

DriveNet: A deep learning framework with attention mechanism for early driving maneuver prediction

Mohamed M'haouach¹, Abdellatif Sassioui², Afaf Bouhoute¹, Khalid Fardousse¹

¹LPAIS Laboratory, Department of Computer Science, Sidi Mohammed Ben Abdellah University, Fez, Morocco

²C3S Laboratory, Hassan II University, Casablanca, Morocco

Article Info

Article history:

Received Mar 20, 2024

Revised Jul 18, 2024

Accepted Jul 26, 2024

Keywords:

Attention mechanism

Convolutional neural network

Long short-term memory

Maneuvers prediction

Neural network

Recurrent neural networks

Semi-autonomous vehicle

ABSTRACT

Inappropriate driving maneuvers are the leading cause of many car accidents. These accidents can be prevented if they are identified in advance and the driver is given the necessary assistance. Anticipating maneuvers is crucial for driving assistance systems in order to alert drivers and take appropriate measures to avoid or mitigate danger. In this paper, we introduce DriveNet a new approach that combines information about the driver's behavior as well as the driving environment to predict the driving maneuvers. DriveNet utilizes a combination of convolutional neural network (CNN) and long short-term memory (LSTM) with attention mechanism to extract spatial information and capture long temporal dependencies. We evaluate DriveNet by performing a series of experiments using the publicly available Brain4Cars dataset. The findings show that the proposed approach achieves state-of-the-art performance and outperforms most previous methods. DriveNet has achieved an accuracy of 91.24%, a precision of 90.13%, and a recall of 91.44% for anticipation 4 seconds before the maneuvers occur.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Mohamed M'haouach

LPAIS Laboratory, Department of Computer Science, Sidi Mohammed Ben Abdellah University

Fez, Morocco

Email: mohamed.mhaouach@usmba.ac.ma

1. INTRODUCTION

Road safety has always been a major concern for governments all over the world. Statistical data illustrates a direct correlation between the surge in vehicle numbers and the subsequent increase in accidents. The World Health Organization (WHO) [1] reports a staggering annual death toll of approximately 1.35 million individuals resulting from road accidents. The primary cause of these accidents is attributed to improper driving behaviors. In 2022, the British Department for Transport (DfT) conducted a survey [2], revealing that 29,795 individuals were either killed or seriously injured in reported accidents within Britain. Furthermore, the National Highway Traffic Safety Administration [3] indicates that 33% of accidents were caused by illegal maneuvers. To increase road safety and decrease the number of accidents, advanced driver assistance system (ADAS) [4]–[9] that can understand the driver's intention before performing any dangerous maneuver have been considered among the most significant advancement in this field of research. Through the assistance of driver intention prediction, it becomes simpler to ascertain the driver's readiness for a safe reaction, based on the relevance of the driver's maneuver intentions in line with the current driving scenario.

Early prediction of driving maneuvers is a fundamental task for numerous ADAS. For example, an ADAS equipped with a driving maneuver prediction feature can proactively notify the driver before executing a dangerous maneuver. This advanced warning grants the driver a slightly extended timeframe to respond to

road situations and potentially prevent accidents. Predicting driving maneuvers entails forecasting future driver actions based on a limited temporal context, presenting a formidable challenge due to the unobservable nature of driver intentions and the intricacies of their interactions with the road environment [10]. In fact, the behavior of drivers is greatly influenced by external factors, including traffic, road conditions, and weather conditions. These factors exert a substantial impact on how drivers respond and behave on the road. Moreover, when preparing for any maneuver, drivers usually follow some common behaviors including hand motion, head rotation, and eye gaze. Several studies show that data from multiple sensors (e.g. cameras capturing the driver's face, the road ahead, in-vehicle sensors, and global positioning coordinates (GPS) can provide contextual information for driver maneuver prediction. Several works [11]–[16] have attempted to predict driving maneuvers based on multimodal sensor data. However, existing research often neglects the time gap between maneuver prediction and maneuver execution. That is, most of studies don't shed light into the important time interval that drivers have to respond to predictions. Motivated by the need to address this issue, this paper introduces a system designed to improve prediction accuracy and developing strategies that take into account the exact time drivers have to react. Tackling this issue, as a result, will lead to the innovation of more effective predictive systems more effective and overall driver safety enhancement.

