

# Detecting aggressive driving patterns in drivers using vehicle sensor data

Michal Monselise<sup>\*</sup>, Christopher C. Yang

College of Computing and Informatics, Drexel University, Philadelphia, PA, United States

## ARTICLE INFO

### Keywords:

Aggressive driving  
ADHD drivers  
Naturalistic driving study  
Time series  
KNearest neighbors

## ABSTRACT

Aggressive driving is known to be a cause of vehicle accidents. Individuals with Attention-deficit hyperactivity disorder (ADHD) are prone to more aggressive behavior and that also leads to aggressive driving. To prevent aggressive driving, we strive to first understand aggressive driving and find patterns in this type of driving behavior. In an effort to uncover to identify patterns in aggressive driving, we examine sensor data and video data of trips taken by drivers with ADHD and identify our distinct aggressive driving patterns. Using the sensor data, we extend our findings to all aggressive trips in our dataset and generate a model to detect aggressive driving patterns. By finding the similarity between trips and then using these distances to produce a KNN model, we are able to model our data and classify it into 4 driving patterns. This analysis can better inform us of the type of driving patterns that appear in aggressive driving. Using this analysis, we can also better understand which patterns are produce better precision and recall using this methodology.

## Introduction

The World Health Organization reported 1.25 million deaths globally from vehicle accidents in 2013 (World Health Organization, 2015). In a 2015 report, the WHO set road accident prevention as an important goal and highlights a number of causes of road accidents including human error (World Health Organization, 2015). One cause of human error is aggressive driving behavior (Zhao et al., 2019). A number of previous research projects have examined methods for classifying aggressive driving using different data and methods. One category of research uses sensor data and then relies on traditional machine learning techniques like random forests or support vector machines in combination with statistical analysis. Ma et al. conducted a compared a number of machine learning techniques including Gaussian mixture models, support vector regression (SVR), and partial least squares regression (Ma et al., 2018). This research project produced an F1 score of 0.77 at most but showed that each of the algorithms had some ability to predict aggressive trips using sensor data. Manzoni et al. used a model based on variables like acceleration and vehicle position to distinguish between aggressive and non-aggressive driving behavior (Manzoni et al., 2010). Wang et al. produced an algorithm based on support vector machine (SVM) to classify aggressive driving using sensor data (Wang et al., 2017). Abou-Zeid et al. studied aggressive driving patterns using a study including 24 participants and examined their driving using sensor data in a simulator environment (Abou-Zeid et al., 2011). There are also

research projects applying deep learning to sensor data for detection of aggressive driving. Lee et al. studied the classification of driver emotion during aggressive and non-aggressive driving using a convolutional neural network (CNN) (Lee et al., 2018). In a review of aggressive driving detection using deep learning, Alkinani et al. list a number of studies that were able to predict aggressive driving with varying degrees of success but not without some issues (Alkinani et al., 2020). They cited Matousek et al. as producing a deep learning algorithm that relies heavily on the post processing of the recent history of the vehicle to achieve good model performance (Matousek et al., 2019). They also cited Xing et al. who developed a model for estimating energy consumption that includes an analysis of acceleration and deceleration patterns. This model defines aggressive behavior using only acceleration and deceleration patterns which misses aggressive behavior that has other characteristics. These studies focus on classifying aggressive driving using certain criteria to define aggressive driving like acceleration and deceleration. However, they do not identify multiple types of aggressive driving.

Additionally, there is research into classifying specific aggressive driving patterns. Kovaceva et al. examined aggressive driving in situations where the driver is too close to the vehicle in front (Kovaceva et al., 2020). Their study focused mostly on statistical analysis rather than a predictive model. Kim et al. looked at a number of different aggressive behaviors such as driving too close and erratic lane changing (Kim et al., 2018). Their study includes 32 participants and performed a statistical

<sup>\*</sup> Corresponding author.

E-mail address: [michal.monselise@drexel.edu](mailto:michal.monselise@drexel.edu) (M. Monselise).

analysis of the driving data. This study examined data from a driving simulator while our study looks at naturalistic driving. The study had some overlap in the patterns they chose to detect – following too closely, erratic lane changes, and sudden change of speed without changing lanes.

Attention Deficit Hyperactivity Disorder (ADHD) is a condition typically detected during the school years that may affect an individual during their entire life (Swanson et al., 1998). The primary symptoms of ADHD are hyperactivity, impulsivity, and poor sustained attention (Barkley, 1997). Long term effects of the condition include significant risk of low educational achievement, difficulties with relationships, delinquency, and mental illness (Biederman et al., 2006). According to research on ADHD we know that the condition is generally inherited, though some environmental factors may influence it as well (Faraone and Doyle, 2001). The disorder has a prevalence rate of about 2.5% in adults (Simon et al., 2009). Research has shown that ADHD drivers are particularly at risk of accidents from aggressive driving due to the nature of this disorder as causing impulsive behavior and poor sustained attention (Fuermaier et al., 2017). Over the past few decades, multiple studies have shown that individuals with ADHD are indeed more prone to driving aggressively which can lead to more vehicle accidents (Jerome et al., 2006; Richards et al., 2006; Romo et al., 2019; Zamani Sani et al., 2020).

In addition to identifying a relationship between ADHD and road accidents, other researchers have also identified a relationship between ADHD and certain types of risky driving behavior – particularly aggressive driving (Deshmukh and Patel, 2019) which is a known cause of vehicle accidents (Ma et al., 2018). To prevent these behaviors, we must identify them as they occur. By examining the multivariate time series (MTS) that is produced from vehicle sensors, we investigate how to identify this behavior in order to prevent it. A significant challenge that is presented when trying to identify aggressive driving behavior is that the behavior may occur at any time during the trip. This creates a challenge for researchers when studying this MTS data since it is harder to take data that has previously confirmed to contain aggressive behavior and compare our newly observed MTS driving data to find similarities between the two trips. In this research, we aim to characterize the different patterns of aggressive behaviors that ADHD drivers may engage in during a trip. We compare pairs of trips and group together the most similar trips in terms of driving patterns. Using the insight from the pairwise similarity of trips, we develop a classification algorithm that will accurately distinguish between the types of trips and identify the different types of aggressive behavior. This algorithm will then be generalized to classify the behaviors of all aggressive drivers. This classification algorithm will be able to identify the different types of aggressive behavior regardless of when they happen during a trip.

In this research, we examine multivariate time series data to identify patterns in naturalistic driving data collected from vehicle sensors. By comparing trips in a holistic manner, we are able to detect patterns in the data that may not be detected by examining the multivariate time series data alone. Our analysis includes two steps to identify patterns in aggressive driving. In the first step of our analysis, we examine aggressive driving in ADHD drivers to search for patterns that can be defined in an empirical manner rather than relying on human annotation. Since multiple studies have shown that individuals with ADHD are more prone to aggressive behavior, and particularly aggressive driving, we have chosen to examine ADHD drivers in the first step of this analysis. ADHD drivers are shown to be significantly more aggressive and therefore, these patterns are more pronounced in their driving. By applying this methodology to aggressive driving performed by drivers with ADHD, we are able to identify distinct driving patterns that can be observed in the videos of the trips. In the second step of the analysis, we use sensor data to generate a model that can identify distinct behavior patterns that occur throughout a trip. This model can be applied to other aggressive driving behavior in the future and produces labels for naturalistic driving data without the need to analyze video recordings of the trips

algorithmically. This research eliminates the need for subjective coding of the data and instead allows us to utilize sensor data to detect aggressive behavior patterns. Additionally, the patterns that emerge using this methodology can identify behaviors that happen at any point in the trip and allow us to examine whole trips rather than identifying when a behavior occurred in each trip.

