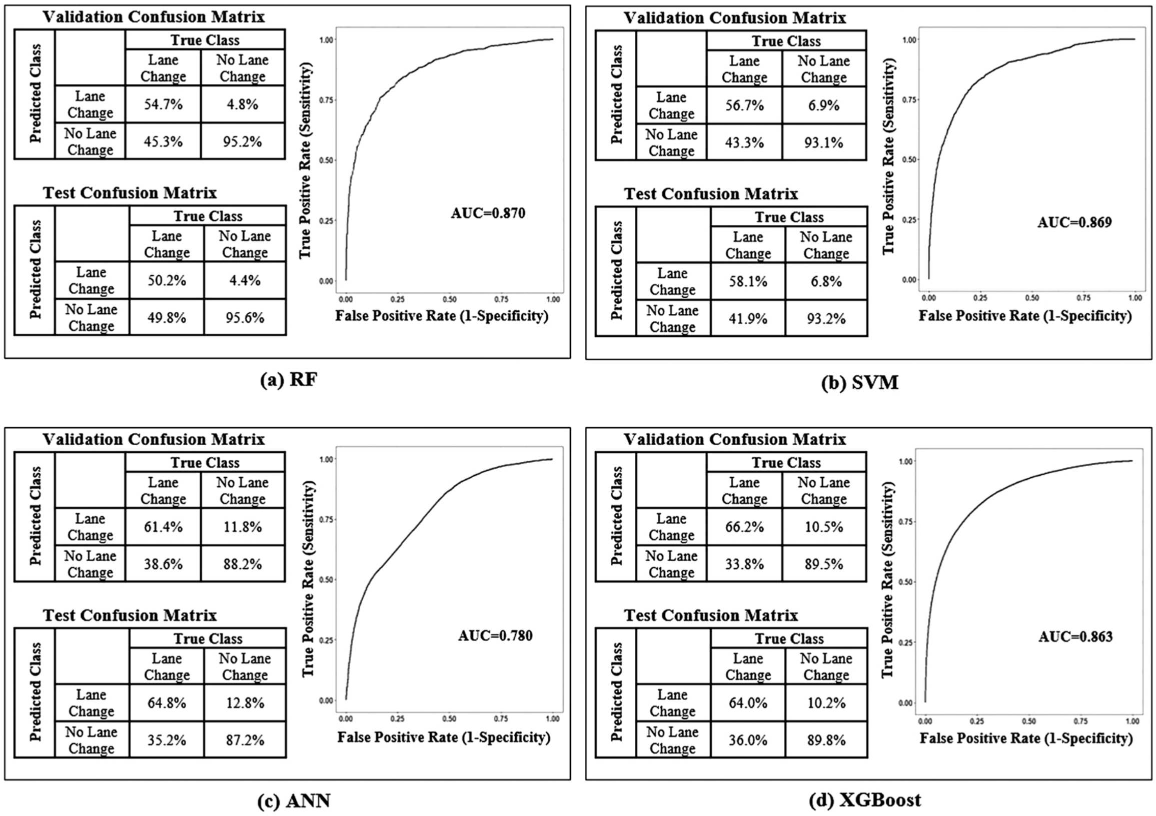


Fig. 9. Detection Summary of the Four Machine Learning Models Using Features Based on Category 4.

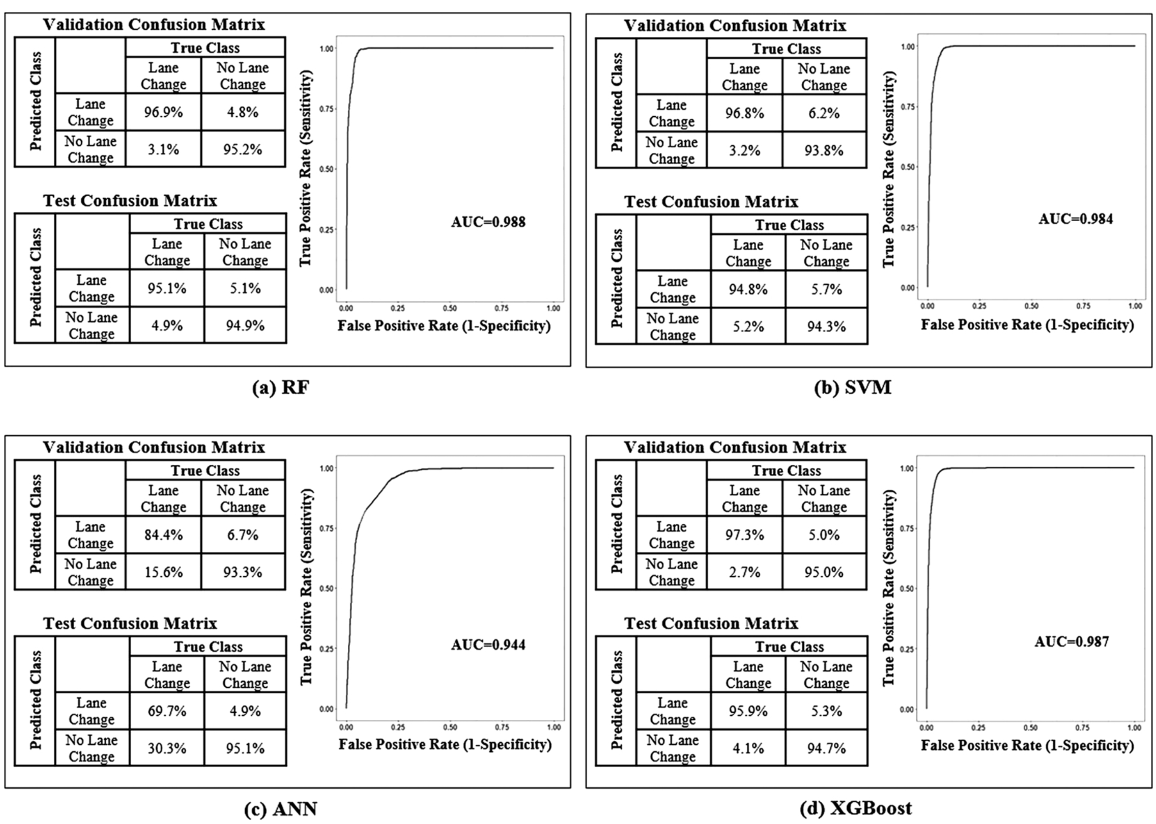


Fig. 10. Detection Summary of the Four Machine Learning Models Using Features Based on Category 5.