This paper tackles the challenge of predicting driver maneuvers by introducing and designing an end-to-end deep learning architecture called DriveNet. DriveNet addresses the existing issues and challenges in maneuver prediction by integrating driver information from videos captured by a driver-facing camera with environmental information, such as details about empty lanes, road artifacts, and speed limits. The main objective of this comprehensive approach is to develop a highly accurate system for predicting driver intentions. This innovative architecture leverages a combination of advanced techniques to achieve significant performance in maneuver prediction. Specifically, it integrates i) the VGG19 model for extracting rich spatial information from the different frames of videos, ii) the OpenFace framework [17] to extract a comprehensive face-based features, and iii) a bidirectional long short-term memory (BiLSTM) network enhanced with an attention mechanism to extract temporal dependencies of the different inputs. DriveNet was validated on the publicly available Brain4Cars dataset. The findings reveal that DriveNet shows superior performance compared to most of the previous methods. It achieved an accuracy of 91.24%, precision of 90.13%, and recall of 91.44% for predicting maneuvers 4 seconds before they occur. The contributions proposed in this paper are summarized as:

- We propose DriveNet as a new approach for early driving maneuvers' prediction. this approach consists of a VGG19 model for extracting spatial features, OpenFace framework for extracting face-based features, and a BiLSTM with an attention mechanism to extract temporal features.
- We conduct exhaustive experiments of the proposed architecture using the aforementioned datasets with different configurations and times to maneuvers (1s to 4s before the maneuver). Evaluation outcomes on the publicly available dataset Brain4Cars demonstrate that DriveNet outperforms other common models for driving maneuvers' prediction.

The remainder of this paper is divided into the following sections. Section 2 discusses related work on prediction of driving maneuvers. The section 3 introduces and details the DriveNet architecture. A presentation of the experimental results is given in section 4. Section 5 presents some open challenges with regard to data and user privacy. Finally, section 6 concludes the paper by summarizing our contributions and draws future research directions.

2. RELATED WORK

Over the past decade, there has been a growing interest among researchers in the field of driving maneuver prediction. This section is devoted to investigating and discussing some existing approaches for driving maneuver prediction using deep learning. The Brain4Cars team was among the first teams that worked on driving maneuvers anticipation [10]. Among its main contributions, the team released the first dataset of natural driving collected using a driver-facing camera to track the driver's head movements, a camera for the outside view, and some information about the environment such as road artifacts, empty lanes, and speed. To model the driving maneuvers, the authors use a Hidden Markov Model variant called AIO-HMM to jointly model the contextual information along with the maneuvers. The proposed AIO-HMM consists of three layers (input, hidden, and output). The input layer represents the outside vehicle features, the hidden layer represents the driver's intention, and the output layer represents features of the vehicle's inside. This system uses models,

which are trained for different types of maneuvers, to anticipate the probability of each maneuver. This method can anticipate maneuvers within 3.5 seconds before they occur with a precision of 77.5% and a recall of 71.4%.

An improved prediction approach is proposed by the Brain4Cars team in their second work [18]. In this latter, the team develops a deep learning sensory-fusion approach for maneuver anticipation. Instead of simple sensor fusion such as feature vector concatenation, their approach uses recurrent neural networks (RNNs) with long short-term memory (LSTM) units to jointly learn to anticipate maneuvers. This is done by using separate RNNs to learn high-level representations from the sensor streams. These representations are then fused via a fully connected layer. Using a deep sensory fusion learning technique, maneuvers could be predicted on average 3.5 seconds in advance, with 84.5% precision and 77.1% recall. By including additional data, like merging the driver's head orientation in 3D, accuracy and recall were able to rise to 90.5% and 87.4%, respectively.

A deep learning framework, which combines the information from the driver's monitoring videos with the outside view was proposed by [19]. This framework consists of two branches as inputs and a classifier that takes the output of the two branches. A ConvLSTM-based [20] encoder is utilized in the first branch to extract motion data, which is then interpreted into optical flow images. The second branch, a 3D ResNet-50 [21] network, uses the driver's face video to extract features. The classifier is composed of a motion decoder for outside motion and fully connected layers to predict the maneuver. This framework achieved an accuracy and f1-score respectively of 83.98% and 84.3% on the Brain4Cars dataset.

More recently, a model was proposed in [22] that utilizes both inside and outside videos as data sources. This model consists of four input sources, with the first two sources containing the main frames and the last two sources representing the optical flow [23], [24] of frames from inside and outside the cabin. To ensure a representative sample, frames were selected at a rate of 10, resulting in 15 frames for each 5-second video. The authors also incorporated four different data augmentation methods, namely translation, flip-left-to-right (FlipLR), cutout, and Augmix. Spatial feature extraction was performed using Densenet121 in the first two branches, while LSTM was employed to extract temporal features from all inputs. Remarkably, this architecture achieved exceptional performance metrics, with an accuracy of 98.90%, precision of 98.96%, and recall of 98.88% in accurately predicting maneuvers within the specified time to maneuver of 0.

A novel method of prediction that uses the SHRP2 Naturalistic Driving Study and roadway information dataset to train several models aimed at predicting driving maneuvers was proposed [25]. To select the most relevant features, the Boruta algorithm was employed. Among the various models examined, the XGBoost model emerged as the top performer, achieving an impressive prediction accuracy of 97% and an F1-score of 95.5% when considering all features. Notably, when focusing solely on vehicle kinematics features, the XGBoost model exhibited even higher accuracy, reaching 97.3%, with an F1-score of 95.9%. The researchers also developed simplified versions of the XGBoost model for practical implementation. This prediction model exhibits promising potential for trajectory planning in autonomous vehicles and can enhance ADAS within a connected and automated vehicle environment.