## Materials and methods

### The SHRP2 dataset

The Second Strategic Highway Program (SHRP2) is a Naturalistic Driving Study (NDS) that followed approximately 3000 drivers from multiple states between the years 2010–2013. The study has recorded data about approximately 40,000 trips and contains video recordings as well as time series data and metadata for each trip. Additionally, the study included the collection of the drivers' medical history and demographics as well as information each driver's vehicle. The metadata for each trip contains human annotated variables, including variables that describe the driver behavior observed during the trip. There are 59 types of driver behavior labels that have been assigned to each trip. Up to 3 driver behaviors may be assigned to a trip. The most common label assigned is that no risky behavior is observed. However, the remaining 58 labels indicate that risky behavior has been observed. There are two labels that indicate aggressive driver behavior. The labels are "aggressive driving, specific directed menacing actions" and "aggressive driving, other" ("SHRP2 Event Detail Table," 2015). These trips are defined in the SHRP2 data dictionary as trips where the driver displays aggressive behaviors like road rage or reckless behavior like weaving in and out of lanes. The trips in this group have been used in this research. When examining the subgroups of aggressive drivers and ADHD drivers in the SHRP2 dataset, we find that ADHD drivers are overrepresented in the subset of trips where aggressive driving behavior is detected. There are 271 trips where aggressive driving is detected in the SHRP2 dataset. After removing trips containing missing data, we are left with 237 aggressive driving trips. Of the 237 aggressive driving trips, 38 trips are performed by drivers with ADHD. This means that ADHD drivers account for approximately 16% of aggressive driving trips while their overall number of trips account for only 5% of trips in the study. Table 1 shows the breakdown of drivers in the dataset.

### Data processing

The data used in this research is extracted from the time series dataset in SHRP2. Our data contains both raw variables from the SHRP2 dataset as well as derived variables. Fig. 1 demonstrates the axis in the sensor data.

Raw Variables (Variable Notations).

Accel X ( $v_{1,t}$ ) – Vehicle acceleration in the longitudinal direction versus time in units of g.

Accel Y ( $v_{2,t}$ ) – Vehicle acceleration in the lateral direction versus time in units of g.

Accel Z ( $v_{3,t}$ ) – Vehicle acceleration vertically (up or down) versus time in units of g.

Gyro X ( $v_{4,t}$ ) – Vehicle angular velocity around the longitudinal axis in units of degrees per second.

Gyro Y ( $v_{5,t}$ ) – Vehicle angular velocity around the lateral axis in units of degrees per second.

**Table 1**

Distribution of Trips Containing Aggressive Driving and Performed by ADHD Drivers.

|                          | All Drivers | ADHD Drivers |
|--------------------------|-------------|--------------|
| All Trips                | 35,578      | 1787         |
| Aggressive Driving Trips | 237         | 38           |

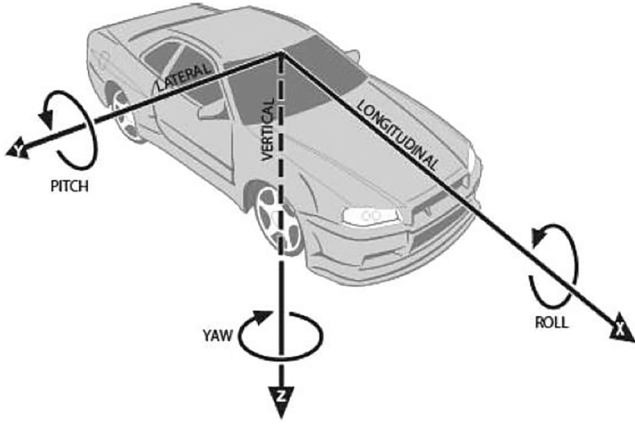


Fig. 1. A Diagram Illustrating the 3 Axes in Recording the Sensor Data.

Gyro Z ( $v_{6,t}$ ) – Vehicle angular velocity around the vertical axis in units of degrees per second.

Derived Variables.

Range X ( $v_{7,t}$ ) - This variable is derived from the range variables which measure the velocity of the nearest 8 vehicles from the subject vehicle. Each of the 8 variables measures the velocity of the vehicles around our vehicle in m/s. However, we only consider whether a vehicle is present in each of the 8 tracks. Using this data, we derive a variable that counts how many vehicles are at a distance of one second away or less from our vehicle in the longitudinal axis. The goal of measuring the number of vehicles less than one second away from our vehicle at its current velocity is to count how many vehicles are too close to our vehicle which helps us identify when sudden breaking happens or when a safe distance is not maintained.

Range Y ( $v_{8,t}$ ) - Likewise, we count the number of vehicles at a distance of one second away or less from our vehicle on the lateral axis.

## Theory and calculation

In the process to identify patterns in the data recorded during trips that contained aggressive driving, we looked at the similarity between trips by comparing entire trips and finding the pairwise similarity of all pairs of trips. This was computed using the Eros similarity score. We then developed a model for predicting the type of risky driving using the Eros similarity score. This model enables us to identify 4 different aggressive driving patterns in ADHD drivers as well as other drivers.

### Eros similarity

The Eros (Extended Frobenius Norm) similarity metric is a metric primarily used for comparing multivariate time series data (Yang and Shahabi, 2004). Intuitively, the Eros similarity metric compares two multivariate time series by measuring the distance between the principal components using the aggregated eigenvalues as weights and considering the variance for each principal component. The research by Yang and Shahabi has shown the Eros similarity measure to be an improvement in performance compared to other techniques like Euclidean distance and dynamic time warping. Therefore, since the Eros similarity metric is a metric that is geared towards MTS, and Yang and Shahabi's research shows that it performs better than multiple other similarity metrics including dynamic time warping, PCA similarity factor and weighted sum SVD, we have selected the metric to compare all trips in our data and group together all similar trips.

The Eros similarity metric is computed using the eigenvalues and eigenvectors of the two multivariate time-series compared. We represent a single trip as a multivariate time series  $V = \{v_{i,t}\} \in \mathbb{R}^{M \times T}$  where we have  $i = 1, \dots, M$  variables and  $t = 1, \dots, T$  time observations. We have  $N$

trips that were taken by the ADHD drivers in our study, each containing a multivariate time series. We select two multivariate time series  $V_k$  and  $V_l$  from our dataset of trips and compute the Eros similarity of the two MTS using the equation:

$$\begin{aligned} Eros(V_k, V_l, \omega) &= \sum_{i=1}^n \omega_i |\langle u_{k,i}, u_{l,i} \rangle| \\ &= \sum_{i=1}^n \omega_i |\cos \theta_i| \end{aligned} \quad (1)$$

where  $u_k = [u_{k,1}, \dots, u_{k,m}]$  and  $u_l = [u_{l,1}, \dots, u_{l,m}]$  are the sets of eigenvectors of  $V_k$  and  $V_l$  respectively and  $\omega$  is the aggregated weight vector generated based on the eigenvalues of the entire MTS dataset (in other words, while the Eros similarity only compares one pair of trips, the weight vector  $\omega$  is computed using all trips). We assert  $\sum_{i=1}^n \omega_i = 1$  and  $\omega_i \geq 0$  for all  $i$ . The computed pairwise similarities of MTS items are used for projecting each MTS item into low-dimensional (2-D or 3-D) space, which will be calculated using multivariate t-SNE (t-distributed Stochastic Neighbor Embedding). The Eros similarity score ranges between 0 and 1 where a value closer to 1 means the two multivariate time series are more similar. While other research has used different similarity metrics to compare time series data, we will be using the Eros similarity metric to compare and group trips. In our case, we would like to compute the Eros similarity between two trip vectors. By comparing each pair of trips, patterns in the data emerge and we can identify similar trips, even when the similar behavior occurs at different points in the trip.

