Optimising driver profiling through behaviour modelling of in-car sensor and global positioning system data

Gabriela Ahmadi-Assalemi[1], Haider M. al-Khateeb[1,*], Carsten Maple[2], Gregory Epiphaniou[2], Mohammad Hammoudeh[3], Hamid Jahankhani[4], Prashant Pillai[1]

[1] Wolverhampton Cyber Research Institute (WCRI), School of Mathematics and Computer Science, University of Wolverhampton, WV1 1LY, West Midlands, U.K.
[2] Warwick Manufacturing Group (WMG), University of Warwick, International Manufacturing Centre, CV4 7AL, Coventry, U.K.
[3] School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, M15 6BH, Manchester, U.K.
[4] Northumbria University London Campus, E1 7HT, London, U.K.

[*] Corresponding author. Email address: H.Al-Khateeb@wlv.ac.uk

ABSTRACT - Connected cars have a massive impact on the automotive sector, and whilst this catalyst and disruptor technology introduce threats, it brings opportunities to address existing vehicle-related crimes such as carjacking. Connected cars are fitted with sensors, and capable of sophisticated computational processing which can be used to model and differentiate drivers as means of layered security. We generate a dataset collecting 14 hours of driving in the city of London. The route was 8.1 miles long and included various road conditions such as roundabouts, traffic lights, and several speed zones. We identify and rank the features from the driving segments, classify our sample using Random Forest, and optimise the learning-based model with 98.84% accuracy (95% confidence) given a small 10 seconds driving window size. Differences in driving patterns were uncovered to distinguish between female and male drivers especially through variations in longitudinal acceleration, driving speed, torque and revolutions per minute.

Keywords: driver identification, behaviour profiling, classification, machine learning, connected cars, random forest, GPS, cybersecurity threat, incident response

Highlights -

- Successful classification based on a short segment of driving data
- Driver gender was identifiable through profiling driver groups
- Downloadable dataset built from actual (non-simulated) experiment
- Feature Selection optimised to increase the accuracy of the result
- Piloting lessons learned shared to facilitate future experiment design

1. INTRODUCTION

The Internet-of-Things (IoT) has a massive impact on society and business, particularly in the area of Connected Cars (CC) with anticipation of significant opportunities and challenges in the automotive sector as the transport experience changes [1]. According to Gartner, it is expected that, by 2020, wireless connectivity will be present in a quarter of a billion cars, representing roughly one in five vehicles worldwide. Many applications can be enabled by advanced driver profiling for CC including novel charging models, wherein an insured driver is only authorised if they comply with an insure-as-you-go model [2]. The motivation for this paper arises from existing and emerging cybersecurity threats in CC. For example, carjacking is a serious concern. Advances in key-less car technologies result in innovative car theft techniques, the more conventional car antitheft methods have become less effective. Although the Office of National Statistics (ONS) reports that car keys are the predominant method of stealing cars, the keyless theft of cars will likely increase as more cars adopt keyless technology making them susceptible to innovative theft [3].

In modern cars, Event Data Recorders (EDR) and the On-Board Diagnostics (OBDII/OBD2) produce and process a myriad of data from in-car sensors that are capable of complex processing. OBD2 has gradually evolved since the 1960s with contributions from several organisations namely: California Air Resources Board, the Society of Automotive Engineers, the International Organisation for Standardisation, and the Environmental Protection Agency (EPA). The early OBDs communicated over proprietary protocols with varied feature sets. OBD was standardised and became mandatory in the USA from 1996, and in the European Union (EU) from 2001 for all petrol vehicles, and from 2003 for all diesel vehicles. Additionally, the OBD had further harmonisation and standardisation using the Controller Area Network (CAN) protocol specified in ISO 15765 from 2008 onwards. A coherent set of specifications for OBD2 port location, specification, and use is set out in ISO 15031-3. Another source of critical data is the Global Positioning System (GPS) which enables vehicles to receive signals from the GPS satellite to calculate the vehicle's three-dimensional position and time. Then, the Electronic Stability Control (ESC) which initially introduced in 1995 quickly caught on. From 2005 a growing number of cars were fitted with ESC, and by 2012 all cars under 10,000lb gross weight had an ESC. Consequently, the European Parliament legislated that all new cars from 2014 are fitted with an ESC. ESC uses a unit to monitor the steering wheel angle and the car’s rotation over its vertical axis recording acceleration and direction. Furthermore, car scanners became more accessible and economically viable. The car we utilised for this study is equipped with ESC, GPS, and OBD2 diagnostics port, this being the norm for cars manufactured from 2012 onwards, and in many cases earlier than that.

Advanced interconnectivity in CC combined with valuable information extracted from in-car data poses significant cybersecurity threats [4]. The infotainment system of a Jeep Cherokee was remotely exploited, BMW’s ConnectDrive was compromised, remotely unlocked, and its location and speed tracked [3]. Nissan Leaf was accessed through its Application Program Interface (API) and through the in-car system, control was gained over the battery life, usernames, trip schedules and distances. Although the exploits were carried out ethically (pen-testing), such vulnerabilities could be utilised by attackers maliciously compromising the drivers’ privacy which could include theft of personal information [1]. Therefore, it is within the scope of this study to investigate if further private information can be learned about the driver if sensor data is analysed by an attacker.

Whilst the data from CC can be used for car-GPS stalking, it can also be utilised for better detection and prevention techniques to create a safer driving environment. Recent research explored opportunities to improve road safety by innovatively combining data and methodologies to monitor drivers, for example, [5] asserts useful insight into drivers' behaviour and concludes that car turns are well suited to identify drivers. However, more experiments are needed in this growing research area to streamline prominent findings.

There are some attempts in the literature to report driving habits in relation to age [6] or when drivers are distracted [7]. However, there is little or no research covering other demographic aspects of behavioural profiling. As an emerging area, it is crucial to building a better understanding that could also lead to enhanced consumer trust and confidence towards the underpinning technology. In this study, we build a new dataset from a car-based experiment including a pool of volunteer drivers. We then investigate how we can better understand their driving behaviour. We analyse the impact of feature selection on the accuracy and detection time in driver profiling, then we apply our model to a short segment of driving data from in-car sensors that reflect driver behaviour to produce authorized driver profiles.

The remainder of this paper reviews related work in Section 2, describes the experiment and the method of driver classification in Section 3, presents and analyses the results in Section 4 that are then discussed in Section 5 in relation to the original research questions. Finally, Section 6 concludes this paper and recommends future research directions.