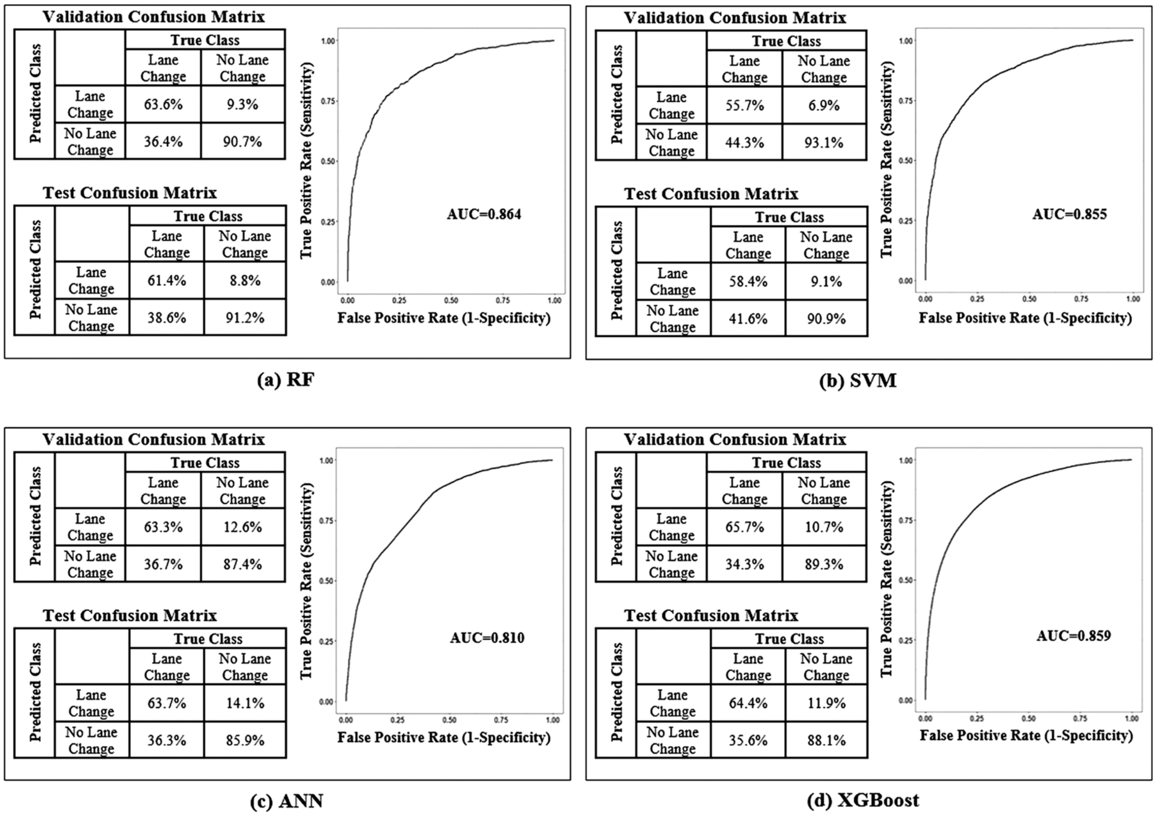


Fig. 11. Detection Summary of the Four Machine Learning Models Using Features Based on Vehicle Kinematics Only.

features from vehicle kinematics and driver demographics based on the highest overall accuracy and the AUC value of the trained models.

4.1.5. Category 5 – fusing machine vision and roadway characteristics

The overall detection accuracy of the four Machine Learning models was found to be similar, as observed in Category 1, utilizing features from machine vision and roadway characteristics. The XGBoost model produced the highest overall detection accuracy of 95.8 % and 95.1 % during validation and testing, respectively. It is worth mentioning that the highest overall detection accuracy was also observed in RF model for validation dataset. Fig. 10 displays the overall detection summary for Category 5 features. The highest true positive rate (97.3 %) and lowest false negative rate (2.7 %) were found in the XGBoost model during validation indicating that only 3% of lane changes have been misclassified by the model. In addition, it was observed that RF and SVM had similar true positive rates during validation. Similar to the validation, the highest true positive rate (95.9 %) and lowest false negative rate (4.1 %) were found in the XGBoost model for the test dataset. However, the highest true negative rate of 95.2 % and lowest false positive rate of 4.8 % were found in the RF model during validation. The AUC values of the trained models were found to be greater than 0.9 representing high accuracy, as can be seen in Fig. 10. The findings were expected as machine vision-based features were considered in this category. According to the overall performance, the study recommended both XGBoost and RF models for detecting lane changes when using features from machine vision and roadway characteristics.

4.1.6. Category 6 – features based on vehicle kinematics only

Category 6 consists of features related to only vehicle kinematics. The detection summary of the Machine Learning models is shown in Fig. 11. While the highest overall detection accuracy was observed in the RF model (81.9 % during validation and 80.1 % during testing), the lowest detection accuracy of 79.6 % and 77.6 % were found for the ANN model during validation and testing, respectively. The highest true negative rate (93.1 %) and the lowest false positive rate (6.9 %) were found in SVM model during validation indicating that 93 % of no lane change maneuvers have been detected correctly by the classifier. On the contrary, the highest true negative rate (91.2 %) and the lowest false positive rate (8.8 %) were found in RF model during testing. Moreover, the XGboost model had the lowest false negative rate (34.3 %) and the highest true positive rate (65.7 %) during validation. The result suggested that around 66 % of lane changes have been detected as lane changes. The overall AUC values corresponding to the four models were found to be greater than 0.8. The finding indicated that the use of only vehicle kinematics based features could also produce moderate accuracy without including other features such as roadway characteristics or driver demographics. Considering the overall performance of all the Machine Learning models, the study suggested that the RF model would provide better lane change detection when only vehicle kinematics are available.

4.2. Effect of weather conditions on lane change detection

In order to observe the effect of weather conditions on lane change detection, additional analyses were conducted in clear, snow, rain, and fog weather conditions. In each weather condition, 300 "lane changes" and 600 "no lane changes" were considered for the analysis. Similar procedures used previously were applied for relevant feature selection and hyperparameter tuning. Table 3 shows the overall detection accuracy of the RF, SVM, ANN, and XGBoost models obtained for clear, snow, rain, and fog weather conditions in each data fusion category. As can be seen in Table 3, RF outperformed all other classifiers in detecting lane change maneuvers in Category 1, 2, and 4 in clear weather. However, the highest detection accuracy was obtained using the XGBoost model for rest of the categories under clear weather conditions during validation. Considering rainy weather conditions, RF model had the highest overall detection accuracy for all categories during validation. Moreover, the highest detection accuracy was obtained for the

Table 3
Lane Change Detection Summary of the Four Machine Learning Models Considering the Effect of Weather Conditions.