Mersch *et al.* [26] introduced a lane change prediction method that utilizes a data representation based on the surrounding to capture interactions between vehicles in highway driving scenarios. By integrating convolutional neural networks (CNNs), this system leverages spatial and temporal correlations, enabling accurate prediction of vehicle trajectories up to a five-second horizon. Notably, the model takes into consideration various potential maneuver intentions and their corresponding motions. The efficacy of this approach was evaluated using the HighD dataset [27] and NGISM dataset [28]. The results demonstrated a mean squared error (MSE) of 1.34 on the HighD dataset and 4.05 on the NGISM dataset in time to maneuver of 5 seconds.

3. PROPOSED APPROACH

In this section, we introduce DriveNet, a new approach for driving maneuver anticipation. The workflow of DriveNet is illustrated in Figure 1. As the figure shows, DriveNet consists of several steps, starting from data acquisition to maneuver classification. These steps are detailed in the following subsections.

3.1. Data acquisition

Nowadays, cars are equipped with different types of sensors that can be used to collect information about the driver, the vehicle, as well as the driving environment. In the context of driving maneuver prediction, relying on a single source of information (i.e. a single sensor) is not sufficiently rich. For instance, predicting maneuvers using only a camera capturing the driver's face may be challenging. Combining data from multiple

sensors (e.g., face camera, road camera, and GPS) is very helpful as it enables information about the whole driving situation to be used for prediction. DriveNet is designed to incorporate two distinct streams of data

- Face camera stream: this stream comprises data captured by the driver's facial camera, providing insights into the driver's facial expressions and movements.
- Environment information stream: this stream encompasses data related to the driving environment coming from vehicle sensors, such as speed, empty lanes, GPS coordinates, and potential road artifacts.

3.2. Preprocessing

The goal of data preprocessing is to transform raw stream data coming from the sensors and the camera videos into a format that is suitable for building a predictive model. In the preprocessing phase, DriveNet applies histogram equalization to the video frame images. This technique, which emphasizes contrast and intensity adjustments, is pivotal in enhancing the quality of input data for the following stages. Histogram equalization guarantees a more balanced spread of pixel intensities throughout images, thus improving their sharpness and the visibility of features. Figure 1, step 2 demonstrates the effect of this preprocessing step by comparing an image before and after histogram equalization. Furthermore, DriveNet incorporated normalization on environment features specifically for the speed variable. This normalization scales the speed values to a range between 0 and 1, ensuring consistency in the magnitude of this feature.

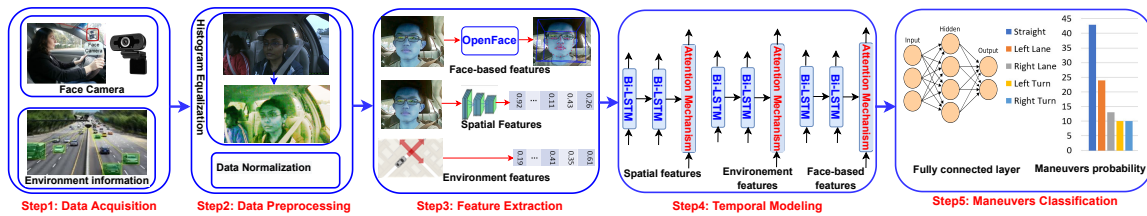


Figure 1. The overview of DriveNet

3.3. Feature extraction

The effectiveness of machine learning techniques is largely affected by how the input data is represented, or the features are chosen. The goal of feature extraction is to extract useful features from the input data that can improve the performance of the prediction model. DriveNet extracts three types of features from the raw data stream:

- Face-based features: Leveraging the face tracking framework OpenFace [17], DriveNet extracts 54 features from the face camera stream encompassing facial landmarks, eye gaze, and head pose. Figure 1, step 3 illustrates examples of face-based features.
- Spatial features: DriveNet uses the CNN VGG19 [29] architecture, extracting 256 features from frames extracted from the face camera stream.
- Environment features: DriveNet extracts from the input stream of sensors three pieces of information (speed, empty lanes, and artificial existence).

3.4. Temporal modeling

Driving maneuver prediction requires learning the temporal dependencies in data. Temporal dependencies are the relationships between the values of a variable in sequential data. These dependencies can be characterized by the way that the value of a variable at a given point in time is influenced by its past values. In the context of predicting driving maneuvers, learning temporal features from raw stream data can allow the car to anticipate and respond to changes in the driver's behavior or road conditions. DriveNet employs two Bi-LSTMs layers and an attention mechanism layer. For each of the three features extracted from the previous component, choosing Bi-LSTM over standard LSTM enhances DriveNet ability to capture long-term dependencies and temporal patterns.