2. BACKGROUND AND RELATED WORK

2.1. Inherited and emerging threats facing CC

Connected cars could transform the industry and society, but it is not without risk. From the early 1990s, proliferation of technological advances in modern vehicles to maximise the consumers' experience, comfort, and safety led to introducing various sensors measuring a wide variety of aspects. Whilst technology is widely used in social and individual dimensions, the fusion of in-car, social and driver aspects is still developing, and the subject of further research. Consumers desire cutting-edge technology, but few of them understand the associated risks and vulnerabilities. Currently, evidence indicates that commercial cost control, safety and introducing new attractive features rather than security-by-design (SbD) drives the automotive industry. At times, this leads to the re-use of older technology such as the CAN-bus developed in 1988 or OBD2 maintenance port disregarding the flaws of such an approach. This asserts that a holistic approach is required and more needs to be known between the different stakeholders, suppliers, manufacturers and security experts. Likewise, better regulation is needed amongst stakeholders.

Environments in CCs are highly heterogeneous, and use a variety of protocols to interconnect through gateways as shown in Figure 1. However, the Local Interconnect Network (LIN), CAN, Media Oriented System Support (MOST), ECU, and FlexRay create an opportunity for a protocol layer attack over 4G, Bluetooth, Tyre Pressure Monitoring Systems (TPMS), Wi-Fi hotspot or keyless entry system. Once the in-car network is compromised, the level of vulnerabilities and damage will depend on the approach to any additional security measures. An exploratory study [1] analysed attack vectors of in-car bus systems, described the external attack surface, and argued that all cybersecurity aspects should be addressed within the context of Cyber-Physical Systems (CPS). The experiment also recognises authentication as a measure of protection from external attackers. A comprehensive risk analysis model was proposed demonstrating feasible cyber-attack scenarios including but not limited to direct access, software-driven, Bluetooth-based attacks.

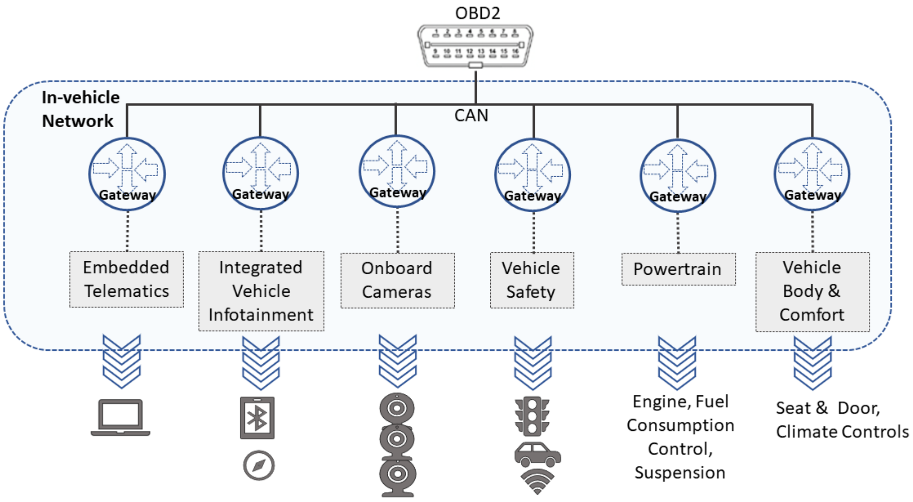

Figure 1. Typical in-car network architecture

2.2. *Car antitheft methods*

Car theft has long been identified as a serious concern with no strategic defence mechanism to mitigate attack vectors. According to published figures by the ONS for the year ending March 2017, vehicle theft in England and Wales rose consistently since 2006. In the United States (US), according to the Federal Bureau of Investigation (FBI) Uniform Crime Report in 2016, motor vehicle theft was estimated at 765,484, a rise of 7.4% on the previous period, and this increase continued in the first half of 2017.

Car hijacking, another form of car theft, deviates the car from its intended route and destination. Initially, Tencent Keen Security Lab reported in 2016 that Tesla-manufactured cars have this vulnerability, specifically the S model remotely controlling the car's controls compromising the CAN bus. Then, in 2017, they found a similar vulnerability with CAN and ECU bus in the Tesla X model. The vulnerability was caused by a malicious Wi-Fi hotspot. Although this type of attack required proximity to the car, the risk was not negligible. In BMW, ConnectedDrive was reverse engineered and vulnerabilities were exploited to remotely unlock the car, tracking its location, speed and viewing emails. Vulnerabilities were exposed relating to insecure data transmission unprotected from replay attacks that allowed connectivity to rogue mobile phone networks. Attackers could harvest data and seize control because BMW used symmetric keys in vehicles, which caused the ConnectDrive control unit to expose the Vehicle Identification Number (VIN) in the Next Generation Telematics Protocol (NGTP) error message, as some services did not encrypt communication whilst NGTP messages were only encrypted with data encryption standard (DES). In modern cars, hijacking does not necessarily take place remotely. It can, in fact, be carried out by someone who is initially an authorized driver but may not use a GPS predefined route. This is apparent in the rise of recent widely reported terror-related attacks utilizing vehicles to drive into crowds in London, Munster, and Barcelona. Modern CC require a proactive approach to help detect and prevent such attack.

Nonetheless, a distributed firewall system was proposed by [4], and a 3G based video surveillance and Carbon Dioxide (CO2) car defence system were presented by [8]. Likewise, [9] proposed a Radio Frequency Identification (RFID) module triggered by vibration and pyroelectric infrared sensors if the car is stolen. A Global System for Mobile (GSM) communication will then be utilised to pass the GPS coordinates of the car to the registered owner's mobile phone. Such security systems, whilst crucial, are not driver-centric, they are not aware of the driver's behaviour characteristics or changes in behaviours.

2.3. Driver classification through behavioural analysis

Drivercentric proactive driving anomaly detection is presented by [10] to improve cyber-resilience through driver behavioural profiling, using Machine Learning (ML) when vehicles deviate from a defined route to prevent hijacking. Methods that analyse driving patterns are presented in [3] and [11] to identify unauthorized drivers. However, these methods are usually limited by false positives, accuracy, and the window size required to collect and analyse driving data. Driver behaviour signals were studied by [12]; the behaviour was observed while following another car, and the best driver identification rate for the sample (n=30) was 73% using a Gaussian Mixture Model. In another study by [13], the AutoSim driving simulator was utilised to build a virtual road scenario and collect data for 20 subjects. Behaviour was modelled using a Hidden Markov Model (HMM) and the accuracy was up to 85%. Traffic safety improvement was one of the motivations for many studies such as [3, 5, 14]. Few datasets were made available for research purposes including driver behavioural analysis, an in-car multimedia data collection of audio-video and vehicle data was shared by [15].

An experiment conducted in the streets of Seoul, South Korea by [3] incorporated additional car's mechanical features that were excluded in previous works and demonstrated their value in reflecting drivers' behaviour. The authors tested a model to decrease time and increase detection in driver identification as a cybersecurity measure. The experiment, which included a limited number of participants (n=10) revealed an opportunity to use in-car sensors to identify different drivers with a size of a sliding window of 60 seconds. Furthermore, [14] and [5] showed that short driving sessions could reveal enough behavioural differences.

