Features that Distinguish Drivers: Big Data Analytics of Naturalistic Driving Data

Bruce Wallace1,2, Frank Knoefel1,2,3, Rafik Goubran1,2, Michelle M. Porter5,6, Andrew Smith4, Shawn Marshall3,4
1Systems and Computer Engineering, Carleton University, Ottawa, Canada
2Bruyere Research Institute
3Faculty of Medicine, University of Ottawa
4Ottawa Hospital Research Institute, University of Ottawa, Ottawa, Canada
5Faculty of Kinesiology and Recreation Management, and Centre on Aging
6University of Manitoba, Winnipeg, Canada
{wally, goubra, fnknoefel}@sce.carleton.ca

Abstract

The unique behaviours of drivers have many emerging applications. These include the personalization of automated/self-driving vehicles so the owners are more comfortable with them, and the identification of changing driving behaviours that may be associated with aging or disease. This paper explores measures of driving behaviours that might allow for the differentiation of drivers based on their individual driving characteristics. An emerging challenge within longitudinal studies of drivers is to distinguish between different drivers of a shared vehicle. It also has application in the insurance industry where insurance risk and associated owner premium depends on the diversity or lack thereof of drivers for a vehicle such as a vehicle driven/never driven by secondary drivers that have higher risk driving behaviours. In this paper, a big data set of driving data for 14 older drivers is analyzed - a single year of data includes over 250,000 km and almost 5000 hours of driving for the 14 drivers. A set of 162 trip level calculated features are analyzed to determine their ability to be used to distinguish between two drivers of a vehicle. The results show that features based on road choice and driver chosen velocity provide the best performance individually and in feature pairs with 2 features providing error rates less than 5% for some driver pairs. The set of features that provided the best performance differed for each driver pair and was found to include features from measures of a driver’s road choice, velocity and velocity ratio in addition to the features measuring trip similarity to two phase acceleration and deceleration relationships for the driver. The best error rate obtained was 1.5% for a driver pair. On the other hand, the results suggest that a number of features and feature groups do not allow for older driver differentiation. For instance, overnight driving and high rates of acceleration are not sufficiently exhibited by these drivers to be useful.

Keywords: Acceleration, Global Positioning System (GPS), finite difference equations, driving signature, data analytics, big data.

1. INTRODUCTION

This paper explores measures of driving behaviours to indicate driver unique behaviours, expanding on previous work on measures of acceleration behaviours [1]. The work specifically explores the performance of features measured for trips to be indicative of a driver’s behaviour, allowing drivers to be differentiated. The emergence of low cost sensor and recording technology now allows long term longitudinal studies of drivers [2-4], to understand the individual behaviours of the drivers and how these behaviours may change. In the case of the Candrive Study data [2-3] used for this work, the sensors were deployed in participant vehicles for up to 7 years resulting in a big data set of almost 1TB that can be studied to understand older driver driving behaviours.

There are many applications for the long term study of drivers, including the study of older drivers and the impacts of aging on driving ability or choices [2-5]. Changes in driving may be due to reduced driving performance because of the effects of age and illness, or drivers mitigating declines by avoiding higher risk driving conditions such as night or high speed driving. Fleet management applications for commercial vehicles [6-8] include analytics to lead to higher efficiency or route optimization in taxis and logistics fleets. Analysis to understand driver risk behaviours [9] has applications within the insurance industry [10], where drivers with lower risk behaviours can receive a reduced premium.

Analysis of vehicle sensor data can also improve traffic modeling [11, 12], that can be used for better fleet routing and also for improved consumer navigation services [13]. In addition, the analysis of an individual’s driving behaviours [14, 15] has applications to emerging digital road and
automated/self-driving systems, where more personalization of the behaviour of the automated system leads to improved user acceptance [16, 17].

2. PROBLEM STATEMENT

Previous analysis has identified a number of measures of specific driver behaviours including choice of road, trip length and trip duration [18]. A focused study of acceleration [19] for drivers with stable general, physical and cognitive health [20] that are sole drivers of their vehicles showed drivers have differing acceleration and deceleration behaviors [1, 21, 22].

This paper explores the trip level calculated features individually and in combination for their ability to allow two different drivers to be distinguished. Linear Discriminant Analysts (LDA), [23] which is an application of the Fischer technique for classification, is used to evaluate the features for their classification performance.