### Multivariate t-SNE

To visualize our data, we opt to reduce the dimensions of the Eros similarity matrix using multivariate t-distributed stochastic neighbor embedding (or m-TSNE). T-SNE is a dimensionality reduction technique developed primarily for the visualization of high dimensional data (van der Maaten and Hinton, 2008). We opt to use a variation of t-SNE called m-TSNE since this technique treats the MTS data as a whole and captures the correlations between the variables in the data (Nguyen et al., 2017). To visualize our Eros similarity matrix that was derived from multivariate time series data, we use m-TSNE which is a technique further developed by Nguyen et al. for visualizing multivariate time series data.

T-SNE reduces the dimensions of our data by computing the Euclidean distance between two points and then converting the distance between the points to a conditional probability distribution. m-TSNE visualizes the data by projecting it into lower dimensional space while preserving the pairwise similarity relation between observation. This means that two similar points in high dimensional space will remain closer or farther apart after the transformation. We compute the probability that two observations will remain similar  $p_{kl}$  in equation (2).

$$p_{kl} = \frac{\exp(-\|V_k - V_l\|^2 / \sigma_k^2)}{\sum_{r \neq k} \exp(-\|V_k - V_r\|^2 / \sigma_k^2)} \quad (2)$$

In equation (2), we describe the probability that point  $v_k$  will be the neighbor of point  $v_l$  if both points are in proportion to their probability density using a Gaussian distribution with variance  $\sigma_k$  (Nguyen et al., 2017).

When reducing the data to a lower dimension, the probability  $q_{kl}$  for our new vectors  $W_k$  and  $W_l$  in the lower dimension space. In this case, our variance will be set to  $\frac{1}{\sqrt{2}}$  and our probability  $q_{kl}$  will be equal to:

$$q_{kl} = \frac{\exp(-\|W_k - W_l\|^2)}{\sum_{r \neq k} \exp(-\|W_k - W_r\|^2)} \quad (3)$$

We find the lower dimension representation of the data where the mismatch between  $q_{kl}$  and  $p_{kl}$  is minimized. To do this, we use the

Kullback-Leibler divergence metric to construct a cost function  $C$ .

$$C = \sum_k KL(P_k || Q_k) = p_{kl} \log \frac{p_{kl}}{q_{kl}} \quad (4)$$

where  $P_k$  and  $Q_k$  represent the conditional probability distributions of the high dimensional data and the mapped lower dimension data, respectively.

Research by Nguyen et al. has shown that the Eros similarity metric is more suitable for visualizing MTS data with t-SNE rather than using other dimensionality reduction techniques like PCA. m-TSNE proposes a better than PCA method to minimize the mismatch between high-dimensional and low-dimensional spaces (Nguyen et al., 2017).

Using m-TSNE in this analysis allows us to uncover relationships between the temporal variables and identify risky behaviors in trips regardless of when the behavior happens in each trip.

#### KNN model based on Eros similarity

The K-Nearest Neighbors algorithm (or KNN) is a classification or regression algorithm that determines the label of a point based on the observations closest to it (or its neighbors). We determine proximity of points using a chosen distance metric. In the case of using KNN for classification, we use the K nearest points and apply the most frequent label in those points to the observation we want to label.

In our case, we use the Eros distance as a distance metric. For features  $V$  and labels  $Y = 1, \dots, 4$ , given the distances computed between all trips, we find an optimal  $K = 5$  and assign the label with the largest probability. Given our multivariate time series data  $V = \{v_{i,t}\} \in \mathbb{R}^{M \times T}$ , the probability of a trip  $l$  having label  $j$  is:

$$P(Y = j | V = V_l) = \frac{1}{K} \sum_{i \in A} I(y^{(i)} = j) \quad (5)$$

where  $A$  is the set of  $K$  nearest observations.

We utilize algorithm 1 in our predictive model to assign a label to each trip.

#### Algorithm 1 KNN Model Based on Eros Similarity

---

**Input:**  $n \times n$  distance matrix of Eros similarity between trips  
 $n \times 1$  vector of labels  $y$   
 $k$  – number of neighbors used to apply label  
**Output:** Vector of predicted labels for test data  
 Split the data to train and test samples  
 For each observation  $j$  in test:  
      $d_j = d_{j,1}, \dots, d_{j,k}$  # The  $k$  closest observations in train to  $j$  using the Eros similarity  
      $c_j = c_{j,1}, \dots, c_{j,4}$  # Vector containing counts of the number of closest observations for each label  
      $\hat{y}_{unweighted} \leftarrow \text{argmax}(c)$   
      $a \leftarrow \text{accuracy}(y, \hat{y}_{unweighted})$   
 Generate the optimization vector  $r$   
 For each observation  $j$  in test perform an exhaustive search for hyperparameters for labels 1-4 in the space  $[0.1, 4]$  by optimizing for accuracy  
      $\hat{y} \leftarrow \text{max}(c \bullet r)$   
 Return  $\hat{y}$

---

## Results

In this research, we utilize the Eros similarity metric for analyzing the MTS data and create a model that can identify four patterns of aggressive driving patterns. We compute the Eros similarity between all pairs of trips to get a holistic view of the trips and how they compare to each other which enables us to detect patterns throughout the entire trip. Our analysis will use the MTS data to extract insight about the data where we look for patterns that happen at any point in the trips. Therefore, we do not compare trips second by second. Instead, we look at larger trends in the MTS data.

Our initial goal is to classify all ADHD aggressive trips. We aim to perform this analysis by visualizing all aggressive trips using m-TSNE

and particularly highlighting the ADHD driver trips. Since t-SNE is a technique meant for visualization and not for inference (van der Maaten and Hinton, 2008), we use the visualization as a guideline to identify the clusters of aggressive driving.

We start by plotting the m-TSNE visualization of the ADHD aggressive trips. By looking at the plot in Fig. 2, we observe a separation between 4 groups. However, as seen in Fig. 2, the groups on the bottom and right are clearly separated while the top and left are separated but without a clear boundary. When we examine the two groups on the left side of the graph, we cannot determine where the boundary between the two groups lies. Since t-SNE plots act as a visual guide to detect separation between groups, we use this information and then manually label the data.

After manually labeling the data, we see that the groups in the right and bottom have distinct behavior and the group on the left can be split into two distinct behaviors as well. This is shown in Fig. 3.

When plotting just the ADHD trips, we see separation between 4 groups in the plot for most points. T-SNE plots serve as a visual guide for detecting separation between groups in the data (van der Maaten and Hinton, 2008), therefore, we use the visually separate groups as a guide for labeling the data.

Driving patterns were identified by watching the videos. The videos were viewed 3 times to ensure correct classification. The videos showed that distinct driver patterns emerged from the data that were identified as 4 distinct groups. These groups were assumed to produce different time series patterns in the 8 different variables that have been used in this analysis. There are a few observations that did not fit perfectly into either of the groups in the t-SNE visualization; however, after labeling the data, we were able to apply an appropriate label to these observations that matched the content of the videos. These labels are shown in Fig. 3.

The patterns are further observed in Fig. 4 after reintegrating all aggressive driving trips into the visualization. We can better observe the separation into 4 groups. We notice that two observations do not exactly fit into a cluster, but these were labeled as well and found to have a consistent pattern with their adjacent cluster.

We can also confirm that the points group together around a centroid by generating 4 clusters of points from our Eros data using k-means and plotting the clusters in Fig. 5. While we get 4 distinct clusters centered similarly to the manual labels, some points in the boundaries are associated with different clusters in this case.

When examining the labeled points, we see 4 distinct driving patterns occurring during the trips.

Pattern 1 – stop and go driving. This pattern occurs when the aggressive driver does not maintain a safe distance with the vehicle in front. This causes the aggressive driver to hard press the brakes as they approach the vehicle in front. Table 2 contains the mean of the absolute value of the raw variables in our dataset for each driving pattern. We observe that due to the constant acceleration and deceleration in this driving pattern, it has the highest mean value for the absolute value of the X acceleration and the Y acceleration.