Categories	Models	Overall Accuracy (%)											
		Clear		Snow		Rain		Fog		Combined Data (Excluding Weather as Feature)		Combined Data (Including Weather as Feature)	
		Validation (%)	Testing (%)	Validation (%)	Testing (%)	Validation (%)	Testing (%)	Validation (%)	Testing (%)	Validation (%)	Testing (%)	Validation (%)	Testing (%)
Category 1	**RF**	96.7 %	99.4 %	92.6 %	93.9 %	96.7 %	98.3 %	98.1 %	97.8 %	95.7 %	95.0 %	95.9 %	95.0 %
	SVM	94.9 %	98.3 %	86.5 %	87.8 %	94.9 %	98.9 %	97.2 %	97.8 %	94.3 %	94.4 %	94.5 %	94.3 %
	ANN	96.1 %	96.7 %	85.6 %	85.0 %	95.3 %	98.3 %	96.7 %	96.7 %	93.9 %	92.9 %	93.8 %	93.1 %
	XGBoost	96.3 %	98.3 %	91.9 %	92.8 %	96.4 %	98.3 %	97.9 %	98.3 %	95.9 %	95.0 %	96.1 %	94.9 %
Category 2	**RF**	82.6 %	78.9 %	86.5 %	88.3 %	80.4 %	80.6 %	91.0 %	87.8 %	83.0 %	81.3 %	84.0 %	81.8 %
	SVM	79.4 %	81.7 %	85.4 %	82.8 %	77.8 %	81.1 %	88.9 %	86.7 %	81.4 %	80.1 %	83.2 %	81.5 %
	ANN	74.9 %	77.2 %	78.8 %	75.0 %	75.6 %	75.6 %	80.1 %	77.2 %	75.9 %	74.0 %	77.4 %	74.6 %
	XGBoost	82.5 %	80.0 %	86.9 %	85.6 %	79.7 %	81.7 %	91.8 %	85.6 %	82.5 %	80.0 %	84.5 %	82.9 %
Category 3	**RF**	81.7 %	78.3 %	86.1 %	89.4 %	80.8 %	79.4 %	88.1 %	85.6 %	82.6 %	80.3 %	83.9 %	82.4 %
	SVM	81.1 %	80.0 %	85.0 %	84.4 %	77.1 %	81.1 %	85.7 %	86.7 %	82.0 %	81.3 %	83.2 %	80.6 %
	ANN	78.3 %	80.6 %	76.8 %	62.8 %	72.1 %	78.3 %	78.5 %	77.2 %	76.9 %	73.8 %	77.2 %	73.1 %
	XGBoost	81.9 %	77.2 %	85.7 %	87.2 %	79.4 %	81.7 %	89.6 %	86.1 %	83.0 %	81.7 %	84.5 %	84.2 %
Category 4	**RF**	81.0 %	80.6 %	86.0 %	87.8 %	79.7 %	78.9 %	90.7 %	86.7 %	82.0 %	78.8 %	84.0 %	81.4 %
	SVM	80.0 %	81.7 %	84.2 %	84.4 %	77.1 %	79.4 %	89.6 %	86.7 %	81.3 %	80.1 %	83.4 %	81.1 %
	ANN	78.3 %	80.6 %	84.2 %	84.4 %	75.4 %	74.4 %	86.8 %	85.6 %	79.5 %	78.9 %	80.6 %	79.7 %
	XGBoost	80.8 %	77.8 %	85.1 %	86.1 %	79.3 %	78.9 %	90.6 %	87.2 %	82.0 %	80.3 %	84.1 %	80.0 %
Category 5	**RF**	96.1 %	98.3 %	91.3 %	91.1 %	96.7 %	98.3 %	97.9 %	96.7 %	95.8 %	95.0 %	96.0 %	94.6 %
	SVM	96.0 %	98.3 %	86.9 %	87.2 %	95.8 %	98.9 %	96.0 %	96.7 %	94.8 %	94.4 %	95.2 %	94.2 %
	ANN	96.0 %	97.2 %	85.3 %	90.0 %	95.8 %	98.3 %	97.1 %	96.1 %	90.4 %	85.7 %	94.6 %	93.8 %
	XGBoost	96.4 %	98.3 %	91.4 %	88.9 %	96.3 %	98.9 %	97.8 %	97.2 %	95.8 %	95.1 %	96.2 %	95.3 %
Category 6	**RF**	80.4 %	78.9 %	85.0 %	86.7 %	79.2 %	79.4 %	87.8 %	85.6 %	81.9 %	80.1 %	83.6 %	81.1 %
	SVM	79.4 %	80.6 %	83.8 %	83.3 %	75.3 %	80.6 %	85.7 %	85.0 %	81.0 %	78.9 %	83.3 %	80.6 %
	ANN	79.2 %	66.1 %	83.6 %	79.4 %	73.6 %	75.0 %	86.0 %	85.0 %	79.6 %	77.6 %	82.3 %	79.4 %
	XGBoost	80.6 %	77.2 %	85.1 %	84.4 %	77.5 %	80.0 %	88.6 %	83.9 %	81.7 %	79.3 %	84.4 %	81.5 %

Note: Category 1: Vehicle Kinematics + Machine Vision + Roadway Characteristics + Driver Demographics; Category 2: Vehicle Kinematics + Roadway Characteristics + Driver Demographics; Category 3: Vehicle Kinematics + Roadway Characteristics; Category 4: Vehicle Kinematics + Driver Demographics; Category 5: Machine Vision + Roadway Characteristics; Category 6: Vehicle Kinematics.

XGBoost model in Category 2 and 6, and RF model in Category 1 and 4 in snowy weather. It is worth noting that similar results were obtained in foggy weather during validation. Table 3 illustrates that while RF and XGBoost outperformed in Category 3 and 5, respectively, under snowy weather, the models had opposite results in those categories under foggy weather. Although all the classifiers were able to detect lane change maneuvers considering Category 1 and 5 with outstanding accuracies under clear, rain, and fog weather conditions, as shown in Table 3; the classifiers did not maintain similar performance when detecting the maneuvers under snowy weather. As discussed earlier, machine-vision based lane position offset might not work accurately where lane markings are not properly visible (such as in extreme snowy condition), which might result in lower detection accuracy compared to other weather conditions. In addition, weather was considered as an additional feature in each data fusion category considering combined dataset. Table 3 also demonstrates the overall detection accuracy of the four Machine Learning models using weather as a new feature in addition to all the previously used features. It was observed that the overall detection accuracy was increased for each data fusion category after including the weather conditions.