3.5. Maneuver classification

The purpose of the final step is to classify the maneuvers. For this purpose, DriveNet aggregates the three outputs by summing them. Then, a fully connected network is used to classify the input into one of the targeted maneuvers. To achieve early prediction, classification is performed by time to maneuver, which

refers to the time before the maneuver happens. Algorithm 1 shows the different steps considered for maneuver prediction. The algorithm takes, as input, data coming from the face camera as well as from other sensors capturing the vehicle outside and follows the steps to predict the maneuver. This latter can be sent to ADAS for further assistance.

Algorithm 1 Maneuver prediction steps

Input: Face camera stream $X = X_0$
 Environment features $Z = Z_0$
Output: Maneuver
for each second t in time of driving **do**
 $X, Z \leftarrow \text{Preprocessing}(X_t, Z_t)$
 $X \leftarrow \text{Extract Features}(X_t)$
 $Z \leftarrow Z_t$
 $X_T \leftarrow \text{Extract Temporal Features}(X)$
 $Z_T \leftarrow \text{Extract Temporal Features}(Z)$
 $\text{Predict}(X_T, Z_T)$
 Send the results to ADAS
end for

4. EXPERIMENTAL RESULTS

This section provides a comprehensive overview of our experiments, including the datasets utilized, experimental settings, and evaluation.

4.1. Dataset description

DriveNet was evaluated using the Brain4Cars dataset [10]. This dataset consists of driving data collected from 10 drivers under real-world conditions, without any interference. The data include a variety of driving maneuvers performed by each driver. Each driver in the study performed at least one maneuver in all situations. The dataset includes videos taken from different angles: i) videos of the inside scenes, and ii) videos of the outside scenes with details of (1088×1920 px, 25 frame per second (fps)) and (480×720 px, 30 fps). It also contains further information about the outside environment, mainly information about empty lanes, road artifacts, and vehicle speed. The whole dataset is structured as follows:

- a. Inside features contain videos of the vehicle inside taken using a face camera.
- b. Outside features contain outside information. It mainly includes the following six features:
 - Id video: the identifier of the video (inside features) corresponding to the outside features.
 - Lane left: the number of empty lanes on the vehicle's left side.
 - Lane right: the number of empty lanes on the vehicle's right side.
 - Road artifact: a binary feature indicating the presence of road artifacts (such as intersections).
 - Speed: the vehicle speed. For each 5-second video, a sequence of 7 speeds ($v_1, v_2, v_3, v_4, v_5, v_6, v_7$) is provided.

Brain4Cars contains a total of 700 maneuvers, among which only 594 of them were accessible. These maneuvers belong to 5 classes, namely [left lane change (L change), right lane change (R change), left turn (L turn), right turn (R turn), straight driving]. In this study, We applied data augmentation using sliding window technique with a window duration of 1 second and stride of 0.6 seconds to increase the size of the dataset and improve the prediction performance. This step was driven by the very limited dataset, which includes only 594 maneuvers.

4.2. Experimental settings

4.2.1. Data splitting and evaluation

To evaluate DriveNet, we adopted a k-fold cross-validation with 5 as the number of folds. This method gives better results compared to the standard train test split. The performance was evaluated using three popular performance metrics, namely accuracy, precision, and recall.

4.2.2. Implementation details

The baseline models were implemented using Tensorflow and Keras libraries with Python 3.7.12 on a Kaggle environment. With the specifications of Intel(R) Xeon(R) CPU @ 2.30 GHz, 16 CPU cores, 12 GB RAM, and NVIDIA TESLA P100 GPU. The hyper-parameters used to train are presented in Table 1.

Table 1. Hyper-parameters for training the model

Hyper-parameter	Value
Epochs	100
Optimizer	RMSprop
Learning rate	0.001
Batch size	32
Loss function	Categorical entropy

4.2.3. Evaluation setup

According to some researchers [30], [31], the drivers' reaction time varies from person to person. In our evaluation, we assess the performance of DriveNet across the five different time to maneuvers intervals measured in second $t \in \{4, 3, 2, 1, 0\}$. We employ two methods for maneuver prediction:

- The first one is based on the current sequence only. That is, to predict the driver's intention with a time to maneuver t , we should classify the sequence $5 - t$. The final prediction is the result of the classification of this sequence.
- The second one is based on the current and the previous sequences. The driver's intention prediction with a time to maneuver t requires a classification of the sequence $5 - t$ (i.e. the third sequence in time to maneuver 2) and the previous sequences. Then, we aggregated the results using soft voting. This means that the final prediction is the class with the highest average probability.

4.3. Performance evaluation

In this section, the performance of DriveNet is studied. Various experiments were conducted studying the model performance in predicting different types of maneuvers, with varying time to maneuver values. The following paragraphs present and discuss the results of the sets of experiments performed. First, we study the model performance considering all types of maneuvers followed by a study focusing only on lane change and turn maneuvers, separately. A last experiment comparing the obtained results with existing approaches is presented.