3. METHOD AND RESULTS

The Candrive study data collection protocol and sensor system [2, 3, 5] includes an in-vehicle 1 Hz sensor system that connects to a Global Positioning System (GPS) antennae to collect longitude, latitude, date, time and velocity information, and to the On Board Diagnostics II (OBDII) port on the engine computer for access to engine and dashboard parameters such as inertial (dashboard) velocity and accelerator pedal position. In addition the OBDII connection provides power to the sensor system. The processing to ready the data for analysis [24] included techniques required to validate, anonymize and augment the data with additional information such as posted speed limit from geographic information systems.

3.1. CANDRIVE DATA AND DRIVERS

The Candrive study [2] of older drivers provides a big data set of driving data that provides unique research opportunities. For this work, driving was analyzed leading to features measuring behaviours of specific individual drivers. To evaluate their ability to differentiate drivers, all Ottawa drivers were chosen [20] that were sole drivers of their vehicle and also exhibited stable general, physical and cognitive health over their first year in the study. This latter requirement controls for any effect changes in these factors may have on their behaviours. Ensuring all drivers are from the same location reduces any effect of differing locales for day to day driving. The sole driver attribute allows gold standard shared vehicles to be created through combinations of the drivers.

The size of the year 1 driving data for the resulting set of 14 drivers is summarized in Table I. All were active drivers providing a big data set for analysis with over 250,000 km and almost 5000 hours of driving data.

3.2. TRIP FEATURES MEASURED

Each trip driven, where a trip is defined as vehicle being turned on through to being shut off, was analyzed for a number of trip level features. The list of features is shown in Table II and includes a set of 162 features. The “Time of Trip” feature measured the time of day driving patterns of drivers. This includes the tendency a driver has to make trips at similar times of day, and measures of the percentage of a trip that occurs in each of the 24 hours of the day and also for each of the days of the week [18]. Drivers may also adjust their driving patterns based on other influences, such as avoiding high traffic that occurs at rush hour [22] or avoiding driving when it is dark [24]. The northern latitude location of Ottawa and the rate at which they choose to drive and the rate at which they choose to accelerate and decelerate. This leads to a number of measures including those based on actual velocity and velocity relative to posted speed limit. Features have been measured for the percentage of the trip velocity within 10km/hr [18] wide bins and also velocity relative to posted limit in 10% of posted limit wide bins with allowance for driving over the posted limit. Acceleration and jerk (rate of change of acceleration) is measured with features for both the maximum and minimum (largest negative) value in addition to percentage time across a set of histogram bins.

<table>
<thead>
<tr>
<th>Driver</th>
<th>Year 1 distance</th>
<th>Year 1 duration</th>
<th>Year 1 trips</th>
<th>Mean trip length</th>
<th>Mean trip duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>km</td>
<td>hours</td>
<td>count</td>
<td>km</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11500</td>
<td>270</td>
<td>1397</td>
<td>8.3</td>
<td>11.6</td>
</tr>
<tr>
<td>2</td>
<td>39500</td>
<td>297</td>
<td>1175</td>
<td>33.7</td>
<td>15.2</td>
</tr>
<tr>
<td>3</td>
<td>34600</td>
<td>514</td>
<td>1627</td>
<td>21.3</td>
<td>18.9</td>
</tr>
<tr>
<td>4</td>
<td>15300</td>
<td>251</td>
<td>715</td>
<td>21.4</td>
<td>21.0</td>
</tr>
<tr>
<td>5</td>
<td>18300</td>
<td>376</td>
<td>2310</td>
<td>7.9</td>
<td>9.8</td>
</tr>
<tr>
<td>6</td>
<td>7700</td>
<td>208</td>
<td>1046</td>
<td>7.4</td>
<td>12.0</td>
</tr>
<tr>
<td>7</td>
<td>8100</td>
<td>252</td>
<td>1367</td>
<td>5.9</td>
<td>11.1</td>
</tr>
<tr>
<td>8</td>
<td>13200</td>
<td>290</td>
<td>1028</td>
<td>12.9</td>
<td>16.9</td>
</tr>
<tr>
<td>9</td>
<td>19400</td>
<td>457</td>
<td>1748</td>
<td>11.1</td>
<td>15.7</td>
</tr>
<tr>
<td>10</td>
<td>18000</td>
<td>379</td>
<td>1462</td>
<td>12.3</td>
<td>15.5</td>
</tr>
<tr>
<td>11</td>
<td>13800</td>
<td>288</td>
<td>1045</td>
<td>13.3</td>
<td>16.5</td>
</tr>
<tr>
<td>12</td>
<td>13300</td>
<td>247</td>
<td>1033</td>
<td>12.9</td>
<td>14.4</td>
</tr>
<tr>
<td>13</td>
<td>36100</td>
<td>756</td>
<td>3007</td>
<td>12.0</td>
<td>15.1</td>
</tr>
<tr>
<td>14</td>
<td>15200</td>
<td>323</td>
<td>1206</td>
<td>12.6</td>
<td>16.0</td>
</tr>
<tr>
<td>total</td>
<td>264000</td>
<td>4908</td>
<td>20166</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>18900</td>
<td>351</td>
<td>1440</td>
<td>13.8</td>
<td>15.0</td>
</tr>
<tr>
<td>st dev</td>
<td>10300</td>
<td>145</td>
<td>598</td>
<td>7.3</td>
<td>3.1</td>
</tr>
</tbody>
</table>