Other studies utilised smartphones; they are inexpensive and available with inertial sensors that can extract data to enable the classification of driving events. Some of these applications that use smartphone sensors technology for intelligent transportation were reviewed and analysed by [16] to highlight weaknesses and suggest improvements. In [17], a Dynamic Time Warping (DTW) algorithm was used to determine the type of driving style and manoeuvres with similar accuracy to the car's CAN bus, and [11] researched a driver authentication system as an anti-theft measure based on data collected by a smartphone.

A time-optimized driver pattern fingerprinting method by [18] was recently proposed. They investigate 3 different datasets and a number of different ML techniques: K-nearest Neighbour (KNN), Random Forest (RF), Extra Trees, Decision Tree (DT) and Gradient Boosting with Random Forest (GBRF). The study found that GBRF performed better than other algorithms and demonstrated the possibility to identify drivers within 3 minutes.

An algorithm review by [19] shows that clustering techniques such as k-means can be used for driving-style distinction and labelling [2] and particularly driver identification [20]. Fuzzy logic was also used in [21]. Support Vector Machines (SVM) and K-mean were successfully used by [22] but, despite encouraging results, their sample consisted of just two drivers. Furthermore, to solve a driver prediction problem, [5] applied an RF classifier successfully. Therefore, a number of effective driver modelling techniques such as SVM, RF and Naïve Bayes (NB), can be applied to in-car sensor data. A comparison of these techniques [3] showed that RF consistently provides the most accurate results in driver identification including when feature sets were varied.

3. METHODOLOGY AND EXPERIMENTS

To address the research questions which can be expressed as: How fast can we distinguish between the different drivers of a car utilising widely available in-car sensors based on their behavioural patterns? What else can we learn about these drivers? What should be an example of a working combination of parameters, characteristics, and features to increase the accuracy in driver profiling? We utilise empirical data collected from a real passenger car to facilitate the experiments. We apply our model to relatively short segments of driving data from widely available in-car sensors, and produce a dataset that reflects driver behaviour to produce authorised driver profiles. Authorised drivers are considered to have regular or permitted use of the vehicle. Therefore, authorised drivers have an existing ML-trained driver profile, unlike unauthorised drivers. Identifying drivers in this study is within the realm of distinguishing different drivers based on patterns emerging as a result of their driving behaviour. The problem is solved as, all drivers with trained ML driving profiles for the vehicle compared to everyone else who does not have a relevant trained driving profile for the vehicle. Therefore, based on the scenario in our experiment, we do binary classification. However, we also test binary classification to show that we can recover a personal detail about the driver e.g. a male or female, which was possible due to the demographic information collected.

3.1. Experiment Design

The experiment was conducted as discussed below.

Participants. Email and social media were utilised to invite volunteers. All participants were required to be over 18, have a valid driving license, and be covered by full and comprehensive car insurance. All participants had to complete a consent form and were given the option to opt-out at any time.

Environment (the route). The route is 8.1 miles long and includes urban roads with roundabouts, traffic lights, and streets with higher volume traffic, but also zones with less traffic with 20, 30 and 40mph speed limits , a combination of local roads with focus on turning and manoeuvring at low speed and roads, with the higher speed limit and less traffic, enabling a wider range of characteristics typical of a driver to be captured. It was the aim for all participants to have similar driving conditions in terms of traffic volumes and weather, and a key criterion was to keep the driving track simple but versatile including many different facets, turns, junctions, roundabouts, road humps, and various speed zones. The route is shown in Figure 2.

Instruments. The car used for this study is equipped with Electronic Stability Control (ESC), Global Positioning System (GPS) and On-Board Diagnostics (OBD2) port. The driving data was recorded using a right-hand drive 2009 Mercedes Benz CLS coupe passenger car fitted with a single-moulded BAFX A5-58GD-LWQN Bluetooth enabled OBD2 scanning tool, that received power from the car’s internal battery and was connected to a Lenovo Android tablet via Bluetooth. An Android application, Torque

Pro, collected the data from the scanning tool as demonstrated in Figure 3. The car profile was set up on the Torque Pro application. The tablet was battery-powered during the experiments. The data was recorded at 1 second intervals, and files were exported and stored chronologically in .csv and .xlsx formats. Additionally, a paper-based questionnaire was used to collect demographic information from the participants.

Finally, Microsoft Excel 2016 version 16.15, WEKA 3. 8.2 and R implementation version 1.1.442 was used on an Intel Core i7 3.3 GHz 16GB RAM MacBook Pro to process and analyse the dataset

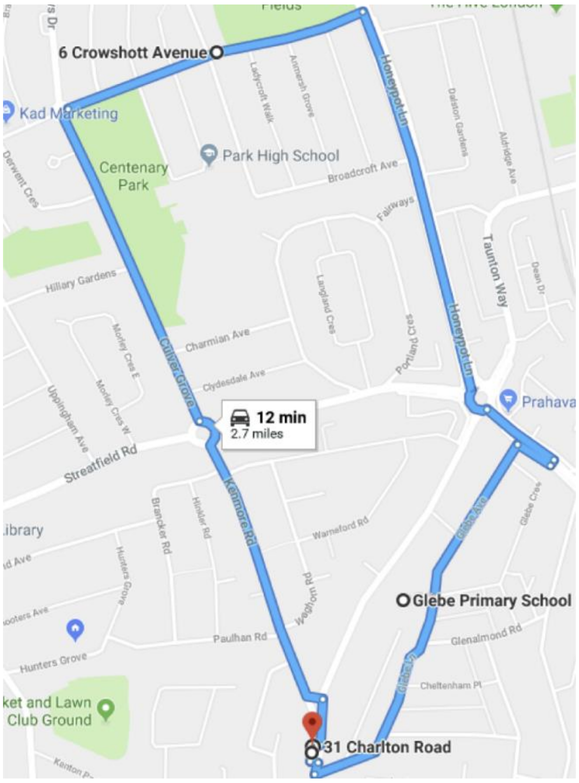

Figure 2. The experiment predefined route in London is 8.1 miles.

Procedure. Firstly, the participants completed a questionnaire gathering demographic information and their driving habits through self-reflection before the driving session, so that the completed driving session would not bias their responses. Besides, the drivers had an opportunity to get used to the vehicle controls before completing the experiment. Next, the participants completed the driving along the predefined route shown in Figure 2 and returned to the original start point with an overall distance of 8.1 miles. The driving session for each participant lasted approximately 45 minutes during one of the pre-identified time allocations where road traffic is relatively similar. These time-slots were 10am-12:30pm, 1-5:30pm or 7-10:30pm London time. A session was split into 3 laps, each lap was driven during the same driving session in quick succession, gathering 153.9 miles of driving data.

At the start of each driving session, the OBD2 scanning tool was connected to the car's OBD2 port located under the steering wheel before the ignition was started. Once the engine was running, the OBD2 scanning tool was automatically paired with the Android tablet and the Torque Pro application ready to collect the data from the ECU via CAN protocol. The data collection started automatically as the car moved.