4.3.1. Model Performance on all maneuver types

In this first scenario, we studied the performance of the DriveNet in predicting all maneuver classes. As described in subsection 4.2.3., we evaluated the prediction performance using two prediction methods: based on the current driving sequence only and based on the current and previous sequence. A summary of the 5-fold mean and standard deviation based on the model's findings is shown in Table 2. The obtained scores, for precision, recall, and accuracy, range between 90.23% and 92.34% which reflect the good performance of DriveNet. The results are presented at different times to maneuvers.

Table 2. Results on all maneuvers without aggregating the results of the current and the previous sequences

Time to maneuver (sec)	All maneuvers		
	Precision (%)	Recall (%)	Accuracy (%)
4	90.23 ± 2	90.91 ± 2	91.24 ± 3
3	92.01 ± 2	92.13 ± 2	91.66 ± 2
2	92.11 ± 2	91.81 ± 2	91.57 ± 2
1	91.57 ± 2	91.67 ± 2	91.67 ± 2
0	91.34 ± 2	91.41 ± 2	92.34 ± 2

The results obtained with our second prediction method, i.e. prediction based on current and previous sequences, are presented in Table 3. The table shows the precision, recall, and accuracy scores obtained for different times to maneuvers. The obtained scores show that the model performance improves when the time to maneuver decreases. This improvement is explained by the fact that the predictions, for each time to maneuver, take also into consideration the predictions made at the preceding instants.

Table 3. Results on all maneuvers aggregating the current and the previous sequences

Time to maneuver (sec)	All maneuvers		
	Precision (%)	Recall (%)	Accuracy (%)
4	90.63 \pm 3	90.32 \pm 2	91.24 \pm 3
3	92.21 \pm 2	92.44 \pm 2	92.51 \pm 2
2	93.18 \pm 2	93.01 \pm 2	94.33 \pm 2
1	94.27 \pm 3	94.01 \pm 2	96.23 \pm 2
0	95.34 \pm 2	95.41 \pm 2	96.23 \pm 2

4.3.2. Model performance on lane change and turn maneuvers

In the second scenario, the performance of the model is studied separately for each type of maneuver. We distinguish the two types of maneuvers presented in the following:

- Lane changes: we evaluate the model performance in anticipating left lane changes and right lane changes. This prediction is of relevance in the case of freeway driving.
- Turns: we evaluate the model performance in anticipating right and left turn maneuvers.

Similar to the previous subsection, the results are discussed when considering prediction based on the current sequence only, and on the current and previous sequences.

Table 4 shows the results of the model's performance in predicting lane changes and turn maneuvers, based only on the current sequence. By comparing the scores of the two types of maneuvers, we can see that DriveNet achieves high performance in anticipating right and left turns compared to the left and right lane changes. This means that the model can find more prominent features that differentiate turns and lane changes.

Table 4. Results on lanes change and turns based on the results of the current sequence

Time to maneuver (sec)	Lane change			Turns		
	Precision (%)	Recall (%)	Accuracy (%)	Precision (%)	Recall (%)	Accuracy (%)
4	89.82 \pm 7	89.82 \pm 5	86.67 \pm 7	96.19 \pm 2	96.45 \pm 4	96.46 \pm 3
3	90.13 \pm 7	90.13 \pm 5	86.69 \pm 8	97.21 \pm 2	97.34 \pm 4	96.46 \pm 3
2	91.42 \pm 5	91.42 \pm 6	87.67 \pm 5	97.21 \pm 2	97.31 \pm 4	96.46 \pm 3
1	90.51 \pm 6	90.51 \pm 7	87.10 \pm 6	96.85 \pm 7	96.55 \pm 5	92.85 \pm 7
0	89.42 \pm 7	89.42 \pm 8	89.67 \pm 8	97.33 \pm 4	97.21 \pm 2	96.42 \pm 3

Table 5 shows the results of the model's performance in predicting lane changes and cornering maneuvers based on current and previous sequences. By comparing Tables 4 and 5, we can see that, even though the performance of lane change maneuvers has been slightly improved, it remains far from that of turn maneuvers. We notice that the results remain consistent with those in Table 4.

Table 5. Results obtained on lanes change and turns based on aggregating the results of the current and previous sequences

Time to maneuver (sec)	Lane change			Turns		
	Precision (%)	Recall (%)	Accuracy (%)	Precision (%)	Recall (%)	Accuracy (%)
4	89.82 \pm 4	89.83 \pm 3	88.42 \pm 4	96.19 \pm 2	96.45 \pm 2	96.44 \pm 2
3	90.13 \pm 7	89.98 \pm 5	88.23 \pm 5	97.21 \pm 2	97.34 \pm 2	97.24 \pm 2
2	91.20 \pm 5	90.31 \pm 6	89.42 \pm 5	97.21 \pm 2	97.31 \pm 2	97.21 \pm 2
1	91.65 \pm 6	90.68 \pm 7	89.51 \pm 6	96.85 \pm 2	96.55 \pm 2	96.85 \pm 2
0	91.12 \pm 6	90.55 \pm 6	89.42 \pm 5	97.33 \pm 4	97.21 \pm 2	97.32 \pm 3