4.3. Analysis of early lane change maneuvers

Lane change detection time is an important index for evaluating lane change detection algorithms. The performance of a detection algorithm might be related to real-time identification of lane changes. Real-time detection of lane change is also essential for the planning of Autonomous Vehicles (AVs) in an attempt to improve cooperative driving. AVs can share information, intentions, and plans for changing lanes, overtaking, merging, etc. to other surrounding vehicles. Based on the shared information, specific vehicle makes decisions to cut-in or yield to the right-of-way (i.e., change lanes, slow down, etc.) to other vehicles in the vicinity (Wang et al., 2018). If the detection algorithm can identify the lane change maneuvers earlier (i.e., before the vehicle crosses the lane line), the information could be provided to enhance the safety and mobility of cooperative and automated vehicles. Considering the importance of detection time, further analysis was conducted to observe the capability of the developed algorithms to detect lane change maneuvers in real time. It is worth mentioning that the detection time of lane change maneuvers is different for AVs compared to human-driven vehicles. Unlike the AVs that can detect the lane change maneuvers within a fraction of seconds, detection time would be higher for human-driven vehicles. Note that the overall detection accuracy would be decreased for higher detection time. A previous study using naturalistic driving data suggested that drivers initiated lane change maneuvers within 5 s of the vehicle crossing the lane line (Chen et al., 2017). Therefore, prediction of drivers' intention to change lanes as far as 5 s in advance were considered and data within 5 s prior to lane change were extracted for further analysis. A 2400 "no lane changes" segments and 1200 segments correspond to "within 5 s before lane changes" were considered for the analysis. In order to obtain the segments correspond to "within 5 s before lane changes", the time stamps of the start of lane changes have been identified using the algorithm/procedure described in Data acquisition and processing section. All the segments from dynamic segmentation were checked in MATLAB environment, and if those segments fall within 5 s of lane changes, they were considered for the analysis of lane change prediction. The same previously used features related to vehicle kinematics, machine vision, roadway characteristics, and driver demographics were considered for the six data fusion categories. Relevant features were selected using Boruta wrapper algorithm and hyperparameters of the four Machine Learning algorithms were tuned utilizing similar processes in each category. The detection summary of the RF, SVM, ANN, and XGBoost models using features based on six categories are presented in Table 4 in the form of overall detection accuracy. As can be seen in Table 4, the overall detection accuracy of real-time lane change maneuvers ranged from 67 % to 77 % with an average of about 72 % during validation. Considering the test dataset, the overall detection accuracy was 66%–75% with an average of around 71 %. The findings summarized that the developed detection algorithm could predict lane changes within 5 s before a vehicle crosses the lane line if the detection accuracy could be sacrificed to some extent, as supported by a previous study (Katrakazas et al., 2020). Moreover, the overall detection accuracy was increased with an average of 73 % during validation and an average of 72 % during testing after considering weather as a feature, as observed previously.

5. Conclusions

This study explored the potential of different Machine Learning approaches to detect lane change maneuvers. The data were selected randomly from the NDS trips that were acquired from the massive USDOT-FHWA SHRP2 Implementation Assistance Program (IAP). These datasets comprise data from six US states. Therefore, the results from this study are more reliable, geographically representative, and relevant to the US.

Considering the fact that utilizing fixed time interval could have the potential of missing lane change events with durations longer than the specified time, this study adopted a dynamic segmentation method to capture all the lane changes ranging from 1 s to 16 s. A combined dataset of 1200 "lane changes" using manual annotation process and 2400 "no lane changes" utilizing dynamic segmentation approach were created and considered for the analysis.

Six data fusion categories were introduced, for the first time, in this study considering different data availability and utilizing a comprehensive set of features. These features encompass vehicle kinematics (i.e., speed, acceleration, and yaw rate), machine vision (i.e., lane position offset), and driver demographics (e.g., age, gender, vehicle type, driving experience, etc.) under different weather conditions from the SHRP2 NDS. In addition, roadway geometrics features, such as presence of curve, curve radius, superelevation, number of lanes, speed limit, etc. were also extracted from the Roadway Information Database (RID). These features were considered to train lane change detection models.

Subsequently, four Machine Learning algorithms including, Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and a cutting-edge classifier, eXtrem Gradient Boosting (XGBoost) were utilized in this study. The classifiers are extensively recognized among researchers due to their interpretability, accuracy, and simplicity and have unique capabilities to solve several classification problems (Bakhit et al., 2017; Khan and Ahmed, 2019; Soleimani et al., 2019). The four classifiers have been used to compare their performance in terms of detecting lane change maneuvers and find the best possible model in each data fusion category. The application of different classifiers for detecting lane change maneuvers was conducted for the first time in this study. All the Machine Learning algorithms were trained, validated, and tested using the extracted features in six categories. In each category, wrapper algorithm was employed to select relevant features and plots (i.e., Figs. 2–4) were developed to observe the final confirmed and rejected features. Therefore, features considered in each category were relevant based on the plots and had significant contributions in detecting lane change maneuvers. In addition, parameters of all the classifiers were tuned utilizing 5-fold cross-validation technique for each step. Afterward, the validation and testing results were comparatively evaluated using confusion matrix across the models in terms of model performance. Moreover, the effect of weather conditions on lane change detection were investigated through overall detection accuracy obtained from separate analysis in each weather condition and including weather as a new feature in each data fusion category. Finally, the capability of the developed detection algorithm was observed for the "no lane changes" segments and separately created segments correspond to "within 5 s before lane changes" using the overall detection accuracy of the models.

Table 4
Early Lane Change Detection Summary of the Four Machine Learning Models.