Fig. 6 demonstrates this behavior with screenshots from a video of a trip containing this pattern. The screenshots are extracted from the trip at intervals of one second between shots. In screenshot 1 we see a white vehicle in front of the subject vehicle. By screenshot 11, the vehicle has moved farther away and by screenshot 20 we see the vehicle in front to be closely in front of the subject vehicle. This is what typical stop and go driving looks like in the videos recorded of the trips.

Pattern 2 – Abruptly changing lanes. This driving pattern mainly occurs when an aggressive driver suddenly changes lane due to not paying attention to the road conditions. This behavior could happen when there is a slower vehicle ahead or another driver legally attempts to change lane and the aggressive driver only notices at the last moment. This causes a sudden change in the driving angle due to a sharp swerve. In Fig. 7, we see screenshots from a trip containing this behavior pattern that were taken from a video representative of pattern 2. The

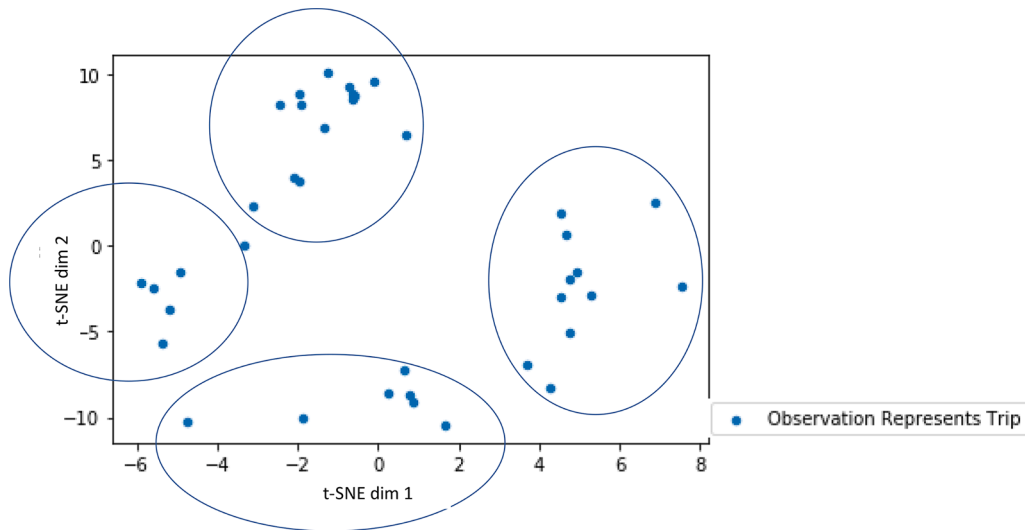


Fig. 2. Two-dimensional m-TSNE plot of the Eros Similarity Between all Pairs of Trips in the Dataset.

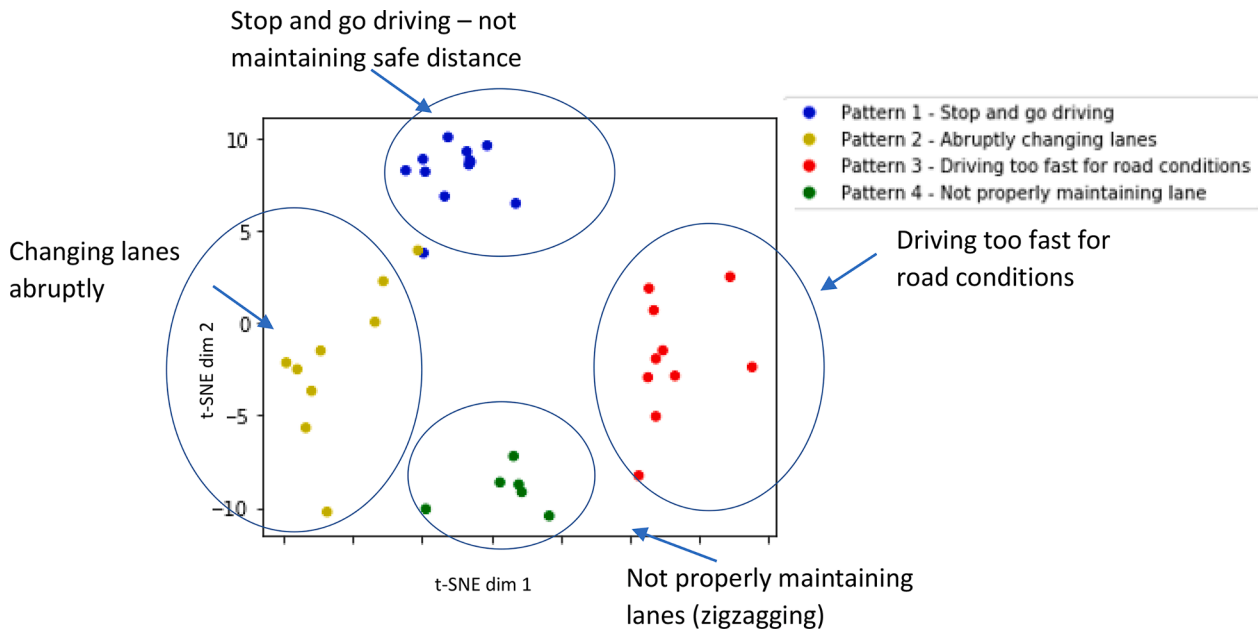


Fig. 3. Two-Dimensional m-TSNE Plot of the Eros Similarity of ADHD trip pairs.

screenshots were taken at a 1 s interval. In screenshot 1, we see a white vehicle in front of the subject vehicle. By screenshot 8, the vehicle becomes closer and closer to the subject vehicle. The subject vehicle changes lanes so abruptly that by screenshot 9, the white vehicle is to the left of the subject vehicle. We continue to see the white vehicle to the left of the subject vehicle till the end of the clip.

**Pattern 3 – Driving too fast for road conditions.** This pattern is mostly observed in weather conditions that should limit a vehicle's speed. Drivers do not necessarily make any sudden maneuvers like in the previous driving pattern. Instead, they engage in aggressive driving by driving too fast in conditions like rain or fog. There are also drivers who engage in aggressive behavior in this driving pattern who are driving too fast in a lane that seems to be momentarily free flowing even though there is upcoming traffic congestion or in a parking lot where there may be pedestrians.

**Pattern 4 – Not properly maintaining lane or zigzagging.** This driving pattern is characterized by aggressive driving behavior where the driver changes lane multiple times to avoid slowing down. This means that

these drivers engaged in constant angle change throughout the trip due to their zigzagging. Zigzagging has been documented in the literature to be observed when there is a change angular velocity in the y axis (Jeong et al., 2013) as is observed in this research as well. In fact, in each trip, at least 8 changes from positive to negative in this group. This is the highest number of changes in angular direction of all patterns. Additionally, this pattern produced the highest mean absolute value of change in angle in the y axis, as can be observed in table 2.