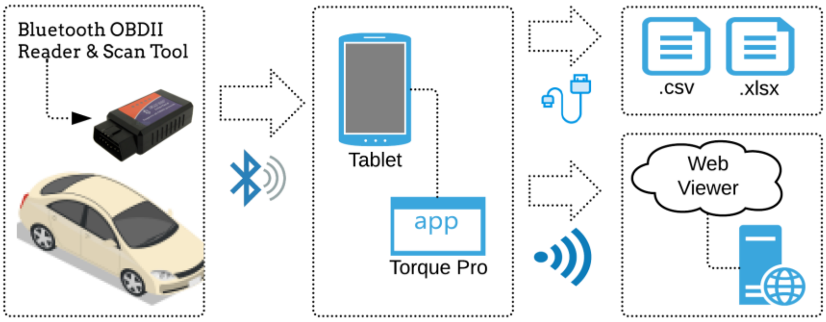

Figure 3. Data Collection Process.

Collected data. Firstly, the questionnaire was designed to gather the following demographics: age group, gender, occupation, ethnicity, and questions related to driving experience (daily driving frequency, daily distance, time since the first driving licence was granted, and the number of accidents in the past year). Furthermore, participants were asked to self-describe their driving style with one of the following options: "Dissociative", "Anxious", "Risky", "High-Velocity", "Distress-Reduction", "Patient" and "Careful". Descriptors have also been shared to explain each style. These were adopted from [23]. Secondly, data collected via the in-car sensors have been fully described in the dataset (download link shared in the next section).

3.2. *The Dataset*

The data was captured with 100 milliseconds and 1 second rate. The formal experiment collected 153.9 miles of data consisting of 57 individual tracks and 14 hrs of driving. The dataset can be download from https://doi.org/10.13140/RG.2.2.14505.49765.

3.3. Piloting

While the above describes the formal experiment for this study, a pilot experiment took place at an earlier stage. It involved 3 participants and helped to test and evaluate the instruments and procedures planned for the formal experiment.

4. RESULTS AND DATA ANALYSIS

4.1. Participants' Demographic Distribution

The formal experiment recruited 19 participants (m=6, f=13) from several age groups (<30, n=3; 30-40, n=4; 34-50, n=4; >50, n=8). All participants were employed and most of them had a driving licence for many years (1-5 years, n=1; 6-10, n= 3; >10, n=15). When asked about their driving frequency on average, the majority (n=13) said they would drive 2-3 sessions per day; we define the driving session here as an independent driving episode with at least 30 minutes time gap from the next one. Otherwise, 1 participant claimed 1 session per day, and 1 participant claimed more than 5 sessions per day. The average daily distance driven was: <5 miles, n= 4 drivers; 6-10, n=4; and >10, n=11. None of the participants reported an accident within the last year. This sample of participants self-described their driving style as Patient (n=17), but they have also selected Careful (n=12) and High-Velocity (n=5). These demographics were used to investigate possible correlations with driving behaviour. Participants were given the option to be excluded from this study at any time, but none asked to opt-out.

4.2. Feature Pre-Processing

This stage combines the raw data collected from the experiment and questionnaire generating 18 features. Numeric identification (ID) has been assigned to each participant to link with each lap within the experiment dataset which was built in a .csv format.

4.3. Feature Selection

This study aimed to collect features that were not specific to any type of car manufacturer, gear-box type or fuel type; features including the torque and the revolutions per minute (RPM) that could be collected with a generic and widely available non-manufacturer specific OBD2 scanning tool or application in order to support wider applicability.

Duplicate or irrelevant features were excluded from the dataset. For example, the *"GPS Satellite Lock Count"* did not represent a driver or a car characteristic; the "Route_ID" was only used as a flag to differentiate between the pilot and formal experiment and was irrelevant, hence removed.

The *"Session date/time-stamp"* and *"Date/Time-stamps"* recorded by the ECU were recorded in GPS format, and converted to Coordinated Universal Time (UTC) to verify the time interval. For this data analysis, a simple time interval counter was required, this was recorded in the *"Time_1s"* (t=1s, ti= t +ti) and used in the dataset.

Feature selection in high-dimensional regression or classification frameworks is a challenging task in the presence of highly correlated predictors. Earlier studies on supervised ML using RF indicates that a high correlation between features will split importance as some of the data segmentation will be done on other similar features [3]. Therefore, we produced a feature correlation matrix (Figure 4) and removed highly correlated values to improve accuracy. The performance improves as there will be fewer data to analyse and reduces the risk of overfitting.

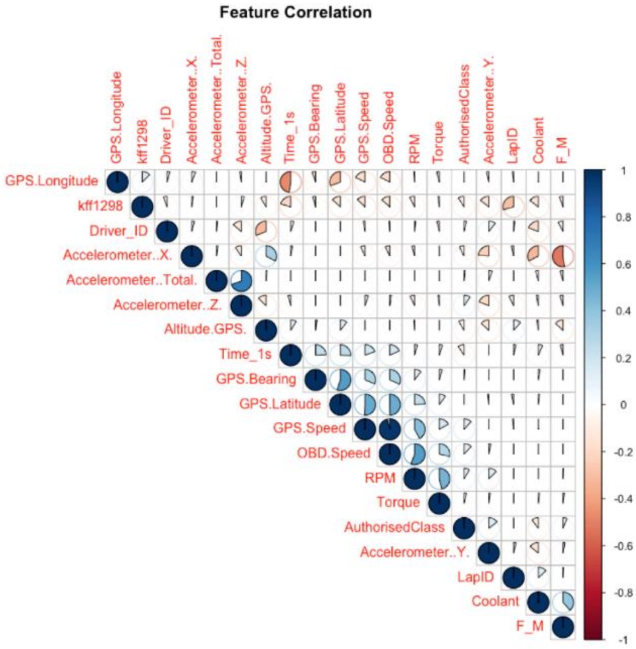

Figure 4. Feature Correlation Matrix.

Figure 4 demonstrates the feature correlation matrix for our dataset and illustrates the relationship between the different variables. For instance, *"GPS_Speed"* and *"OBD_Speed"* were highly correlated, with the correlation coefficient, r= 0.95. The values of both features were tested over a short distance at a constant speed. GPS accuracy can be impacted by several factors as the car travels depending on weather conditions, interference, and environmental factors, such as heavy tree-lined areas, tunnels or other similar conditions. Whilst GPS speed is known to be more accurate compared to many cars' speedometers, it was found that the GPS speed records are lower than what the speedometer shows by up to 5%. It is widely acknowledged that this may be by-design from the vehicle manufacturer, but other factors could be influential, for example, temperature, the pressure of the tyres and wheel size. Given the high correlation value in this example, and for this analysis with a focus on cybersecurity, it was paramount to record the car's speed consistently without the risk of signal jamming or loss. Therefore, the *"OBD_Speed"* feature was retained and the *"GPS_Speed"* was removed from the dataset.