Categories	Models	Overall Accuracy (%)			
		Combined Data (Excluding Weather as a Feature)		Combined Data (Including Weather as a Feature)	
		Validation (%)	Testing (%)	Validation (%)	Testing (%)
Category 1: Vehicle Kinematics + Machine Vision + Roadway Characteristics + Driver Demographics	RF	77.3 %	72.5 %	77.4 %	76.3 %
	SVM	72.2 %	70.6 %	72.8 %	72.4 %
	ANN	67.1 %	66.3 %	67.2 %	67.1 %
	XGBoost	75.1 %	73.8 %	75.7 %	76.4 %
Category 2: Vehicle Kinematics + Roadway Characteristics + Driver Demographics	RF	76.3 %	74.0 %	76.9 %	75.3 %
	SVM	72.0 %	70.4 %	72.5 %	71.7 %
	ANN	68.0 %	68.9 %	69.2 %	66.3 %
	XGBoost	75.0 %	72.5 %	75.0 %	74.2 %
Category 3: Vehicle Kinematics + Roadway Characteristics	RF	76.5 %	74.3 %	76.5 %	74.6 %
	SVM	71.2 %	69.6 %	71.6 %	71.3 %
	ANN	68.3 %	66.1 %	69.0 %	66.3 %
	XGBoost	74.4 %	72.4 %	74.4 %	74.2 %
Category 4: Vehicle Kinematics + Driver Demographics	RF	75.9 %	73.2 %	76.4 %	73.2 %
	SVM	71.7 %	69.9 %	72.2 %	72.2 %
	ANN	70.4 %	69.7 %	71.3 %	72.1 %
	XGBoost	73.2 %	73.5 %	74.8 %	72.1 %
Category 5: Machine Vision + Roadway Characteristics	RF	73.4 %	72.9 %	73.5 %	72.5 %
	SVM	68.6 %	67.8 %	68.3 %	68.8 %
	ANN	67.3 %	65.8 %	67.2 %	65.3 %
	XGBoost	70.6 %	68.3 %	71.0 %	70.0 %
Category 6: Vehicle Kinematics only	RF	75.9 %	74.9 %	76.5 %	73.1 %
	SVM	70.8 %	68.5 %	71.1 %	70.6 %
	ANN	72.1 %	67.8 %	71.9 %	70.8 %
	XGBoost	72.9 %	72.2 %	74.3 %	72.8 %

The highest overall detection accuracy was found around 96 % using XGBoost, which is higher compared to most of the previous studies (Yang et al., 2017; Zheng et al., 2014). More precisely, with respect to features based on vehicle kinematics, machine vision, roadway characteristics, and driver demographics (i.e., Category 1), the XGBoost model provided the highest overall detection accuracy of around 95.9 % and 95 % during validation and testing, respectively. In addition, it was found that the trained RF model provided the highest accuracy of 81.9 % and 80.1 % during validation and testing, respectively, utilizing features based on only vehicle kinematics (i.e., Category 6), indicating that the vehicle kinematics features have the potential to detect lane change maneuver. Moreover, XGBoost was found to be the best model when considering Category 1, and 3, and RF was found to be the best model for Category 2, 4, and 6. In addition, both XGBoost and RF models turned out to be the best models for Category 5. Considering the results from the detection models, the study recommends to use XGBoost model if the machine vision-based data (e.g., lane position offset) are available and RF model if only vehicle kinematics data are available. In addition, the effect of weather conditions on lane change detections was investigated by integrating weather as an additional feature in all the categories, which revealed that weather could improve the lane change detection accuracy. Therefore, the study suggested to include weather in addition to the features related to vehicle kinematics, machine vision, roadway characteristics, and driver demographics for detecting lane change maneuvers. Moreover, separate analysis in each weather condition revealed that machine-vision based features in extreme weather conditions could negatively affect the lane change detection performance. The analysis of early lane change detection suggested that the developed detection algorithm could predict the lane changes within 5 s before the vehicles cross the lane line if the detection accuracy could be sacrificed to some extent.

As mentioned earlier, weather may impact the accuracy and quality of data, specifically for machine-vision based lane position offset. In order to assess the reliability of the lane position offset, probability of correctly interpreting right or left-side lane markings was utilized. Note that, the reliability variables were provided in the DAS developed by the VTTI. The probability values range from 0 to 1024, where higher values indicate better probability (Hallmark et al., 2015). This reliability could go down due to adverse weather conditions. Although significant patterns of lane position offset can be observed during the lane change maneuver, it might provide erroneous value during harsh weather conditions where the machine vision algorithm might not work properly due to poor visibility of lane marking. It is worth mentioning that the reliability of the lane position offset in this study was checked, which revealed that the probability was affected during heavy snow and rain. Therefore, it is not recommended to use the variable in extreme adverse weather. However, lane position offset could be considered if decent reliability is achieved in heavy rain or snow. If desired reliability could not found, other variables such as yaw rate and steering wheel angle might be considered. These variables could be extracted from CAN-bus and are not affected by adverse weather conditions.

The developed models can detect lane change maneuvers based on the available data with promising detection accuracy, which will provide valuable insights and appropriate guidance to transportation researchers in conducting lane change-related research. The study revealed that the trained models in Category 2, 3, 4, and 6 could detect lane change in any weather condition in absence of machine vision-based data. However, if accurate machine vision-based data were available, the study suggested to use the detection model in Category 1 and 5. By analyzing the lane changes detected from this study, it would be possible to examine the effect of weather conditions on lane change behavior, such as lane change duration, speed and acceleration during lane change, number of lane changes per mile, and so on.

In addition, using similar trajectory-level data like NDS, the developed detection models could be used to monitor and control driver behavior in a Connected Vehicle (CV) environment. Aggressive driving behavior (e.g., high frequency and high-speed lane changes) could be obtained through analyzing their lane change events detected by the proposed models. Similarly, safe drivers in terms of lane change behavior could be identified and incentives, such as lower insurance rates, could be provided to encourage safe driving. Moreover, in a CV environment, if unusual traffic patterns are detected in terms of lane changes, these roadway segments could be flagged, and appropriate countermeasures could be provided in a timely manner to reduce/prevent the risk of crashes.

6. Limitations and future studies

While the study exhibited the capability of Machine Learning algorithms to detect lane change maneuvers from the SHRP2 NDS and RID datasets, some limitations should be mentioned. The first limitation is related to the data segmentation approach. The study considered a dynamic segmentation approach to select non-lane change segments. This approach is appropriate and necessary when machine vision-based features (e.g., lane position offset in this study) are included in the dataset. Therefore, the study recommends to use fixed time window approach in the absence of machine vision data in the future. The second limitation is related to the steering wheel angle variable, which was not considered for detecting lane change behavior due to excessive missing values in most of the trips. In addition, data from all surrounding vehicles were not available as front-mounted radar of NDS vehicle cannot detect the presence of vehicles behind the NDS vehicle's lane. Consequently, the lane change was considered as a single behavior of the subject vehicle. Future studies can focus on incorporating data from all surrounding vehicles as the input of the detection algorithm using similar trajectory-level data with information from all surrounding vehicles. Moreover, the study is only limited to trips on freeways. Lane change detection models using the SHRP2 NDS data on urban roadways could be considered in future studies.

Furthermore, the continuation of this study is important to include other driver behavioral features in addition to driver demographics in the developed detection algorithm. To be specific, drivers' aggressiveness could be considered in future studies. The lane change events database developed in this study contained a number of features that could be used to classify drivers' aggressiveness. Features associated with driving behavior, such as speed, acceleration, yaw rate, speed differences from speed limit, number of lane changes per mile, etc. could be obtained, and then cluster analysis could be adopted with possible features to classify drivers as aggressive or conservative. Once all drivers are classified, they could be introduced as conservative or aggressive in the current model. Finally, the expansion of this study would be to include other relevant features and develop more advanced lane change detection models using Artificial Intelligence and Deep Learning for specific lane change types and extend the work to predict lane changes.