TABLE I. Summary table for 1 year of driving data for each of the 14 drivers

Drivers may also have unique behaviors in their choice of roads, such as having preference for or choosing to avoid some types of road [18]. As an example, a driver may choose to not drive on expressways or highways. The percentage of each trip taken on roads of various types was measured. The following road types were used: city street, highway, expressway and low speed zones (such as a school zone) by leveraging GIS map [24] augmentation of the data set.

A key aspect of driver choice is the actual speed at which they choose to drive, and the rate at which they choose to accelerate and decelerate. This leads to a number of measures including those based on actual velocity and velocity relative to posted speed limit. Features have been measured for the percentage of the trip velocity within 10km/hr [18] wide bins and also velocity relative to posted limit in 10% of posted limit wide bins with allowance for driving over the posted limit. Acceleration and jerk (rate of change of acceleration) is measured with features for both the maximum and minimum (largest negative) value in addition to percentage time across a set of histogram bins.
The 14 drivers allowed 91 gold standard pairs of drivers to be formed through use of all combinations of two drivers from the 14. Each of these driver pairs is an independent case with this work focusing on performance of individual features and feature subsets across all driver pairs and not identification of unique features or feature subsets to distinguish a specific driver pair.

3.3. Two Phase Acceleration / Deceleration Features

The 1Hz GPS velocity information captured by the Candrive sensor allows for acceleration to be derived using the central two-point difference equation [19] with a 1 second sample period \( h \) (equation 1) providing a record of the driver's acceleration.

\[
\text{Acceleration}[n] = \frac{\text{velocity}[n+1] - \text{velocity}[n-1]}{2h}
\]

Acceleration and deceleration events are defined by equations 2 and 3 respectively. This defines these events as a continuous set of measures with a total velocity change of 4 km/hr or more while ensuring that events that contain sample gaps are split or discarded. Small velocity changes are discarded as these will provide little insight into a driver's preferences. Data analytics signal processing techniques were then applied to each event to measure and capture a number of features describing the event as summarized in Table III.

Acceleration event specification:

\[ \text{Acceleration}[n] > 0 \text{ for } n=k, k+1, \ldots, m \text{ AND } V_{GPS}[m] - V_{GPS}[k] \geq 4 \text{ km/hr} \text{ AND } \]

No sample gaps > 2 seconds (i.e. max 1 sample gap)

Deceleration event specification:

\[ \text{Acceleration}[n] < 0 \text{ for } n=k, k+1, \ldots, m \text{ AND } V_{GPS}[m] - V_{GPS}[k] \geq 4 \text{ km/hr} \text{ AND } \]

No sample gaps > 2 seconds (i.e. max 1 sample gap)