Most trips contained only one aggressive driving pattern; however, in the case of more than one pattern appearing in a single trip, we classified the trip by selecting the driving behavior that appeared the longest in the trip. We defined this behavior to be the dominant pattern. Typically, the non-dominant pattern was much shorter in comparison. For example, if a driver was displaying stop and go driving throughout the trip but also zigzagged for a small portion of the trip, we would determine that the trip contained stop and go driving for the purpose of assigning a label. After labeling our data with the different driving patterns, we then expand our analysis to all aggressive trips in the

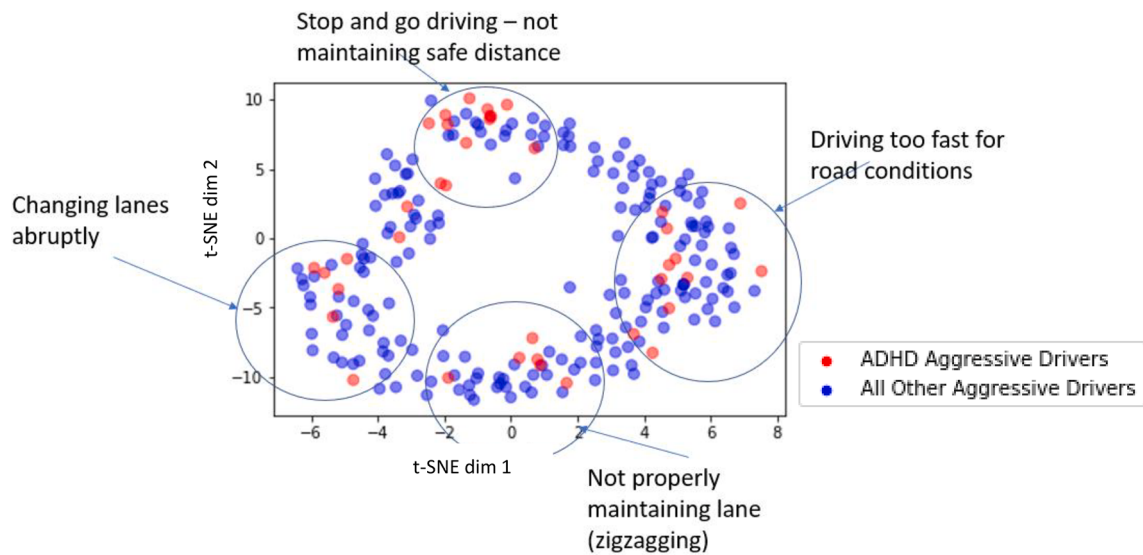


Fig. 4. Two-dimensional m-TSNE plot of the Eros Similarity Between all Pairs of Trips in the Dataset.

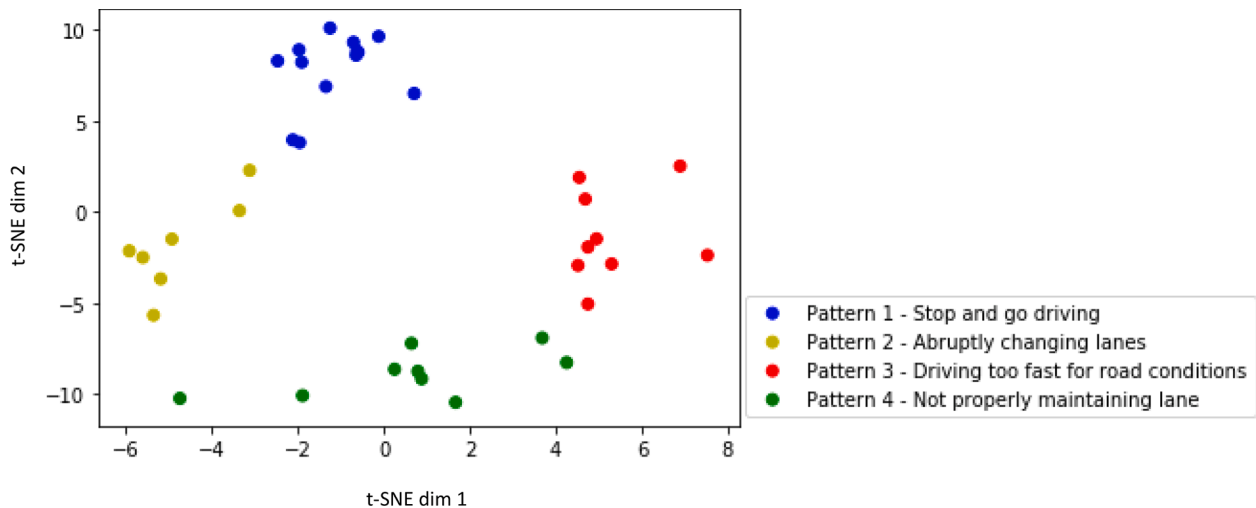


Fig. 5. Two-Dimensional m-TSNE Plot of the Eros Similarity of ADHD trip pairs grouped using K-Means.

Table 2

Mean of the Absolute Value of Raw Variables for Each Driving Pattern.

| Driving Pattern         | X Acceleration (g) | Y Acceleration (g) | Z Acceleration (g) | X Angular Velocity (degrees/second) | Y Angular Velocity (degrees/second) | Z Angular Velocity (degrees/second) |
|-------------------------|--------------------|--------------------|--------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Stop and Go Driving     | 6.093              | 64.788             | 0.315              | 12.233                              | 1.722                               | 1.571                               |
| Changing Lanes Abruptly | 4.665              | 44.569             | 0.480              | 11.326                              | 1.555                               | 2.902                               |
| Driving too fast        | 4.423              | 47.248             | 0.504              | 1.167                               | 10.488                              | 15.056                              |
| Zigzagging              | 4.287              | 49.025             | 0.509              | 3.569                               | 15.313                              | 2.617                               |

dataset. The primary reason for this is because the sample of ADHD aggressive trips is very small and creating a predictive model using less than 40 trips would not produce any meaningful insight. The study performed by Lambert for NHTSA compared drivers with ADHD to drivers who did not have ADHD (Lambert and University of California, 1995) and found that there were no significant difference in many areas related to driving such as the age of obtaining a license for men as well as the prevalence of driving under the influence between the two groups. The main difference between drivers with and without ADHD is in the prevalence of aggressive driving behavior in drivers with ADHD which led to more moving violations and an increased risk of accidents.

Therefore, we extend our initial result to all aggressive trips using the ADHD driver data. The trips were labeled as aggressive by human annotators. These annotators examined the trip data which includes video of the front and rear windows, the facial expressions of the driver, and the sensor data. Based on these data sources they determined whether the driving was aggressive. However, the data dictionary does not mention the criteria for judging whether the driving is considered aggressive. By identifying patterns in the sensor data that correspond to the aggressive driving label, we are able to extend the finding to automatically assign the aggressive driving label based on a predictive model.



Fig. 6. Screenshots from a Trip Containing Stop and Go Behavior.

To find a pattern in our data, we must first look at the similarity between the trips. As previously mentioned, the Eros distance has been shown to produce better results as a similarity measure for multivariate time series data. Therefore, we will look at the Eros distance between the MTS data for all trips. Using a combination of Eros and t-SNE for visualizing MTS data has been outlined in the research by Wong and Chung (Wong and Chung, 2019) and will be employed in this research. The data in this study has been reduced to all aggressive trips. After

eliminating observations with missing videos, we are left with 148 trips. We compute the Eros distance between each pair of trips and then visualize the data using m-TSNE. All trips were viewed 3 times and one of the four labels was applied to the trip. The trips were then separated into test and training data. To classify our data to one of the four patterns, we create a case-based reasoning model where we use a KNN model of the Eros distance to classify the trips.