CRediT authorship contribution statement

Anik Das: Conceptualization, Methodology, Software, Formal analysis, Validation, Writing - original draft, Writing - review & editing. **Md Nasim Khan:** Conceptualization, Methodology, Software, Validation, Writing - review & editing. **Mohamed M. Ahmed:** Conceptualization, Methodology, Validation, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was conducted under the second Strategic Highway Research Program (SHRP2), which is administrated by the Transportation Research Board of the National Academies of Sciences, Engineering, and Medicine, and it was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials (AASHTO) and the Wyoming Department of Transportation (WYDOT).

References

Ahmed, M.M., Ghasemzadeh, A., Eldeeb, H., Gaweesh, S., Clapp, J., Ksaibati, K., Young, R., 2015. Driver performance and behavior in adverse weather conditions: an investigation using the SHRP2 naturalistic driving study data—phase 1. Publ. FHWA-WY-16/08F. Wyoming Dep. Transp. 307, 1–73.

Ahmed, M., Young, R., Ghasemzadeh, A., Hammit, B., Ali, E., Khan, N., Das, A., Eldeeb, H., 2017. Implementation of SHRP2 Results within the Wyoming Connected Vehicle Variable Speed Limit System: Phase 2 Early Findings Report and Phase 3 Proposal. Wyoming Department of Transportation (WYDOT).

Ahmed, M.M., Ghasemzadeh, A., Hammit, B., Khan, N., Das, A., Ali, E., Young, R.K., Eldeeb, H., 2018. Driver performance and behavior in adverse weather conditions: an investigation using the SHRP2 naturalistic driving study data-phase 2. Final Rep. WY-18/05F, Wyoming Dep. Transp.

Ali, E.M., Ahmed, M.M., Wulff, S.S., 2019. Detection of critical safety events on freeways in clear and rainy weather using SHRP2 naturalistic driving data: Parametric and non-parametric techniques. Saf. Sci. 119 (November), 141–149. https://doi.org/10.1016/j.ssci.2019.01.007.

Ayres, G., Wilson, B., LeBlanc, J., 2004. Method for identifying vehicle movements for analysis of field operational test data. Transp. Res. Rec. J. Transp. Res. Board 1886, 92–100. https://doi.org/10.3141/1886-12.

Bakhit, P.R., Osman, O.A., Ishak, S., 2017. Detecting imminent lane change maneuvers in connected vehicle environments. Transp. Res. Rec. J. Transp. Res. Board 2645 (2645). https://doi.org/10.3141/2645-18. 14p.

Bambrick, N., 2016. Support Vector Machines: A Simple Explanation [WWW Document]. URL https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html (accessed 7.26.18). .

Bogard, S., Fancher, P., 1999. Analysis of Data on Speed-Change and Lane-Change Behavior in Manual and Acc Driving. Publication DTNH22-94-Y-47016 U.S. Department of Transportation.

Campbell, K.L., 2012. The SHRP 2 Naturalistic Driving Study. Transp. Res. News, pp. 30–35.

Castrejón, L., 2016. Neural Networks Tutorial.

Center for Transportation Research and Education (CTRE), 2015. SHRP2-Roadway Information Database - Background [WWW Document]. URL http://www.ctre.iastate.edu/shrp2-rid/ (accessed 7.31.19). .

Chen, R., Gabler, H.C., Sherony, R., 2017. Predicting driver lane change maneuvers using vehicle kinematic data. 25th Int. Tech. Conf. Enhanc. Saf. Veh. 1–9.

Chicco, D., 2017. Ten quick tips for machine learning in computational biology. BioData Min. 10 (1), 1–17. https://doi.org/10.1186/s13040-017-0155-3.

Collett, D., 2003. Modelling Binary Data, 2nd editio. ed. Chapman and Hall/CRC, New York.

Das, A., Ahmed, M.M., 2019. Exploring the effect of fog on lane-changing characteristics utilizing the SHRP2 naturalistic driving study data. J. Transp. Saf. Secur. 1–26. https://doi.org/10.1080/19439962.2019.1645777.

Das, A., Ahmed, M.M., Ghasemzadeh, A., 2019a. Using trajectory-level SHRP2 naturalistic driving data for investigating driver lane-keeping ability in fog: an association rules mining approach. Accid. Anal. Prev. 129, 250–262. https://doi.org/10.1016/j.aap.2019.05.024.

Das, A., Ghasemzadeh, A., Ahmed, M.M., 2019b. Analyzing the effect of fog weather conditions on driver lane-keeping performance using the SHRP2 naturalistic driving study data. J. Safety Res. 68 (February), 71–80. https://doi.org/10.1016/j.jsr.2018.12.015.

Das, A., Khan, M.N., Ahmed, M.M., Wulff, S.S., 2019c. Evaluating the impact of Rain on lane-changing behavior using naturalistic driving data: cluster analysis and multi-level modeling approach. Proc. 99th Transp. Res. Board Annu. Meet.

Das, A., Khan, M.N., Ahmed, M.M., 2020. Nonparametric multivariate adaptive regression splines models for investigating lane-changing gap acceptance behavior utilizing strategic highway research program 2 naturalistic driving data. Transp. Res. Rec. J. Transp. Res. Board 1–16. https://doi.org/10.1177/0361198120914293.

Fitch, G.M., Lee, S.E., Klauer, S., Hankey, J., Sudweeks, J., Dingus, T., 2009. Analysis of Lane-Change Crashes and Near-Crashes, Final Report DOT HS 811 147. US Department of Transportation, National Highway Traffic Safety Administration doi:DOT HS 811 147.

Ghasemzadeh, A., Ahmed, M.M., 2018a. A framework to enhance the transferability of the SHRP2 naturalistic driving study by considering heterogeneity of driver behavior using spatial-temporal factors in a trajectory level. Proc. 2018 21st Int. Conf. Intell. Transp. Syst. 1826–1833. https://doi.org/10.1109/ITSC.2018.8569927.

Ghasemzadeh, A., Ahmed, M.M., 2018b. Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: non-parametric MARS and parametric logistic regression modeling approaches. Transp. Res. Part C Emerg. Technol. 90 (March), 379–392. https://doi.org/10.1016/j.trc.2018.03.018.