The ranking of features was derived using WEKA's InfoGainAttribEval, a Ranker search method. The ranking of the features is shared in Table 1.

Table 1. Selected Features and their ranking as determined with InfoGainAttribEval.

#	Feature ID	Value	Range	Rank
1	Time_1s	[s]	[1-814]	8
2	Driver ID	[numeric]	[1-23]	1
3	Lap ID	[numeric]	[1-3]	18
4	GPS Longitude	[DD]	[-0.3078->-0.2916]	16
5	GPS Latitude	[DD]	[51.59->51.61]	14
6	GPS Speed	[km/h]	[0-98.26]	9
7	GPS Bearing	[◦]	[0.02-359.96]	15

8	Accel. Sensor (Total)	[g]	[-0.74399->3.10342]	17
9	Accel. Sensor (X)	[g]	[-3.581855->3.210907]	3
10	Accel. Sensor (Y)	[g]	[-2.313659->1.274006]	4
11	Accel. Sensor (Z)	[g]	[-/0.9317->2.7223]	11
12	Engine Coolant Temp.	[F]	[40-2-4.8]	5
13	Engine RPM	[rpm]	[0-3498]	10
14	Altitude	[m]	[-35->911.6]	6
15	OBD Speed	[km/h]	[0-99]	7
16	Engine Idling	[%]	[0-100.04]	2
17	Torque	[ft-lb]	[0-268.96]	12
18	F_M	[char]	[F/M]	13

4.4. Feature Distribution

The feature distribution was analysed between drivers with respect to the 3 laps of the experiment as shown in Figure 5. The *"Acceleration_Sensor_X"* (acceleration and braking) varied during the stages of the experiment, but we can observe a continuation of a unique driving pattern for individual participants. These values are changing in the environment throughout the experiment according to the participants' driving characteristics. There are conditions outside of the driver's control such as traffic volume, pedestrians, traffic lights and the weather that influence driving. The response to these conditions affects the rate of acceleration, consistency of manoeuvres, and other driving aspects which helps profiling participants according to their reactions. Figure 6 demonstrates an example of the participants' driving profiles based on the *"OBD_Speed".*

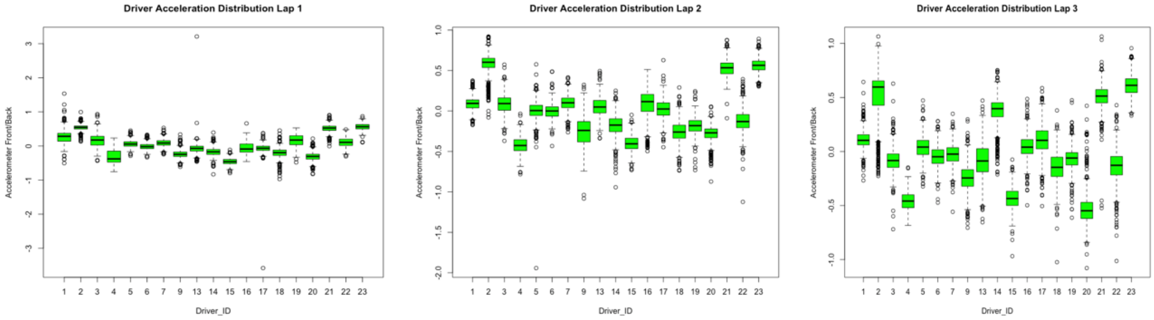

Figure 5. Box plot, feature distribution of Acceleration.

Further analysis utilising demographics recovered differences in driving characteristics between female and male drivers. For example, fluctuation in values within and between laps throughout the experiment was greater for male than female drivers. We have utilised the *"OBD_Speed"*, *"Torque"*, and *"RPM"* values to visualise these differences as shown in Figure 7 and Figure 8.

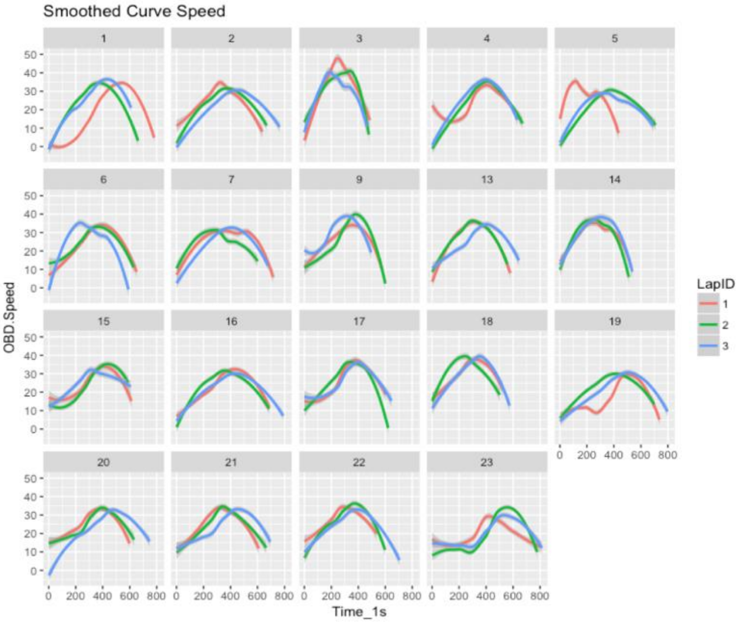

Figure 6. Time series of extracted speed data for all drivers.

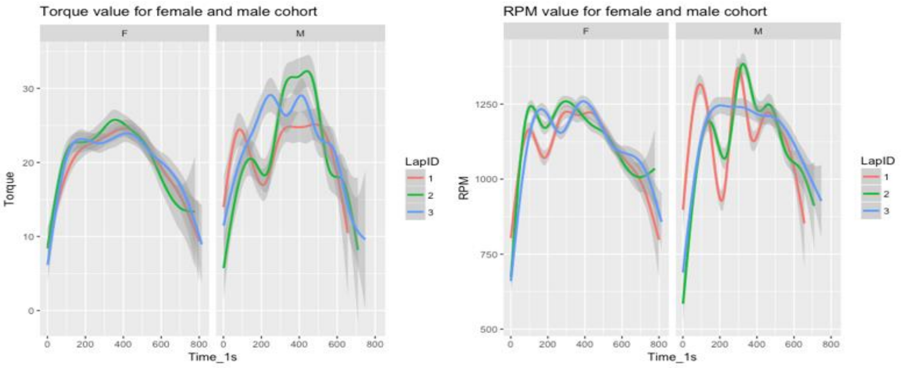

Figure 7. Time series pattern of torque and RPM for female and male drivers, 3 laps.

4.5. Exceptions

Driving exceptions such as excessive speeding, sharp deceleration, swerving, hitting the roadside curb or the wrong direction followed resulting in a wrong turn were recorded as shown in Figure 9.