<table>
<thead>
<tr>
<th>Feature Group</th>
<th>Features</th>
<th>Number of features</th>
<th>Feature Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceleration Events</td>
<td>Trip events correlation to self/other driver for Mean/Max</td>
<td>4</td>
<td>1-4</td>
</tr>
<tr>
<td>Deceleration Events</td>
<td>Trip events correlation to self/other driver for Mean/Min</td>
<td>4</td>
<td>5-8</td>
</tr>
<tr>
<td>Trip Attributes</td>
<td>Length of trip in seconds and kilometers</td>
<td>2</td>
<td>9,10</td>
</tr>
<tr>
<td>Time of Trip</td>
<td>Rush hour driving</td>
<td>8</td>
<td>11-18</td>
</tr>
<tr>
<td></td>
<td>Time of day (1hr bins)</td>
<td>24</td>
<td>19-42</td>
</tr>
<tr>
<td></td>
<td>Day of week</td>
<td>7</td>
<td>43-49</td>
</tr>
<tr>
<td></td>
<td>Solar cycle</td>
<td>4</td>
<td>50-53</td>
</tr>
<tr>
<td>Posted Limit on Road</td>
<td>&lt;=40km/hr, 41-70km/hr, 71-90km/hr, &gt;90km/hr, unknown</td>
<td>5</td>
<td>54-58</td>
</tr>
<tr>
<td>Velocity</td>
<td>Actual velocity (10km/hr bins)</td>
<td>13</td>
<td>59-71</td>
</tr>
<tr>
<td></td>
<td>Velocity ratio to posted (10% bins)</td>
<td>13</td>
<td>72-84</td>
</tr>
<tr>
<td>Acceleration</td>
<td>Min and Max</td>
<td>2</td>
<td>85,86</td>
</tr>
<tr>
<td></td>
<td>Actual Acceleration histogram bins (-3 to 3m/s²)</td>
<td>33</td>
<td>87-119</td>
</tr>
<tr>
<td></td>
<td>Min and Max</td>
<td>2</td>
<td>120,121</td>
</tr>
<tr>
<td></td>
<td>Actual Jerk histogram bins (-0.5 to 0.5m/s³)</td>
<td>41</td>
<td>122-162</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>162</td>
</tr>
</tbody>
</table>

TABLE II. Summary of features for each trip evaluated for their performance to distinguish drivers.
Table IV summarizes the number of acceleration and deceleration events that were identified for each of the 14 drivers in each of their first 2 years in the study. The big data attributes of these acceleration and deceleration events are demonstrated by the number of events identified within a year as the drivers demonstrated an average of between 16 and 26 deceleration events/trip and between 17 and 28 acceleration events with a mean of 21 events in both cases. This large and rich set of events provides a data set that can be explored to identify the behaviours of the drivers for patterns and driver unique attributes.

The identified events for each driver were analyzed and contrasted against various features to identify patterns within the data. One visualization of the data segmented the events based on the size of the acceleration (total velocity gain) in 2 km/hr bins. Figures 1 and 2 show examples of this visualization where histograms of the mean and maximum acceleration values within the acceleration events for each of the velocity change bins is presented. The resulting surface shows the mean and maximum accelerations chosen by the driver compared with the size of the total velocity change. Both of the figures indicate a similar pattern where there is a ridge line in the driver behaviour for the histogram peak that initially increases with the size of the velocity change and then slows for larger velocity change values. A driver’s choice in acceleration behaviour will be influenced by their personal preferences and also by surrounding traffic that may limit their rate of acceleration. It is likely that larger velocity changes would be less influenced by other drivers while small changes would include many events that occur within the ebb and flow of traffic.

Table IV. Summary of the number of acceleration and deceleration events that were identified for each of the 14 drivers over their first 2 years within the Candrive study.

<table>
<thead>
<tr>
<th>Year</th>
<th>Acceleration Events</th>
<th>Deceleration Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>2015</td>
<td>17</td>
<td>28</td>
</tr>
</tbody>
</table>

The Matlab fit function for a Gaussian distribution was applied to each of the histograms to provide measures of the mean and standard deviation of the resulting distributions. The identified mean for the Gaussian distribution indicates an estimate of the ridge line for the driver and the resulting relationships for the example driver are shown in Figure 3.

Deceleration [21] and acceleration events provide a potential measure of a driver's preference as the other 13 drivers also showed similar characteristic ridge lines with Figure 4 showing the mean and maximum lines found for 4 example drivers from the set of 14. This shows that although some drivers have similar relationships, there are differences that make them unique.

The two phase relationships were used to provide a set of measures for the similarity of a trip to each of the driver's typical behaviour and each of the other drivers. This was measured through calculation of the correlation of the features a trip’s acceleration and deceleration events with the two phase relationships for each of the drivers. Within a given trip, the number of events can be highly varied, as typically shorter duration trips will have fewer events but also extremely long duration trips can also have a relatively low number of events if they include long periods of highway driving.