Ghasemzadeh, A., Ahmed, M.M., 2019. Quantifying regional heterogeneity effect on drivers' speeding behavior using SHRP2 naturalistic driving data: a multilevel modeling approach. Transp. Res. Part C Emerg. Technol. 106 (June), 29–40. https://doi.org/10.1016/j.trc.2019.06.017.

Ghasemzadeh, A., Hammit, B.E., Ahmed, M.M., Eldeeb, H., 2019. Complementary methodologies to identify weather conditions in naturalistic driving study trips: lessons learned from the SHRP2 naturalistic driving study & roadway information

database. Saf. Sci. 119 (November), 21–28. https://doi.org/10.1016/j.ssci.2019.01.006.
Goh, Y.M., Ubeynarayana, C.U., 2017. Construction accident narrative classification: an evaluation of text mining techniques. Accid. Anal. Prev. 108 (July), 122–130. https://doi.org/10.1016/j.aap.2017.08.026.
Hallmark, S.L., Oneyear, N., Tyner, S., Wang, B., 2015. Analysis of Naturalistic Driving Study Data : Roadway Departures on Rural Two-Lane Curves. SHRP2 Rep. S2-S08D-RW-1 91.
Hammit, B., James, R., Ahmed, M., 2018. Radar-vision algorithms to process the trajectory-level driving data in the SHRP2 naturalistic driving study. In: IEEE Conf. Intell. Transp. Syst. Proceedings. ITSC 2018-Novem. pp. 2126–2133. https://doi.org/10.1109/ITSC.2018.8569688.
Hammit, B.E., James, R., Ahmed, M., Young, R., 2019. Towards the development of weather-dependent microsimulation models. Transp. Res. Rec. J. Transp. Res. Board 2673 (7), 143–156.
Hanowski, R.I., Wierwille, W.W., Garness, S.A., Dingus, T.A., 1999. A Field evaluation of safety issues in Local/Short haul trucking. In: Proc. Hum. Factors Ergon. Soc. Annu. Meet. 44 September.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. The Mathematical Intelligencerhttps://doi.org/10.1198/jasa.2004.s339.
Hou, Y., Edara, P., Sun, C., 2015. Situation assessment and decision making for lane change assistance using ensemble learning methods. Expert Syst. Appl. 42 (8), 3875–3882. https://doi.org/10.1016/j.eswa.2015.01.029.
Hutton, J.M., Bauer, K.M., Fees, C.A., Smiley, A., 2014. Analysis of Naturalistic Driving Study Data: Offset Left-Turn Lanes. SHRP2 REPORT S2-S08B-RW-1 Analysishttps://doi.org/10.17226/22315.
Insight Website, 2019. SHRP2 NDS Data Access [WWW Document]. URL https://insight.shrp2nds.us/ (accessed 4.15.19). .
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: With Applications in R. https://doi.org/10.1007/978-1-4614-7138-7_8.
Katrakazas, C., Mahajan, V., Antoniou, C., 2020. Prediction of lane-changing maneuvers with automatic labeling and deep learning. Proc. 99th Transp. Res. Board Annu. Meet.
Khan, M.N., Ahmed, M.M., 2019. Snow detection using in-vehicle video camera with texture-based image features utilizing K-Nearest neighbor, support vector machine, and random forest. Transp. Res. Rec. J. Transp. Res. Board 2673 (8), 221–232. https://doi.org/10.1177/0361198119842105.
Khan, M.N., Ahmed, M.M., 2020. Trajectory-level fog detection based on in-vehicle video camera with TensorFlow deep learning utilizing SHRP2 naturalistic driving data. Accid. Anal. Prev. 1–42 In-Press.
Khan, M.N., Ghasemzadeh, A., Ahmed, M.M., 2018. Investigating the impact of fog on freeway speed selection using the SHRP2 naturalistic driving study data. Transp. Res. Rec. J. Transp. Res. Board 2672 (16), 93–104.
Knoop, V.L., Hoogendoorn, S.P., Shiomi, Y., Buisson, C., 2012. Quantifying the Number of Lane Changes in Traffic. Transp. Res. Rec. J. Transp. Res. Board 2278 (1), 31–41. https://doi.org/10.3141/2278-04.
Koziol, J., Inman, V., Carter, M., Hitz, J., Najm, W., Chen, S., Lam, A., Penic, M., Jensen, M., Baker, M., Robinson, M., Goodspeed, C., 1999. Evaluation of the intelligent cruise control system volume II - appendices. Proc. 78th Transp. Res. Board Annu. Meet.
Kumar, S.S., Shaikh, T., 2017. Empirical evaluation of the performance of feature selection approaches on random Forest. In: 2017 Int. Conf. Comput. Appl. ICCA 2017. pp. 227–231. https://doi.org/10.1109/COMAPP.2017.8079769.
Kumar, P., Perrollaz, M., Lefevre, S., Laugier, C., 2013. Learning-based approach for online lane change intention prediction. IEEE Intell. Veh. Symp IV, 797–802.
Kursa, M.B., 2018. Package 'Boruta.' R J. 15.
Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the boruta package. J. Stat. Softw. 36 (11), 1–13. https://doi.org/10.18637/jss.v036.i11.
Lek, S., Park, Y.S., 2008. Multilayer perceptron. Encycl. Ecol.
Li, X., Wang, W., Zhang, Z., Rötting, M., 2018. Effects of feature selection on lane-change maneuver recognition: an analysis of naturalistic driving data. J. Intell. Connect. Veh. 1 (3), 85–98. https://doi.org/10.1108/jicv-09-2018-0010.
Mandalia, H.M., Salvucci, D.D., 2005. Using support vector machines for lane-change detection. Proc. Hum. Factors Ergon. Soc. 1965–1969.
Mattes, S., 2003. The lane change task as a tool for driver distraction evaluation. DaimlerChrysler AG Res. Technol. 1–30.
Mazibuko, T., Mashao, D.J., 2006. Feature Extraction and Dimensionality Reduction in SVM Speaker Recognition. Margin 1, S2.
McDowell, I., 2006. Measuring Health: A Guide to Rating Scales and Questionnaires. Oxford University Press, USA.
Miller, R., Srinivasan, G., 2005. Determination of lane change maneuvers using naturalistic driving data. Proc. 19th Int. Tech. Conf. Enhanc. Saf. Veh. 1–5.
NHTSA, 2015. Traffic Safety Facts 2015: a Compilation of Motor Vehicle Crash Data From the Fatality Analysis Reporting System and the General Estimates System. U.S. Department of Transportationhttps://doi.org/10.1016/j.annemergmed.2013.12.004.
Nishida, K., 2017. Introduction to Extreme Gradient Boosting in Exploratory [WWW Document]. URL https://blog.exploratory.io/introduction-to-extreme-gradient-boosting-in-exploratory-7bbec554ac7.