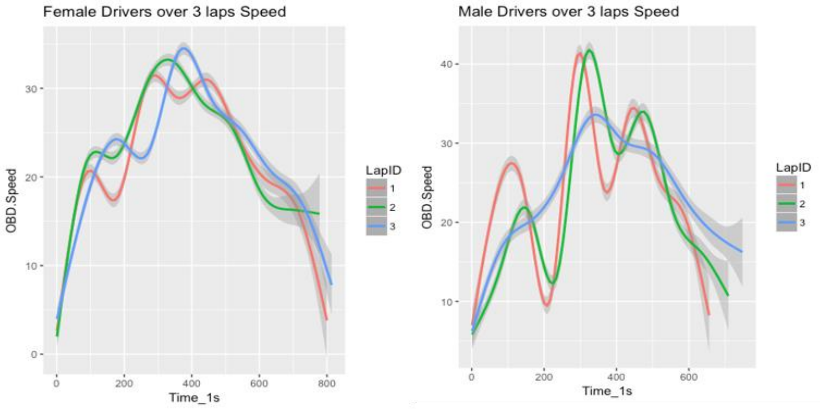

Figure 8. Time series pattern Female (left), Male (right) drivers, 3-laps speed characteristics.

4.6. Classification Algorithm Selection

The research question in this paper focuses on optimising driver profiling to minimise the window-size, the time required to verify a known driver, and investigates demographic clustering of drivers based on their driving behaviour. Therefore, an ML classifier was required that consistently provided accurate results and had a comparative baseline with other classifiers. Related research review (Section 2.3) informed the choice of the ML classifier based on consistency and accuracy of the results. To solve the driver classification, we utilise RF ensemble supervised ML, as it has shown to have consistently high performance in previous studies. This ML algorithm uses a training dataset to construct a set of decision trees detecting the driver's unique driving patterns classifying the driver.

4.7. Driver Classification Results

The R implementation of the algorithm was used on an Intel Core i7 3.3 GHz 16GB RAM MacBook Pro that processed the dataset. The dataset recorded the driving data every second and the driver classification was performed every second. The training and test datasets were prepared for the algorithm. The model was trained with 70% of the data and 30% was kept for testing the algorithm.

Using a high-level language like R was very effective, taking a 10s window-size to classify drivers based on the full trained-dataset containing 3 "authorised" and 16 "unauthorised" drivers. The selection is based on a typical family arrangement. As a method of RF model validation, the RF provides an unbiased estimate of the test set error internally by calculating the misclassification rate, the Out-of-Bag (OOB) score. The OOB score was calculated for the full driving track and individual driving laps. Table 2 shows the OOB error rate for 19 drivers. The variables were randomly sampled at each RF split by assigning different mtry values and results were examined. The full driving track, with 3 random variables and 15 trees (mtry =3/ntree=15), has an OOB error rate of 0.03%, 99.97% accuracy classifying 99/100 drivers. For individual driving laps 1 and 2 OOB error rate of 0.06% and 0.05%. Classification of drivers using a full driving track dataset, 3 random variables and 20 trees (mtry = 3/ntree=20) yielded 100%. The algorithm within the first 10s of driving had 99.52% confidence (mtry = 3/ntree=15) and 99.76% confidence (mtry = 3/ntree=20). The confidence increased with longer driving reaching confidence levels similar to the classification of the full train-dataset after 300s of driving (Table 2).

Furthermore, variable tuning was performed to improve the results without introducing significant implications to the performance or accuracy of the model. The model was trained based on the OOB with different mtry values to tune the forest as a function of mtry and evaluate the model's performance in terms of accuracy in relation to the model's computational complexity. RF was tuned for the full track searching for the optimal value of mtry with respect to the OOB error estimate. In this case, the mtry parameter was adjusted, a baseline of mtry=2 was created and tuned with results of its impact on the model examined. The mtry parameter was selected as part of this model optimisation due to its impact on the final accuracy and complexity of the model.

Classification of Female/Male groups show female drivers' classification with (mtry = 2/ntree=10) and (mtry = 3/ntree=15) has higher confidence than that of male drivers, also the accuracy of classification of female drivers is closer to the confidence rate of all drivers' classification across the entire dataset. The data shows higher OOB between the first 10s of the full track driver classification of female and male drivers as compared to the classification of all drivers over the same time for (mtry = 2/ntree=10), (mtry = 3/ntree=15) and (mtry = 3/ntree=20).

Furthermore, differences in female/male drivers show that it takes longer to gain similar consistency and confidence for the individual cohorts than for all drivers across the entire track, and 0% OOB is more prominent in the classification of male drivers for (mtry = 2/ntree=10), (mtry = 3/ntree=15) and (mtry = 3/ntree=20) than female over the first 10s of driving.

4.8. Variable Importance

The Mean DecreaseGini ranks the usefulness of the variables, by measuring the gain of purity by splits for a given variable with the lowest Gini Index. The relative importance of various features in the dataset across the full track and the individual laps correlate. The importance of each variable with the Mean DecreaseGini showing is demonstrated in Figure 10, they show how each variable contributes to the purity on each node in a tree. Whilst maintaining the same number of trees but reducing the number of variables by 1 has increased the OOB error rate by 0.13% for the full track, the OOB error rate increased by 0.34% on average for the individual laps given the same conditions. Therefore, any attempt to reduce computational time by reducing the number of variables will affect our model's accuracy.

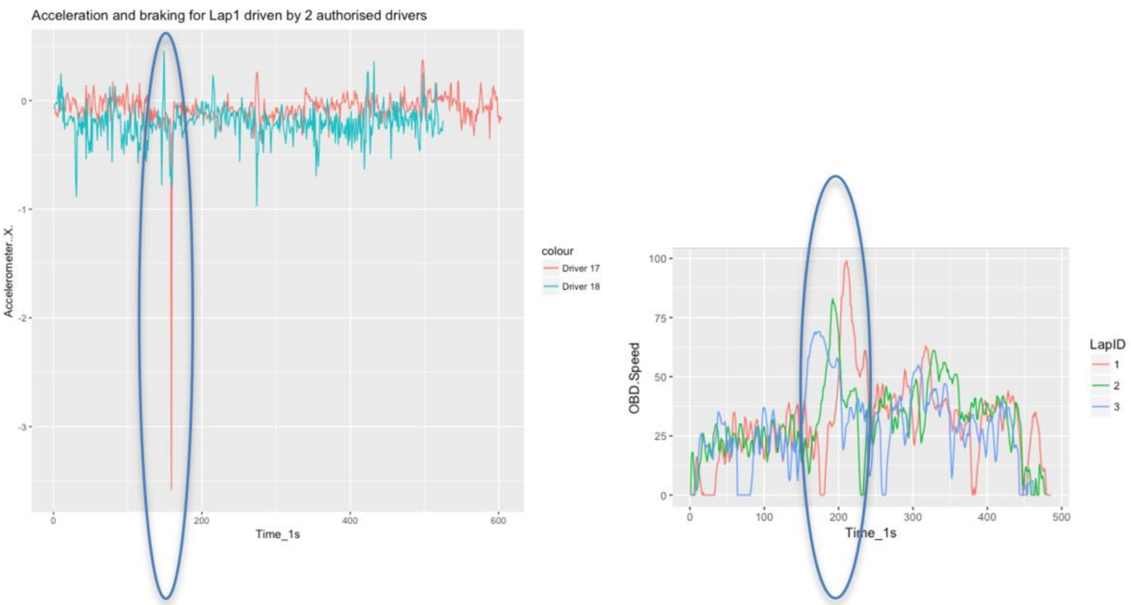

Figure 9. Time series patterns; emergency break exception (left); and speeding exception (right) during lap 1 but also lap 2 and 3 during the 40mph zone (graph in km/h).