Our initial analysis revealed that there are 4 distinct labels when

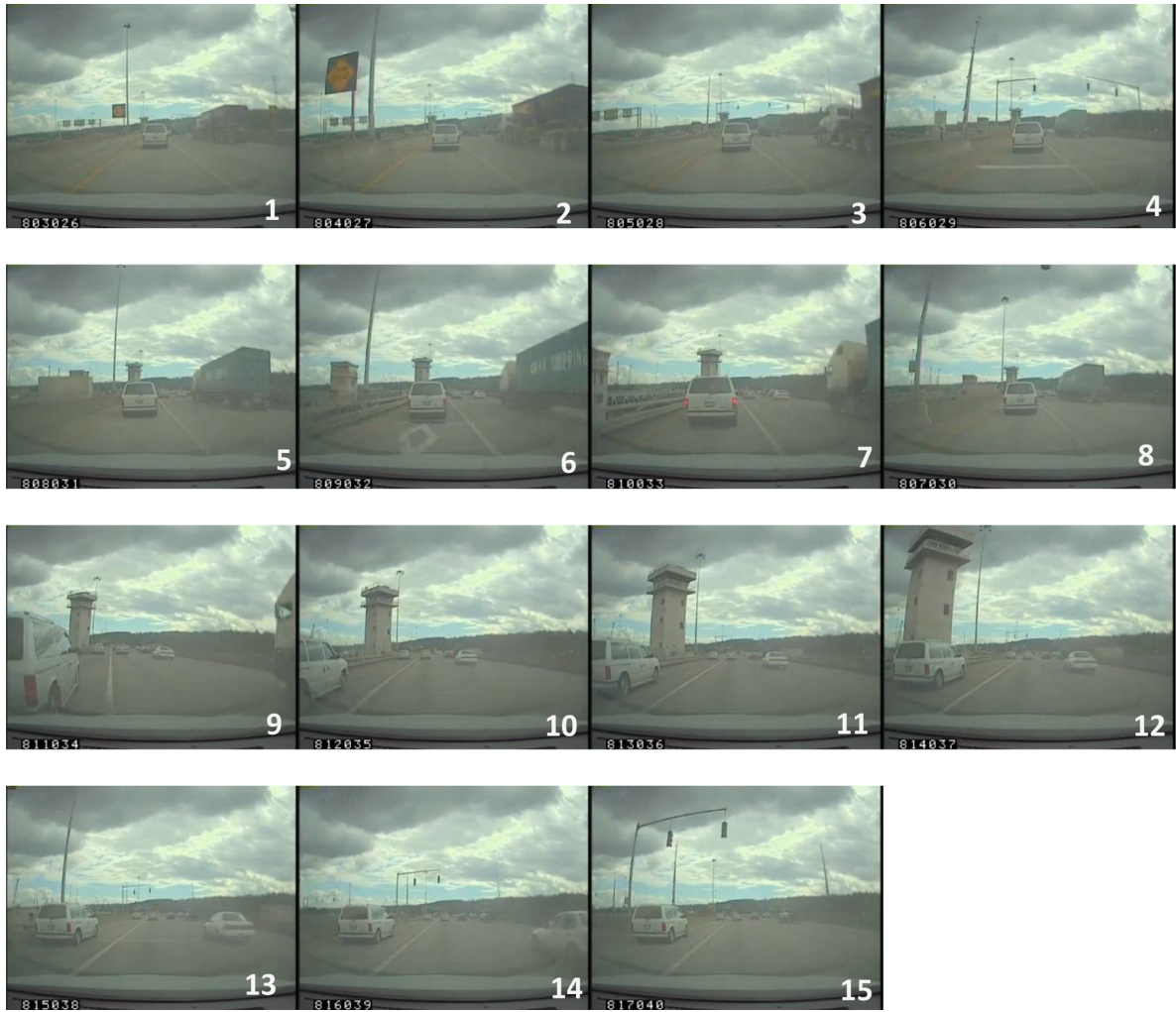


Fig. 7. Screenshots from a Trip Containing an Abrupt Change in Lane.

examining the Eros similarity between the trips. Therefore, our prediction model will be a KNN model that is based on the Eros similarity between the trips. We start by computing the pairwise Eros similarity of all pairs of trips in the data. This produces a  $n \times n$  distance matrix. We select  $k$  based on the accuracy score and find that  $k = 5$  produces the highest accuracy score. Our algorithm then identifies the 5 closest trips in the training data to each trip in the test data. It computes the count of trips with each label and assigns that label to the trip in the test data. Additionally, we add a weight to each label. We optimize the weights that the counts are adjusted by. We end up with a weight of [1,1.5,0.5,1.5] (in other words, the count for label 1 is multiplied by 1, the count for label 2 is multiplied by 1.5, and the count of labels 3 and 4 are multiplied by 0.5 and 1.5 respectively). We compute the accuracy using the count of correct predictions divided by all predictions as described in Table 3. Our model is described in algorithm 1 and results in a test accuracy of 81% and precision and recall described in Tables 4 and 5. We visualize the labels applied by the KNN algorithm in Fig. 8. Even though we used the Eros distances without applying the m-TSNE transformation in the KNN algorithm, when applying the labels to the

Table 3  
Confusion Matrix.

| Predicted | Observed       |                |
|-----------|----------------|----------------|
|           | True Positive  | False Positive |
|           | False Negative | True Negative  |

Table 4  
Total Model Performance.

|             | Accuracy | Precision | Recall | F1      |
|-------------|----------|-----------|--------|---------|
| Total Model | 0.8103   | 0.8448    | 0.8103 | 0.82719 |

Table 5  
Model Performance by Driving Pattern.

|                                                  | Precision | Recall | F1       |
|--------------------------------------------------|-----------|--------|----------|
| Stop and go driving (pattern 1)                  | 1         | 0.3636 | 0.533333 |
| Abruptly changing lanes (pattern 2)              | 1         | 0.8571 | 0.84375  |
| Driving too fast for road conditions (pattern 3) | 0.75      | 0.9643 | 0.923077 |
| Not properly maintaining lanes (pattern 4)       | 0.8333    | 0.8333 | 0.833333 |

observations in the visualizations, we still get a clear separation of the 4 patterns. However, we do observe that the boundary between the patterns has slightly moved.

We compute the accuracy, precision, recall, and F1 score for the entire model as well as for each driving pattern. Recall that when using a prediction model, we get 4 types of predictions shown in the confusion matrix in table 3.

The accuracy is computed by dividing the number of correctly classified observations by the total number of observations. We compute precision, recall, and the F1 statistic using the following equations:

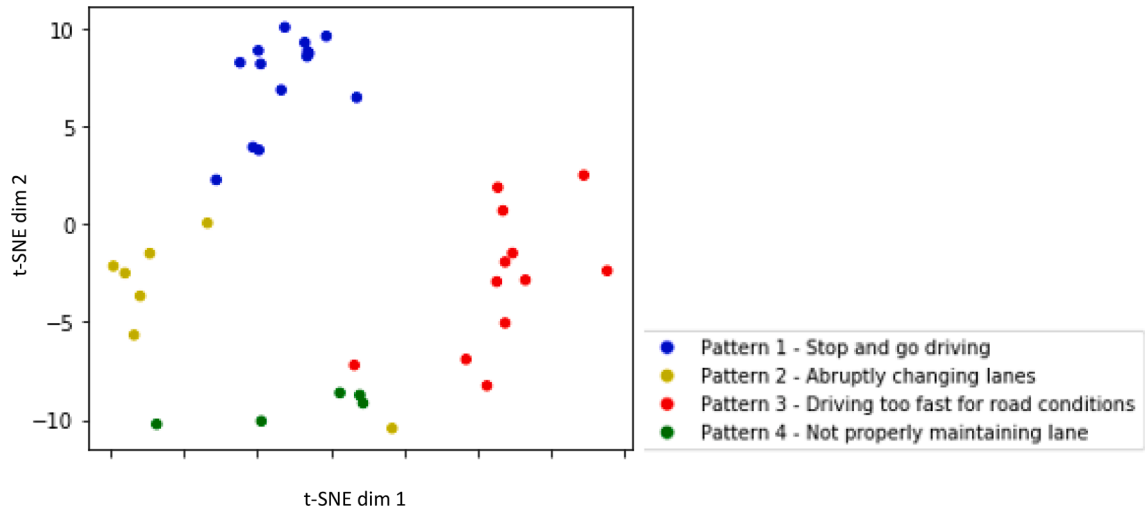


Fig. 8. Two-Dimensional m-TSNE Plot of the Eros Similarity of ADHD trip pairs grouped using the KNN algorithm.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (6)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (7)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

We compute the total accuracy and find that it is 81.03%. While the accuracy we achieve is satisfactory, the model performs better for some patterns than others. When we look at the model performance across all patterns, we observe that pattern 1 had the lowest recall. With a recall of less than 50%, we note that we have more false negatives than true positives for pattern 1. Pattern 3 had the best recall but the lowest precision. This means that pattern 3 had the greatest number of false positives but the smallest number of false negatives.