3.4. Single Feature Performance

The 14 selected drivers provide a set of 91 gold standard cases that can be used to analyze feature performance as the trips for two drivers can be combined to create a "shared" vehicle where the actual driver for every trip is known. Given the large number of trips driven by the 14 drivers within the year as shown in Table I, thousands of trips were available for training and testing. A train and test method was applied where a subset of the data was used for training and the balance for testing. The available data for each case were divided randomly into 10 groups containing 10% of the trips each. This allowed for the train and test to be repeated 10 times where 90% of the data was used for training and each of the 10% groups used for testing. The performance presented is based on the average for the 10 models.

A number of features provided no classification value at all, but this was expected as these features measured driving attributes that were atypical of the older drivers in the sample. A feature that failed for all driver pairs was middle of the night driving between 1am and 5am, while late night (10pm-1am) and early morning (5am-8am) failed for some driver pairs. Similarly driving at dawn based on the sunrise time failed for some driver pairs. Older drivers typically do not drive during these times of day and hence the features contained minimal information.

Night driving due to the early winter sunset in Ottawa did contain enough information for each driver to provide some classification value. Other features that failed in some or all cases included the highest and lowest bins for acceleration and jerk, especially large positive values as these balanced histograms were designed to span the values observed in the drivers and some drivers again did not exhibit behaviors at the extremes. The imbalance in the acceleration and jerk values is expected as hard braking is a more frequent event than matching levels of hard acceleration. The result, up to 22 of the features provided no distinguishing value in some or all of the driver pairs.

The feature group that provided the best overall performance with error rates as low as 3.4% for one driver pair was road type choice as measured through posted speed limit information. Although these provided strong classification performance for some driver pairs, the average error performance for these features individually across all 91 pairs was 30-43% with worst case error rates of over 50%.

Another feature group that shows low error rates performance for some pairs is the velocity ratio. This group had best errors rates for some driver pairs between 5 and 16%, while again having an average error rate performance between 30-
40%, and worst case error performance over 50%. No other feature group exhibited best error performances better than 10%, although many provided error rates between 10 and 25% for best cases. The summary in Table VI shows the number of features providing various levels of best performance.

Figure 1: Distribution of the mean acceleration for example driver against the size of the velocity change in the event. (N= 25261 events).
Figure 2: Distribution of the maximum acceleration for example driver against the size of the velocity change in the event. (N= 25261 events).

Figure 3: Plots of the two-phase relationship for an example driver for the mean surface in Figure 1 (blue) and maximum surface in Figure 2 (red). The estimated peak is shown along with 95% confidence intervals. Best fit two phase relationship with optimal transition between the phases is shown by the lines.
These results lead to the conclusion that there is no one feature that can be used alone to distinguish every driver pair and that combinations of features are likely going to be required to achieve the best performance.

3.5. FEATURE PAIR PERFORMANCE

To further explore the classification potential of the features, LDA classification models similar to the single feature models in the previous section were run for all pairs of features across all 91 driver pairs.

The computation for this research case required careful design given the large data sets being analyzed and the large number of cases under consideration. With 162 features, there are 13,041 unique pairs of features that need to be evaluated for each of the 91 driver pairs using the same 10 repetitions of the 90% train, 10% test models. This leads to >11.8M cases to be computed. Computation on an Intel i7-4770 @3.4Ghz processor was estimated at ~1 year for serial execution of the cases. The independence of the cases allowed the task to be easily spread across a cluster machine greatly reducing computation time.
The results for the 2 feature classifiers are presented in Figures 5 through 7 for each of the feature pairs. The diagonal line on the plots can be ignored as this indicates results for both features being the same which is invalid. Figure 5 shows the best performance from a 2 feature classifier and a number of attributes can be observed. In these plots, all features are presented including those that provided minimal single feature classification ability. In all three figures, the set of yellow lines shows features that failed to provide any classification and match with the features that failed in the 1 feature cases presented previously. Figure 5 shows the best performance from a 2 feature classifier and the darker blue bands show the features bands that provide the best classification performance. Again, these bands align with the features groups that provided best single feature performance, namely velocity ratios and road choice. Figure 6 shows the worst performance for each feature pair and the lack of any dark blue indicates that no feature pair provides good performance across all driver pairs. Figure 7 shows the mean performance for the feature pairs and this also shows that no feature pair provides strong performance across all driver pairs.