Papathanasopoulou, V., Antoniou, C., 2018. Identification of lane changes manoeuvreson mixed traffic trajectory data. Proc. 97th Transp. Res. Board Annu. Meet. 1–17 August 2017.
Peng, F., Feng, F., Mccallum, A., 2004. Chinese segmentation and New word detection using conditional random fields. Proc. 20th Int. Conf. Comput. Linguist.
Poona, N.K., Ismail, R., 2014. Using Boruta-selected spectroscopic wavebands for the asymptomatic detection of fusarium circinatum stress. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (9), 3764–3772. https://doi.org/10.1109/JSTARS.2014.2329763.
Reimer, B., Donmez, B., Lavallière, M., Mehler, B., Coughlin, J.F., Teasdale, N., 2013. Impact of age and cognitive demand on lane choice and changing under actual highway conditions. Accid. Anal. Prev. 52, 125–132. https://doi.org/10.1016/j.aap.2012.12.008.
Rudnicki, W.R., Wrzesień, M., Paja, W., 2015. All Relevant Feature Selection Methods and Applications. Springer-Verlag, Berlin Heidelberg.
Salvucci, D.D., 2004. Inferring driver intent: a case study in lane-change detection. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 48 (19), 2228–2231. https://doi.org/10.1177/154193120404801905.
Salvucci, D.D., Liu, A., 2002. The time course of a lane change: driver control and eye-movement behavior. Transp. Res. Part F Traffic Psychol. Behav. 5 (2), 123–132. https://doi.org/10.1016/S1369-8478(02)00011-6.
Saraswat, M., 2019. Practical Tutorial on Random Forest and Parameter Tuning in R [WWW Document]. URL https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/tutorial-random-forest-parameter-tuning-r/tutorial/.
Sarkar, S., Pramanik, A., Maiti, J., Reniers, G., 2020. Predicting and analyzing injury severity: a machine learning-based approach using class-imbalanced proactive and reactive data. Saf. Sci. 125 (January), 1–23. https://doi.org/10.1016/j.ssci.2020.104616.
Shaheen, A., Iqbal, J., 2018. Spatial distribution and mobility assessment of carcinogenic heavy metals in soil profiles using geostatistics and Random Forest, Boruta Algorithm. Sustainability 10, 3. https://doi.org/10.3390/su10030799.
Soleimani, S., Mousa, S.R., Codjoe, J., Leitner, M., 2019. A comprehensive railroad-highway grade crossing consolidation model: a machine learning approach. Accid. Anal. Prev. 128 (March), 65–77. https://doi.org/10.1016/j.aap.2019.04.002.
Suthaharan, S., 2016. Support Vector Machine - Machine Learning Models and Algorithms for Big Data Classification. Springerhttps://doi.org/10.2514/6.2003-4110.
Thiemann, C., Treiber, M., Kesting, A., 2008. Estimating acceleration and lane-changing dynamics based on NGSIM trajectory data. Transp. Res. Rec. J. Transp. Res. Board 90–101. https://doi.org/10.3141/2088-10.
Tijerina, L., Garrott, W., Stoltzfus, D., Parmer, E., 2005. Eye glance behavior of Van and passenger Car drivers during lane change decision phase. Transp. Res. Rec. J. Transp. Res. Board 1937, 37–43. https://doi.org/10.3141/1937-06.
Toledo, T., Zohar, D., 2007. Modeling duration of lane changes. Transp. Res. Rec. J. Transp. Res. Board 1999, 71–78. https://doi.org/10.3141/1999-08.
Valous, N.A., Sun, D.W., 2012. Image Processing Techniques for Computer Vision in the Food and Beverage Industries, Computer Vision Technology in the Food and Beverage Industries. Woodhead Publishing Limitedhttps://doi.org/10.1533/9780857095770.1.97.
Vandenbroucke, J.P., Von Elm, E., Altman, D.G., Gøtzsche, P.C., Mulrow, C.D., Pocock, S.J., Poole, C., Schlesselman, J.J., Egger, M., 2007. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Epidemiology 18 (6), 805–835. https://doi.org/10.1097/EDE.0b013e3181577511.
Wang, N., Wang, X., Palacharla, P., Ikeuchi, T., 2018. Cooperative autonomous driving for traffic congestion avoidance through vehicle-to-vehicle communications. IEEE Veh. Netw. Conf. VNC 2018-Janua 327–330. https://doi.org/10.1109/VNC.2017.8275620.
Wang, X., Yang, M., Hurwitz, D., 2019. Analysis of cut-in behavior based on naturalistic driving data. Accid. Anal. Prev. 124 (March), 127–137. https://doi.org/10.1016/j.aap.2019.01.006.
Xi, Y., Crisler, M., 2013. A review of lane change definitions and identification methods. Proc. 92th Transp. Res. Board Annu. Meet January.
Xuan, Yiguang, Coifman, B., 2006. Lane change maneuver detection from probe vehicle DGPS data. IEEE Intell. Transp. Syst. Conf. 624–629. https://doi.org/10.1109/ITSC.2006.1706811.
Yang, D., Qiu, X., Liu, Y., Wen, C., Zhu, L., Hong, X., 2017. Modeling the discretionary lane-changing decision behavior using random Forest theory. Proc. 96th Transp. Res. Board Annu. Meet.
You, F., Zhang, R., Lie, G., Wang, H., Wen, H., Xu, J., 2015. Trajectory planning and tracking control for autonomous lane change maneuver based on the cooperative vehicle infrastructure system. Expert Syst. Appl. 42 (14), 5932–5946. https://doi.org/10.1016/j.eswa.2015.03.022.
Zhao, D., Lam, H., Peng, H., Bao, S., LeBlanc, D.J., Nobukawa, K., Pan, C.S., 2017. Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. IEEE trans. Intell. Transp. Syst. 18 (3), 595–607. https://doi.org/10.1109/TITS.2016.2582208.
Zheng, Z., Ahn, S., Monsere, C.M., 2010. Impact of traffic oscillations on freeway crash occurrences. Accid. Anal. Prev. 42 (2), 626–636. https://doi.org/10.1016/j.aap.2009.10.009.
Zheng, J., Suzuki, K., Fujita, M., 2014. Predicting driver's lane-changing decisions using a neural network model. Simul. Model. Pract. Theory 42, 73–83. https://doi.org/10.1016/j.simpat.2013.12.007.
Zhou, Y., 2018. Deterioration and Optimal Rehabilitation Modelling for Urban Water Distribution Systems. CRC Press.