## Discussion

### Identifying patterns from the general population of trips

Some patterns of aggressive driving could be identified by examining the pool of all aggressive trips since they are distinct patterns containing sharp angle changes or extreme changes back and forth in the number of cars around our vehicle. On the other hand, other behaviors would require additional data to identify. Pattern 3 contains all trips that are classified as too driving fast for road conditions. When viewing all aggressive trips that are classified as too fast for road conditions, these trips occur in road conditions like rain, fog, snow, or in heavy traffic. There are few or no risky maneuvers performed by the driver like sharp swerves or zigzagging between lanes. What makes this driving pattern risky is the driver's failure to adapt their driving to the road conditions. Since these trips do not differ in their behavior from trips that were not classified as aggressive, to extend our algorithm to the general population, we would have to add additional data to our model. The SHRP2 dataset currently contains information about external factors like time of day and weather conditions. To extend this work to detecting these patterns of risky behaviors, we would need to include this additional information in our model.

### Difference in classification performance

While we have an 81% accuracy for the model overall, we observe that for pattern 1 – stop and go driving, recall was 36%. We find that these trips are not easily identified by comparing them one to another.

Stop and go trips are most incorrectly classified as driving too fast for road conditions. This means that when tailing the car in front, our subject vehicle might be pressing the brake as much as in the driving too fast for road conditions category. In this case, we may differentiate between the categories using the distance kept between the vehicles in front and behind our subject vehicle. Furthermore, we observe that there are bursts of activity when the driver presses the gas and the brakes over and over. Using these bursts, we can better identify stop and go behavior. Another issue that we notice is that in pattern 3, we observe the lowest precision of all patterns since this pattern contains trips where aggressive driving is characterized by driving too fast for road conditions. This pattern has multiple false positives since these trips are characterized by driving recklessly during poor conditions rather than a specific pattern like swerving or stopping and going.

### ADHD drivers and aggressive driving patterns

Prior research has shown that among aggressive drivers, there is a higher prevalence of ADHD (Malta et al., 2005). Other research suggests this may be caused by the fact that individuals with ADHD are more likely to exhibit anger and aggression in general and not just in driving (Ramirez et al., 1997; Whalen et al., 2002). We also observe this in the SHRP2 dataset where there is an overrepresentation of ADHD drivers in all aggressive trips. Aggressive trips performed by ADHD drivers account for 16% of all aggressive trips while ADHD drivers represent only 5% of all trips in the study. While we extended the model from ADHD aggressive drivers to all aggressive drivers, the model classifies ADHD aggressive trips with an even higher accuracy than the general population of aggressive trips, which indicates that the model is well suited for them since it was initially developed using ADHD aggressive driving data only. When filtering the test data to only ADHD trips, the accuracy score is 0.857. Other studies suggest that ADHD drivers are more prone to risky driving behaviors like sudden lane changing and driving too close to the vehicle in front of them (Reimer et al., 2005).

### Limitations

In our research, we identify patterns in aggressive driving. We use the sensor data to detect patterns that can be used to classify trips that were annotated by human annotators. However, in this research, we have not used nonaggressive trips in our dataset. We found that nonaggressive trips may sometimes contain similar driving patterns (such as speeding and slowing down frequently). Therefore, the inclusion of nonaggressive trips in the dataset may cause us to detect the existence of one of the four patterns in these trips as well.

Furthermore, due to the low prevalence of aggressive driving trips in our dataset, the sample size is small. This means that a few misclassified observations can sway the accuracy of the model greatly. Since the data collection for this project ended in 2013, new data cannot be added to the sample.

## Conclusions

In this research, we examined the aggressive driving behavior of ADHD drivers as well as the general population of aggressive drivers. The SHRP2 data has been labeled by human annotators who described many things about each trip including a field that noted whether a trip was aggressive (Kamrani et al., 2019). Our goal was to detect distinct aggressive driving patterns in both ADHD drivers and the general population of drivers. By doing this, we will be able to circumvent the process of manual labeling and instead detect aggressive driving patterns using the sensor data. We found that using this model, we were able to take the subjective label of aggressive driving that was assigned to the trips by human annotators and detect what is specifically meant by aggressive driving. We observed 4 driving patterns: stop and go driving, abruptly changing lanes, driving too fast for road conditions, and not properly maintaining lane (or zigzagging). By identifying 4 different types of aggressive driving, we were able to find what characterizes these driving behaviors and quantify this information using the sensor data to produce a predictive model. Our predictive model compared all pairs of trips using the Eros distance (Yang and Shahabi, 2004) and then used these distances in a KNN model. We found that some patterns had better precision and recall than others since detecting some patterns required more reliance on the actual information in the sensor data rather than a comparison between the trips. These conclusions may assist in driver assistance vehicle systems. Since this work relies on vehicle velocity and acceleration as well as radar sensors to detect adjacent vehicles, this work can present a lower cost alternative research that proposes using cameras to detect driving behaviors such as the research by Wang et al. (Wang et al., 2013). Our future work will focus on creating a model that will depend on the sensor data and its unique features. For example, stop and go driving had low recall with this model. We plan on detecting this driving pattern more successfully using a model that can better observe the changes in acceleration in the x and y dimensions of the subject vehicle. Similarly, we may integrate environmental features into our model like weather conditions and time of day which may improve the prediction of driving too fast for road conditions.

## CRedit authorship contribution statement

**Michal Monselise:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Christopher Yang:** Resources, Writing - review & editing, Supervision, 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.

## Acknowledgements

This work was supported in part by the National Science Foundation under the Grant NSF-1741306, IIS-1650531, and DIBBS-1443019. Any opinions, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## References