3.6. Feature Set Performance

Table VII summarizes the results for the 91 driver pairs showing the ability for an adaptive number of features beyond two to distinguish drivers. The best performance observed has 1.5% error rate across year 1 trips. The poorest performance was an error rate of over 22.0% with a mean of just over 8.6%. One impact on the performance for trips with more acceleration and deceleration events was analyzed by looking at trips with greater than 5 events as compared to the subset of these trips that had greater than 10 events. Trips with more events are likely longer in duration and will also provide better sample sets as measured against the two-phase behaviour models for the drivers. Performance improved for the set of trips with a larger number of events.

The set of features that provided the best performance for each of the driver pairs was different but they consistently included features from the groups that measured road choice, velocity and velocity ratio in addition to the features measuring trip similarity to the two phase acceleration and deceleration relationships for the drivers.

<table>
<thead>
<tr>
<th>Error Rate</th>
<th>Acc/Dec events &gt;= 5</th>
<th>Acc/Dec events &gt;= 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>1.6%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Worst</td>
<td>23.5%</td>
<td>21.5%</td>
</tr>
<tr>
<td>Mean</td>
<td>8.7%</td>
<td>8.6%</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>4.4%</td>
<td>4.0%</td>
</tr>
</tbody>
</table>

TABLE VII. Summary of performance possible for an adaptive number of features to achieve best possible performance. Effect of number of acceleration and deceleration events in the trips shown.
Figure 5: Best error performance for 2 features for each feature pair using LDA classifiers. Features are numbered according to Table II.

Figure 6: Worst error performance for 2 features for each feature pair using LDA classifiers. Features are numbered according to Table II.
4. DISCUSSION

This paper uses big data analytics on a large dataset that contains a full year of driving by 14 drivers, and shows that trip level big data analytics techniques can be used to create a set of features for each trip. It specifically evaluates the performance of these trip level driving behaviour features in their ability to differentiate between drivers.

The paper shows that features related to overnight driving and extremely high levels of acceleration have no value in the differentiation of the older drivers studied, as there are not enough of these events. However, these features may be of value in other driver agegroups or segments where these behaviours are more prevalent.

The paper shows that features related to road type choice and velocity, including velocity relative to posted limit, provide the best performance individually and when paired with other features. In the study of 14 sole drivers it was shown that combinations of features allowed on average 91.4% of the trips driven within a driver pair to have the driver identified correctly, with a best performance of 98.5%. It was also shown that the performance improved when trips with lower numbers of deceleration and acceleration events were excluded as would be expected since trips with few acceleration events would typically be shorter in duration reducing the opportunities for distinctive driver behaviours to be exhibited.

The features that have been identified that distinguish drivers are all features that measure a drivers behaviour and hence choice while driving. Drivers choose the road types they prefer to use such as a choice to avoid or preferential choose highways. Drivers behaviours will also influence their tendency to drive at, above or below the speed limits and also their preference for acceleration and deceleration such as more aggressive and faster accelerations/decelerations instead of slower changes in velocity.

This work is applicable to longitudinal driving studies where the identity of the driver of a vehicle has to be validated, such as a vehicle shared by spouses. It can also be applied to insurance industry applications where the risks associated with insuring a vehicle. Insurance industry premium discounts or price increases depend on the crash risks associated with a vehicle which will depend on the number of distinct drivers and especially on their differing behaviours, such as aggressive driving or avoidance of higher risk driving situations.

The paper demonstrates processing using big data analytics on a large data set of driver data. Using event classification and signal processing techniques, it measures features of events to create a large set of events for a given individual that can then be summarized into a simple two phase relationship that provides a measure of a driver’s individual driving behaviours.
5. ACKNOWLEDGMENTS
This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC), the Bruyère Research Institute - Bruyère Academic Medical Organization grant and the Ontario AHSC AFP Innovation Fund. The Candrive study was funded by a Team Grant from the Canadian Institutes of Health Research (CIHR) entitled "The CIHR Team in Driving in Older Persons (Candrive II) Research Program" (grant 90429). Additional support was provided by the Ottawa Hospital Research Institute and the Toronto Rehabilitation Institute, University Health Network, and the University of Manitoba.