Table 2. RF model OOB and Test Dataset Confusion Matrix - a snippet of data.

	OOB mtry=2, ntree=10, (%)	p-value, Acc>NIR	Confusion Matrix (%)	OOB mtry=3, ntree=15, (%)	p-value, Acc>NIR	Confusion Matrix (%)	OOB mtry=3, ntree=20, (%)	p-value, Acc>NIR	Confusion Matrix (%)
All Drivers									
full track	0.70	< 2.2e-16	99.97	0.03	< 2.2e-16	99.97	0.00	< 2.2e-16	100
lap1	0.58	< 2.2e-16	100	0.06	< 2.2e-16	100	0.02	< 2.2e-16	100
lap2	0.35	< 2.2e-16	100	0.05	< 2.2e-16	100	0.00	< 2.2e-16	100
lap3	0.48	< 2.2e-16	99.97	0.00	< 2.2e-16	100	0.00	< 2.2e-16	100
first 10s	2.20	< 2.2e-16	97.74	0.48	1.89E-13	100	0.24	1.89E-13	100
first 20s	1.95	< 2.2e-16	99.15	1.33	< 2.2e-16	100	0.85	< 2.2e-16	100
First 90s	1.38	< 2.2e-16	99.94	0.24	< 2.2e-16	100	0.08	< 2.2e-16	100
First 300s	0.38	< 2.2e-16	100	0.02	< 2.2e-16	100	0.01	< 2.2e-16	100
Female									
Full track	0.27	< 2.2e-16	99.96	0.02	< 2.2e-16	99.95	0.01	< 2.2e-16	100
Full track 10s	3.35	3.2480E-09	100	1.10	3.25E-09	100	1.10	3.25E-09	100
Lap1 first 10s	0.98	< 2.2e-16	99.79	3.30	0.0015	100	0.00	0.0015	100
Lap2 first 10s	2.22	0.0015	100	7.69	0.0120	97.44	3.30	0.0120	97.44
Lap3 first 10s	0.65	< 2.2e-16	99.96	3.30	0.0120	97.44	0.15	0.0015	100
Male									
Full track	0.54	< 2.2e-16	99.89	0.03	< 2.2e-16	99.89	0.00	< 2.2e-16	100
Full track 10s	5.00	0.0007	98.30	1.43	5.82E-05	100	2.14	0.0007	98.33
Lap1 first 10s	4.76	0.0376	100	0.24	< 2.2e-16	99.95	0.12	< 2.2e-16	99.95
Lap2 first 10s	4.08	0.0393	100	2.04	0.0393	100	0.06	< 2.2e-16	100
Lap3 first 10s	6.38	0.0393	100	0.00	0.0393	100	0.00	0.0393	100

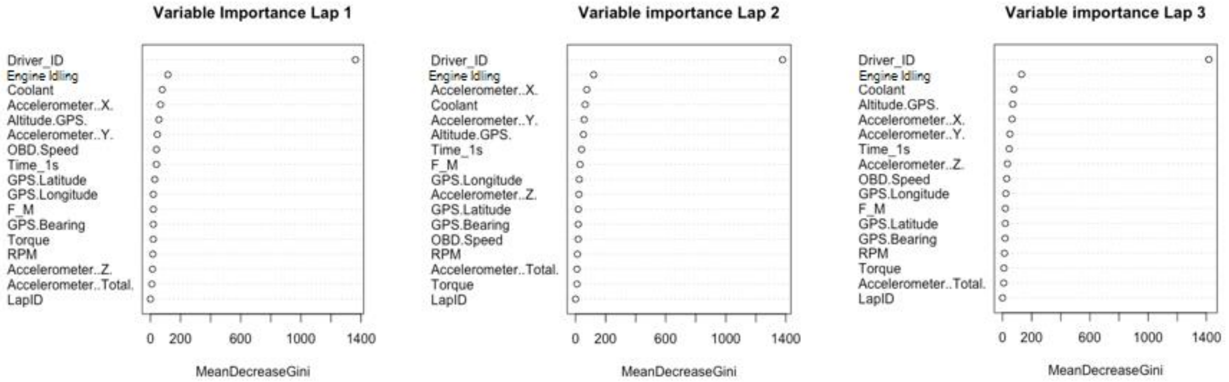

Figure 10. Variable Importance Chart for by Individual Laps.

4.9. Model Accuracy and Predictions

To demonstrate the overall quantitative comparison of the classification model for the two categories, “Yes” (Authorised Drivers) and, “No” (Unauthorised Drivers), a Confusion Matrix was created from the test dataset based on the 95% confidence level, which yielded a summary of prediction results to better understand the performance of the classification model for different values of mtry and ntree. The overall model accuracy based on 95% confidence level improved from 96.7% (mtry= 2/ntree=10) to 99.97% (mtry= 3/ntree=15) accuracy in correctly identifying the authorised drivers when optimising the model by altering the mtry/ntree values as shown in Figure 11 and Tables 3-4.

Table 3. Confusion Matrix 95% Confidence Level for the overall classifier performance with different model complexity.

Confusion Matric Confidence Level 0.95 (95%), threshold 0.8					
mtry2, ntree10	**Drivers**	**Unauthorised**	**Authorised**	**Class Error**	**% correct**
	Unauthorised	21647	65	0.002	99.7006
	Authorised	111	3283	0.032	96.7295
mtry3, ntree15	**Drivers**	**Unauthorised**	**Authorised**	**Class Error**	**% correct**
	Unauthorised	21903	2	9.1303360000E-05	99.9909
	Authorised	1	3425	2.9188560000E-04	99.9708
mtry3, ntree20	**Drivers**	**Unauthorised**	**Authorised**	**Class Error**	**% correct**
	Unauthorised	21927	0	0.00000	100.0000
	Authorised	1	3428	0.00029	99.9708

Furthermore, increasing the trees adds additional computational demand and complexity without significantly improving the model accuracy. The overall model accuracy of female drivers is stronger at 0.27% OOB compared to 0.54% OOB error rate for male drivers.

Table 4. OOB overall and for female/male classifier performance with different model complexity at different reference points across the track.