- Abou-Zeid, M., Kaysi, I., Al-Naghi, H., 2011. Measuring Aggressive Driving Behavior Using a Driving Simulator: An Exploratory Study. Presented at the 3rd International Conference on Road Safety and Simulation. Purdue University Transportation Research Board.
- Alkinani, M.H., Khan, W.Z., Arshad, Q., 2020. Detecting human driver inattentive and aggressive driving behavior using deep learning: recent advances, requirements and open challenges. *IEEE Access* 8, 105008–105030. <https://doi.org/10.1109/ACCESS.2020.2999829>.
- Barkley, R.A., 1997. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol. Bull.* 121, 65–94. <https://doi.org/10.1037/0033-2909.121.1.65>.
- Biederman, J., Monuteaux, M.C., Mick, E., Spencer, T., Wilens, T.E., Silva, J.M., Snyder, L.E., Faraone, S.V., 2006. Young adult outcome of attention deficit hyperactivity disorder: A controlled 10-year follow-up study. *Psychol. Med.* 36, 167–179. <https://doi.org/10.1017/S0033291705006410>.
- Deshmukh, P., Patel, D., 2019. Driving and Road Rage Associated with Attention Deficit Hyperactivity Disorder (ADHD): a Systematic Review. *Curr. Dev. Disord. Rep.* 6, 241–247. <https://doi.org/10.1007/s12205-013-0387-9>.
- Faraone, S.V., Doyle, A.E., 2001. The nature and heritability of attention-deficit/hyperactivity disorder. *Child. Adolesc. Psychiatr. Clin. N. Am.* 10, 299–316 viii–ix.
- Fuermaier, A.B.M., Tucha, L., Evans, B.L., Koerts, J., de Waard, D., Brookhuis, K., Aschenbrenner, S., Thome, J., Lange, K.W., Tucha, O., 2017. Driving and attention deficit hyperactivity disorder. *J. Neural Transm.* 124, 55–67. <https://doi.org/10.1007/s00702-015-1465-6>.
- Jeong, E., Oh, C., Kim, I., 2013. Detection of lateral hazardous driving events using in-vehicle gyro sensor data. *KSCE J. Civ. Eng.* 17, 1471–1479. <https://doi.org/10.1007/s12205-013-0387-9>.
- Jerome, L., Segal, A., Habinski, L., 2006. What We Know About ADHD and Driving Risk: A Literature Review, Meta-Analysis and Critique. *J. Can. Acad. Child. Adolesc. Psychiatry* 15, 105–125.
- Kamrani, M., Arvin, R., Khattak, A.J., 2019. The Role of Aggressive Driving and Speeding in Road Safety: Insights from SHRP2 Naturalistic Driving Study Data. Presented at the Transportation Research Board 98th Annual Meeting. Transportation Research Board.
- Kim, K., Hong, S., Choe, B., Oh, C., 2018. The methodology for identifying aggressive driving using driving behavior patterns. Presented at the 18th International Conference Road Safety on Five Continents (RS5C 2018), Jeju Island, South Korea, May 16–18, 2018, Statens väg- och transportforskningsinstitut.
- Kovaceva, J., Isaksson-Hellman, I., Murgovski, N., 2020. Identification of aggressive driving from naturalistic data in car-following situations. *J. Saf. Res.* 73, 225–234. <https://doi.org/10.1016/j.jsr.2020.03.003>.
- Lambert, N.M., University of California, Berkeley, S. of E., 1995. Analysis of driving histories of ADHD subjects (No. DOT-HS-808-417).
- Lee, K.W., Yoon, H.S., Song, J.M., Park, K.R., 2018. Convolutional Neural Network-Based Classification of Driver's Emotion during Aggressive and Smooth Driving Using Multi-Modal Camera Sensors. *Sensors* 18, 957. <https://doi.org/10.3390/s18040957>.
- Ma, C., Hao, W., Xiang, W., Yan, W., 2018. The Impact of Aggressive Driving Behavior on Driver-Injury Severity at Highway-Rail Grade Crossings Accidents [WWW Document]. *J. Adv. Transp.* <https://doi.org/10.1155/2018/9841498>.
- Malta, L.S., Blanchard, E.B., Freidenberg, B.M., 2005. Psychiatric and behavioral problems in aggressive drivers. *Behav. Res. Ther.* 43, 1467–1484. <https://doi.org/10.1016/j.brat.2004.11.004>.
- Manzoni, V., Corti, A., De Luca, P., Savaresi, S.M., 2010. Driving style estimation via inertial measurements. In: 13th International IEEE Conference on Intelligent Transportation Systems. Presented at the 13th International IEEE Conference on Intelligent Transportation Systems, pp. 777–782. <https://doi.org/10.1109/ITSC.2010.5625113>.
- Matousek, M., EL-Zohairy, M., Al-Momani, A., Kargl, F., Bösch, C., 2019. Detecting Anomalous Driving Behavior using Neural Networks. In: 2019 IEEE Intelligent Vehicles Symposium (IV). Presented at the 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 2229–2235. <https://doi.org/10.1109/IVS.2019.8814246>.
- Nguyen, M., Purushotham, S., To, H., Shahabi, C., 2017. m-TSNE: A Framework for Visualizing High-Dimensional Multivariate Time Series.
- Ramirez, C.A., Rosén, L.A., Deffenbacher, J.L., Hurst, H., Nicoletta, C., Rosencranz, T., Smith, K., 1997. Anger and anger expression in adults with high ADHD symptoms. *J. Attention Disord.* 2, 115–128. <https://doi.org/10.1177/108705479700200205>.
- Reimer, B., D'Ambrosio, L.A., Gilbert, J., Coughlin, J.F., Biederman, J., Surman, C., Fried, R., Aleardi, M., 2005. Behavior differences in drivers with attention deficit hyperactivity disorder: The driving behavior questionnaire. *Accid. Anal. Prev.* 37, 996–1004. <https://doi.org/10.1016/j.aap.2005.05.002>.
- Richards, T.L., Deffenbacher, J.L., Rosén, L.A., Barkley, R.A., Rodricks, T., 2006. Driving Anger and Driving Behavior in Adults With ADHD. *J. Atten. Disord.* 10, 54–64. <https://doi.org/10.1177/1087054705284244>.
- Romo, L., Sweerts, S.J., Ordonneau, P., Blot, E., Gicquel, L., 2019. Road accidents in young adults with ADHD: Which factors can explain the occurrence of injuries in drivers with ADHD and how to prevent it? *Appl. Neuropsychol.: Adult* 1–6. <https://doi.org/10.1080/23279095.2019.1640697>.
- SHRP2 Event Detail Table [WWW Document], 2015. URL <https://insight.shrp2ndns.us/info/printable/38?type=dataset> (accessed 1.5.21).
- Simon, V., Czobor, P., Bálint, S., Mészáros, A., Bitter, I., 2009. Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis. *Br. J. Psychiatry* 194, 204–211. <https://doi.org/10.1192/bjp.bp.107.048827>.

- Swanson, J., Sergeant, J., Taylor, E., Sonuga-Barke, E., Jensen, P., Cantwell, D., 1998. Attention-deficit hyperactivity disorder and hyperkinetic disorder. *Lancet* 351, 429–433. [https://doi.org/10.1016/S0140-6736\(97\)11450-7](https://doi.org/10.1016/S0140-6736(97)11450-7).
- van der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Wang, W., Xi, J., Chong, A., Li, L., 2017. Driving style classification using a semisupervised support vector machine. *IEEE Trans. Hum.-Mach. Syst.* 47, 650–660. <https://doi.org/10.1109/THMS.2017.2736948>.
- Wang, J., Zhang, L., Zhang, D., Li, K., 2013. An adaptive longitudinal driving assistance system based on driver characteristics. *IEEE Trans. Intell. Transp. Syst.* 14, 1–12. <https://doi.org/10.1109/TITS.2012.2205143>.
- Whalen, C.K., Jamner, L.D., Henker, B., Delfino, R.J., Lozano, J.M., 2002. The ADHD spectrum and everyday life: experience sampling of adolescent moods, activities, smoking, and drinking. *Child Dev.* 73, 209–227. <https://doi.org/10.1111/1467-8624.00401>.
- Wong, K.Y., Chung, F., 2019. Visualizing Time Series Data with Temporal Matching Based t-SNE. In: 2019 International Joint Conference on Neural Networks (IJCNN). Presented at the 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. <https://doi.org/10.1109/IJCNN.2019.8851847>.
- World Health Organization, 2015. Global Status Report on Road Safety. World Health Organization.
- Yang, K., Shahabi, C., 2004. A PCA-based similarity measure for multivariate time series. In: Proceedings of the 2nd ACM International Workshop on Multimedia Databases - MMDDB '04. Presented at the the 2nd ACM international workshop. ACM Press, Washington, DC, USA. <https://doi.org/10.1145/1032604.1032616>.
- Zamani Sani, S.H., Fathirezaie, Z., Sadeghi-Bazargani, H., Badicu, G., Ebrahimi, S., Grosz, R.W., Sadeghi Bahmani, D., Brand, S., 2020. Driving Accidents, Driving Violations, Symptoms of Attention-Deficit-Hyperactivity (ADHD) and Attentional Network Tasks. *Int. J. Environ. Res. Public Health* 17. <https://doi.org/10.3390/ijerph17145238>.
- Zhao, X., Xu, W., Ma, J., Li, H., Chen, Y., 2019. An analysis of the relationship between driver characteristics and driving safety using structural equation models. *Transp. Res. Part F: Traff. Psychol. Behav.* 62, 529–545. <https://doi.org/10.1016/j.trf.2019.02.004>.