4.3.3. Comparison with the state of the art

Table 6 shows the performance of DriveNet method compared to other related approaches. The results show the precision, recall, and accuracy scores under three settings: lane change, turns, and all maneuvers. With respect to performance scores, we found that DriveNet performs better in most maneuvers compared to other approaches. It is over 88% accurate in predicting left and right lane change. In the case of anticipating turns, the accuracy of DriveNet is higher than 97%, indicating its efficacy in predicting this particular maneuver type. Moreover, when considering all maneuvers collectively, DriveNet's algorithm exhibits a comprehensive predictive capacity, achieving an accuracy of 91,24% with 4 seconds before the maneuver occurs.

Table 6. Comparison of DriveNet with presented state-of-the-art methods

Approaches	Time to maneuver (sec)	Lane change		Turns		All maneuvers	
		Precision (%)	Recall (%)	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Olabiya <i>et al.</i> [9]	3.5	-	-	-	-	77.4	71.2
Zhou <i>et al.</i> [32]	3.30	87.3	93.8	86.0	81.4	91.7	90.7
Jain <i>et al.</i> [18]	3.5	83.8	86.0	83.8	79.9	84.5	77.1
STA-Net [33]	0	-	-	-	-	90.8	91.1
DriveNet (Ours)	4	89.82	89.83	96.19	96.45	90.63	90.32

Comparing our results with other approaches, Olabiya *et al.* [9] introduced a technique called deep bidirectional recurrent neural network (DBRNN), which attained a precision rate of 77.4% and a recall rate of 71.4% when forecasting turns and lane changes. Although their suggested method allows for flexibility in maneuvering time by framing it as an anomaly detection issue, DriveNet outperforms DBRNN, notably achieving a precision of 90.63% and a recall of 90.32% specifically in predicting lane changes. On the other hand, Zhou *et al.* [32] introduced CF-RNN, achieving an F1-score of 91.2%, a precision of 91.7%, and a recall of 90.7% on the Brain4Cars dataset. Despite similarities in outcomes, DriveNet demonstrates comparable performance while extending the predictive time to maneuver 4 seconds, surpassing CF-RNN's capability, which is limited to 3.30 seconds.

In their study, Jain *et al.* [18] put forward a sensory-fusion technique based on deep learning, employing RNNs with LSTM units. Their method successfully predicted maneuvers with an average lead time of 3.5 seconds, demonstrating a precision rate of 84.5% and a recall rate of 77.1%. By contrast, DriveNet exhibits competitive performance by attaining a prediction accuracy of 91.24% when anticipating maneuvers four seconds ahead. More recently, STA-Net [33], a novel approach utilizing a spatial-temporal joint attention network, achieved a precision of 90.8% and a recall of 91.1%. While STA-Net slightly surpasses our method in terms of these scores, it achieves these results with a time to maneuver of 0 seconds. In contrast, our method achieves a precision of 90.63% and a recall of 90.32% with a time to maneuver of 4 seconds, providing sufficient time for the driver to react.

5. DISCUSSION AND PERSPECTIVES

In this section, we present some perspectives that we believe will help improve research on the prediction of driving maneuvers. We mainly focus on two challenging issues: data availability and data privacy.

5.1. Dataset

The Brain4Cars dataset used in this work contains two major challenges. First, the data contains only 594 samples. Though this number is largely considered to other publicly available datasets, it is still not enough to train and evaluate models. Second, the maneuver class distributions are unbalanced. To address this challenge, we are looking forward to building a new dataset in the same context by recording the natural long driving distances of different drivers.

5.2. Federated learning

In this work, we used the Brain4Cars dataset, which contains drivers' faces, to anticipate maneuvers. This means that we don't respect users' privacy by taking images of drivers to train the model in order to predict maneuvers. To avoid this problem, federated learning can be used to protect the data generated for each device by sharing model updates, e.g. gradient information, instead of raw data. The predictive model of intelligent vehicle behavior must respond quickly to predict driver behavior in complex real-world situations in order to avoid accidents. Thus, federated learning approach can be used to train machine-learning models for the prediction of driving behaviors.

6. CONCLUSION

In this study, the focus lies in anticipating driving maneuvers a few seconds before they are performed by the driver. The outcome of our research empowers ADAS to forewarn drivers prior to executing risky maneuvers, consequently providing drivers with additional response time. We propose an innovative approach that combines deep learning and the attention mechanism to effectively capture long temporal dependencies

and extract spatial information. DriveNet utilizes the CNN, LSTM, and attention mechanism models, which are specifically designed for sequential data handling and image processing. By employing this configuration, we achieved an impressive accuracy score of 91.24% with a 4-second lead time before the maneuver, allowing ample decision-making time for the driver. Moving forward, our future research endeavors will focus on enhancing the results by exploring alternative models that are more adaptable to this problem. Specifically, we plan to incorporate an optical flow model to analyze motion in videos and a vision transformer capable of interpreting movement over extended periods.