The Candrive Investigators thank the Candrive cohort study participants for their dedication. Without their commitment, this publication would not have been possible. The Candrive Investigators thank Lynn MacLeay, Candrive Program Manager, for her role in managing and operationalizing the study for the Canadian sites. The Candrive Investigators thank the Research Associates for their dedication to the full cohort study. Jennifer Biggs and Anita Jessup (Ottawa Coordinating Centre). The Candrive Investigators wish to thank the following individuals for their work associated with the in-vehicle driving data: Akram Alakel (Ottawa), Glenys Smith, Andrew Cull, Satoru Nakagawa, Sandra Webber and Maureen Babb (Winnipeg).

6. REFERENCES
Symposium on Medical Measurements and Applications (MeMeA), pp. 391-396.


Authors

Bruce Wallace received the B.E.Sc. degree in electrical engineering from the University of Western Ontario, in 1987, the M.Eng. degree in systems and computer engineering and a Ph.D. degree in systems engineering (biomedical engineering and signal processing) both from Carleton University, in 1991 and 2017 respectively. He joined Bell Northern Research, in 1987, as a Designer. He joined Nortel, in 1991, where he led a design team followed by leadership roles in technology strategy and business development. He was the Leader of Global Healthcare Solutions, then with Nortel, and continued in same role at Avaya, after acquisition until returning to Carleton University in 2011. He is currently a Post-Doctoral Fellow with Carleton University and Bruyère Research Institute and a Lecturer with the Systems and Computer Engineering Department, Carleton University, where he teaches courses in problem solving, computer programming, and numerical methods. He holds 17 U.S. patents and 2 Canadian patents. His current research interests include the application of sensors, signal processing, and big data analytics to cognition measurement with a specific focus on the cognition issues associated with aging and older adults.

Frank Goubran received the B.Sc. degree in electronics and communications engineering and the M.Sc. degree in electrical engineering from Cairo University, in 1978 and 1981, respectively, and the Ph.D. degree in electrical engineering from Carleton University, in 1987. He joined the Department of Systems and Computer Engineering, Carleton University in 1987, he was the Chair of the Department of Systems and Computer Engineering from 1997 to 2006 and Dean of the Faculty of Engineering and Design from 2007 to 2016. He is currently a Professor and the Vice-President (Research and International). He was involved in several research projects with industry and government organizations in the areas of digital signal processing, biomedical engineering, sensors, data analytics, microphone arrays, and the design of smart homes for the independent living of seniors. He was the Founding Director of the Ottawa-Carleton Institute for Biomedical Engineering, from 2006 to 2009. He is currently a Scientist with the Bruyère Research Institute. Dr. Goubran is a fellow of the Canadian Academy of Engineering and a member of the Association of Professional Engineers of Ontario.
Michelle Porter has been conducting research in the area of older drivers, utilizing a variety of in-vehicle technologies since 1999. She is currently the Director of the Centre on Aging, and a Professor with the Faculty of Kinesiology and Recreation Management, University of Manitoba, Winnipeg, MB, Canada. Her current research interests include driving and aging, mobility, and resistance training.

Andrew Smith received the B.Sc. degree in human kinetics from the University of Guelph, Guelph, ON, Canada, in 2008, and the M.Sc. degree in biomechanics from the School of Human Kinetics, University of Ottawa, Ottawa, ON, Canada, in 2012. He is currently pursuing the Ph.D. degree with the School of Human Kinetics, University of Ottawa. He has been with Biomechanics Laboratories, University of Ottawa, and The Ottawa Hospital Research Institute, Ottawa, and a Researcher with the Ottawa Hospitals Center for Rehabilitation Research and Development, Ottawa. His current research interests include aging, assisted mobility, and biomechanical contributions to joint and walking stability.

Shawn Marshall received the master’s degree in clinical epidemiology from the University of Ottawa, Ottawa, ON, Canada. He was a Founding Co-Investigator for the Canadian Driving Research Initiative for Vehicular Safety in the Elderly (Candrive), and is currently a Co-Principal Investigator for the CIHR funded Candrive Older Driver Prospective Study from 2014 to 2016. He is a Professor with the Department of Medicine, University of Ottawa, Ottawa, ON, Canada. He is a Physician (Physiatrist) specializing in physical medicine and rehabilitation. He is the Medical Director of Acquired Brain Injury Rehabilitation with the Ottawa Hospital Rehabilitation Centre, Ottawa, where he manages both in-patients and out-patient clinics. His current research interests include driving and disability, specifically older driver research, conditional licensing, and assessing medical fitness to drive.