REFERENCES

- [1] M. Peden et al., *World report on road traffic injury prevention*. Geneva, Switzerland: World Health Organization, 2004.
- [2] Department for Transport, "Reported road casualties great britain, annual report: 2022," *Government of the United Kingdom*, 2023.
- [3] S. Singh, "Critical reasons for crashes investigated in the national motor vehicle crash causation survey," *National Highway Traffic Safety Administration*, Washington, DC, 2015.
- [4] S. Lee, W. Choi, C. Kim, M. Choi, and S. Im, "ADAS: A direct adaptation strategy for multi-target domain adaptive semantic segmentation," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 19174–19184, doi: 10.1109/CVPR52688.2022.01860.
- [5] M. A. Al Noman et al., "A computer vision-based lane detection technique using gradient threshold and hue-lightness-saturation value for an autonomous vehicle," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 1, pp. 347–357, 2023, doi: 10.11591/ijece.v13i1.pp347-357.
- [6] A. Al Mamun, P. P. Em, M. J. Hossen, A. Tahabilder, and B. Jahan, "Efficient lane marking detection using deep learning technique with differential and cross-entropy loss," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 4, pp. 4206–4216, 2022, doi: 10.11591/ijece.v12i4.pp4206-4216.
- [7] H. A. Ghani et al., "Advances in lane marking detection algorithms for all-weather conditions," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 4, pp. 3365–3373, 2021, doi: 10.11591/ijece.v11i4.pp3365-3373.
- [8] M. Tonutti, E. Ruffaldi, A. Cattaneo, and C. A. Avizzano, "Robust and subject-independent driving manoeuvre anticipation through domain-adversarial recurrent neural networks," *Robotics and Autonomous Systems*, vol. 115, pp. 162–173, 2019, doi: 10.1016/j.robot.2019.02.007.
- [9] O. Olabiyyi, E. Martinson, V. Chintalapudi, and R. Guo, "Driver action prediction using deep (bidirectional) recurrent neural network," *arXiv-Statistics*, 2017, doi: 10.48550/arXiv.1706.02257.
- [10] A. Jain, H. S. Koppula, B. Raghavan, S. Soh, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3182–3190, doi: 10.1109/ICCV.2015.364.
- [11] M. Bonyani, M. Rahmanian, S. Jahangard, and M. Rezaei, "DIPNet: Driver intention prediction for a safe takeover transition in autonomous vehicles," *IET Intelligent Transport Systems*, vol. 17, no. 9, pp. 1769–1783, 2023, doi: 10.1049/itr2.12370.
- [12] Y. Ma et al., "CEMFormer: Learning to predict driver intentions from in-cabin and external cameras via spatial-temporal transformers," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 4960–4966, 2023, doi: 10.1109/ITSC57777.2023.10421798.
- [13] F. Hasan and H. Huang, "Driver intention and interaction-aware trajectory forecasting via modular multi-task learning," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 1857–1865, 2024, doi: 10.1109/TCE.2023.3321324.
- [14] C. Guo, H. Liu, J. Chen, and H. Ma, "Temporal information fusion network for driving behavior prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 9, pp. 9415–9424, 2023, doi: 10.1109/TITS.2023.3267150.
- [15] Y. Zou, S. Du, H. Han, Y. Liu, and Z. Tian, "Two-Stream (2+1)D CNN based on frame difference attention for driver behavior recognition," in *2023 10th International Conference on Dependable Systems and Their Applications (DSA)*, 2023, pp. 782–788, doi: 10.1109/DSA59317.2023.00110.
- [16] Z. Hou, Y. Fu, S. Wang, D. Liu, H. Liu, and Y. Yang, "Driver-TRN: An approach to driver behavior detection enhanced SOTIF in automated vehicles," *IEEE Vehicular Technology Conference*, 2023, doi: 10.1109/VTC2023-Fall60731.2023.10333638.
- [17] T. Baltrusaitis, P. Robinson, and L. P. Morency, "OpenFace: An open source facial behavior analysis toolkit," *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016, doi: 10.1109/WACV.2016.7477553.
- [18] A. Jain, H. S. Koppula, S. Soh, B. Raghavan, A. Singh, and A. Saxena, "Brain4Cars: Car that knows before you do via sensory-fusion deep learning architecture," *arXiv-Computer Science*, 2016, doi: 10.48550/arXiv.1601.00740.
- [19] Y. Rong, Z. Akata, and E. Kasneci, "Driver intention anticipation based on in-cabin and driving scene monitoring," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–8, doi: 10.1109/ITSC45102.2020.9294181.
- [20] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, 2015, pp. 802–810.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [22] M. Bonyani, M. Rahmanian, and S. Jahangard, "Predicting driver intention using deep neural network," *arXiv-Computer Science*, 2021, doi: 10.48550/arXiv.2105.14790.
- [23] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in *Computer Vision – ECCV 2020*, Cham: Springer, 2020, pp. 402–419, doi: 10.1007/978-3-030-58536-5_24.
- [24] A. Dosovitskiy et al., "FlowNet: Learning optical flow with convolutional networks," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2758–2766, doi: 10.1109/ICCV.2015.316.
- [25] A. Das and M. M. Ahmed, "Machine learning approach for predicting lane-change maneuvers using the shrp2 naturalistic driving study data," *Transportation Research Record*, vol. 2675, no. 9, pp. 574–594, 2021, doi: 10.1177/03611981211003581.




- [26] B. Mersch, T. Hollen, K. Zhao, C. Stachniss, and R. Roscher, "Maneuver-based trajectory prediction for self-driving cars using spatio-temporal convolutional networks," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 4888–4895, doi: 10.1109/IROS51168.2021.9636875.
- [27] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 2118–2125, 2018, doi: 10.1109/ITSC.2018.8569552.
- [28] J. Halkias and J. Colyar, "Next generation simulation fact sheet," *Federal Highway Administration Research and Technology*, 2006. [Online]. Available: <https://www.fhwa.dot.gov/publications/research/operations/its/06135/>
- [29] S. Karen and Z. Andrew, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
- [30] D. V. McGehee, E. N. Mazzae, and G. H. S. Baldwin, "Driver reaction time in crash avoidance research: Validation of a driving simulator study on a test track," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 20, pp. 3-320-3–323, 2000, doi: 10.1177/154193120004402026.
- [31] G. Johansson and K. Rumar, "Drivers' brake reaction times," *Human Factors: The Journal of Human Factors and Ergonomics Society*, vol. 13, no. 1, pp. 23–27, 1971, doi: 10.1177/001872087101300104.
- [32] D. Zhou, H. Ma, and Y. Dong, "Driving maneuvers prediction based on cognition-driven and data-driven method," in *2018 IEEE Visual Communications and Image Processing (VCIP)*, 2018, pp. 1–4, doi: 10.1109/VCIP.2018.8698695.
- [33] B. He, N. Yu, Z. Wang, and X. Chen, "STA-Net: A spatial-temporal joint attention network for driver maneuver recognition, based on in-cabin and driving scene monitoring," *Applied Sciences*, vol. 14, no. 6, 2024, doi: 10.3390/app14062460.

BIOGRAPHIES OF AUTHORS






Mohamed M'haouach    is currently a PhD student at USMBA, Fez, Morocco. He obtained a master's degree in Big Data Analytics and Smart Systems (BDSaS) and a bachelor's degree in computer science and mathematics, from USMBA, in 2017 and 2015, respectively. His research areas are based on computer vision approaches, Business Intelligence, Advanced Driving Assistance Systems (ADAS) and intelligent transportation systems. He can be contacted at email: mohamed.mhaouach@usmba.ac.ma.






Abdellatif Sassioui    is currently a PhD student at Hassan II University of Casablanca, Morocco. He obtained a master's degree in Big Data Analytics and Smart Systems (BDSaS) from Sidi Mohamed Ben Abdellah University (USMBA), Fez, Morocco in 2022 and a bachelor's degree in computer science and mathematics from Mohamed First University (UMP), Oujda, Morocco, in 2020. His research areas are based on computer vision approaches, Advanced Driving Assistance Systems (ADAS), intelligent transportation systems. He can be contacted at email: abdellatif.sassioui-etu@etu.univh2c.ma.



Afaf Bouhoute    holds a is a professor at the Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Morocco. She received a PHD degree in computer science, a master degree in information system, networking and multimedia, and a bachelor degree in computer science, from USMBA, Fez, Morocco, in 2018, 2012 and 2010. DR. Bouhoute had worked as a teaching assistant for 2 years at the National School of Applied Sciences (ENSA) of Fez, Morocco. She also has a 1-year Postdoctoral experience at LaBRI, Bordeaux, France. Her research interests mainly span on modeling and analysis of the driving behavior, using different techniques and algorithms with a focus on their application in intelligent transportation systems. She can be contacted at email: afaf.bouhoute@usmba.ac.ma.



Khalid Fardousse    is a professor at Sidi Mohamed Ben Abdellah University, Fez, Morocco. He obtained a computer science PHD and a master's degree in computer science and decision making, from the Faculty of Science, USMBA, Fez, Morocco, in 2010 and 2002, respectively. Dr. Fardousse' research interests are directed towards the application of computer vision, natural scene pattern recognition and the driving behavior modeling/analysis, in intelligent transportation systems. He can be contacted at email: khalid.fardousse@usmba.ac.ma.