Participants	Track	OOB mtry=2, ntree=10	OOB mtry=3, ntree=15	OOB mtry=3, ntree=20
All	Overall	0.70%	0.03%	0.00%
	lap1	0.58%	0.06%	0.02%
	lap2	0.35%	0.05%	0.00%
	lap3	0.48%	0.00%	0.00%
	Full track 10s	2.20%	0.48%	0.24%
Female	Overall	0.27%	0.02%	0%
	Full track 10s	3.35%	1.10%	1.10%
Male	Overall	0.54%	0.03%	0%
	Full track 10s	5.00%	1.43%	2.14%

The model accuracy improved at a lower rate by increasing the computational complexity for the overall dataset, but that rate of improvement of the model accuracy when analysing the subgroups using smaller datasets varied. The prediction accuracy of individual female drivers is marginally weaker at 99.96%, but stronger than male drivers at 99.89%. This was calculated using a binomial test that demonstrated the p-value greater than the No-Information-Rate of 0.8648.

Cross-validation was applied using the test dataset to fit 10 RF models to a subset of the training dataset calculating the mean, across the entire track being 99.99%, results in line with the binomial test. Receiver Operating Characteristics (ROC) curve could also be produced for the prediction as a measure of performance for a binary classifier.

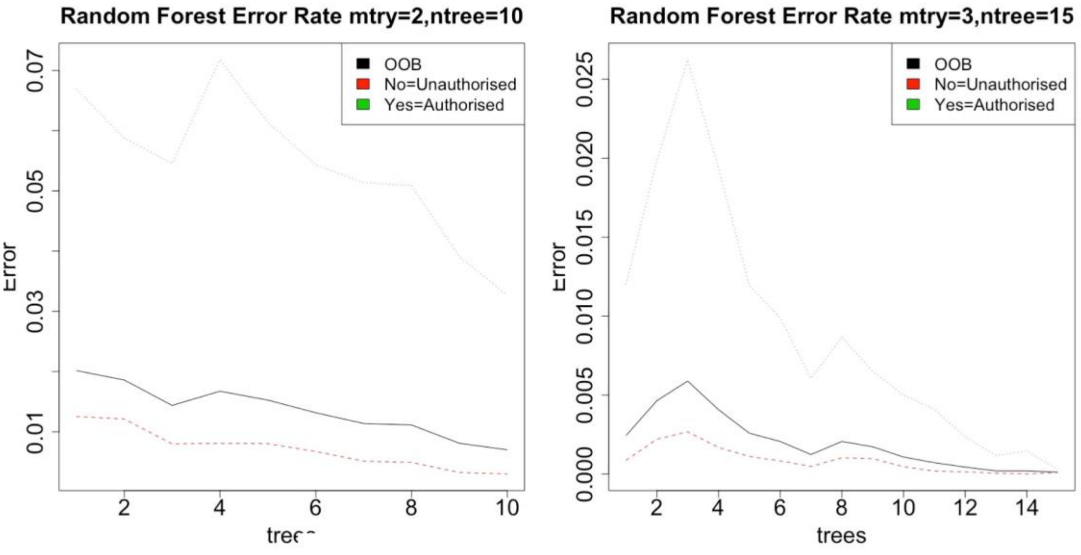

Figure 11. RF model Error Rate data full track and individual laps.

4.10. Sliding Window

Anomaly detection was investigated with testing carried out every second. It was possible to derive classification accuracy for any window size and record the speed of the driver at any given point in time too, this was based on data generated in Near Real-Time (NRT) directly from the car sensors as shown in Figure 9.

5. DISCUSSION

The baseline in this study is set at level 2, defined as partial automation by the Society of Automotive Engineers (SAE), where human drivers monitor and control the environment. Inconsistency detection at the outset or during the driving session is a key element to proactively address cyber-threats [14]. Furthermore, grouping similar driving habits together yield driving patterns that are reflective of drivers or clusters of drivers based on criteria, which is an important step in driver classification [11]. We have addressed the research question via an experiment within the context of connected cars, more precisely through a driver-centric approach based on profiling as demonstrated in the analysis of our results. The choice of techniques to extract driving data was scrutinised. Several key facets were taken into consideration such as the simplicity of the process, accuracy, cost, accessibility, and applicability. For instance, the OBD2 adapter that was utilised for our data extraction phase is independent of the car manufacturer with a variety of applications to export data to a range of file formats. The results demonstrate the feasibility of the drivers' classification in modern cars generated from in-car sensors that are based on general, not the manufacturer or otherwise proprietary data, or data extraction technique. Therefore, our approach can be utilised across a range of vehicles fitted with an OBD2 port using a Bluetooth or hard-wired connection to facilitate data collection and processing. The built-in data-enabled Subscriber Identification Module (SIM) card feature can be utilised for a remote secondary processing and storage in CC and a third-party data-enabled mobile device across a range of other vehicles.

If the process of driver identification is required for an NRT application as part of a layered security approach, such as a proactive countermeasure against carjacking, then the threat model for such a system has to account for connectivity issues if data processing is located remotely. The connection can be disrupted or jammed, resulting in data loss or corruption, hence the attack surface remains exposed. However, using a Bluetooth enabled OBD2 scanning tool, the current computational capabilities of relatively small devices make it reasonable and important to implement data storage and analysis of the extracted data locally using a black-box type device for reliability, robustness and legal forensics [24]. That said, extending the model with a remote secondary processing and storage facility utilising an in-car data-enabled SIM card or third-party data-enabled mobile device helps to mitigate against threats related to data integrity, attribution and denial of service. Therefore, a single measure of protection is insufficient, and several levels of security controls are needed for effective layered security defence [25] including but not limited to host and network controls, physical and data-level security. Within the context of layered security defence, driver profiling and understanding driving behaviour changes is key to prevent or detect attacks such as remote hijacking or where the driver could be initially authorised such as drive-related attacks where other security controls could be ineffective.

The pilot experiment played a critical role to improve our methodology, as it helped to refine the route to capture sufficient data by increasing the number of laps to three. Other methods, such as the dashcam and tablet, were tested alongside the OBD2 method. Furthermore, our observations from the pilot study helped explain or support our findings from the dataset. For instance, it was noted that male drivers were less hesitant on junctions and in giving free-way to oncoming traffic in narrow pass areas than female drivers, whilst female drivers took significantly fewer chances and waited until both left and right approaching traffic was clear or they were clearly given way by other drivers. Also, female drivers were more concerned about being in the correct lane on the roundabout approach than male drivers. The dataset reflected this by showing how the fluctuation of data was consistently higher for male drivers and the time to complete the laps was considerably shorter; Figure 6, Figure 7 and Figure 8 illustrate the time series driving characteristics of all female and male drivers and a subgroup of randomly selected female and male participants.

This study investigated the best working combination of parameters, characteristics, and features to increase the accuracy of driver profiling. One of the objectives of this study was to investigate the relevance, duplicity, and correlation of extracted features. To optimise a model, any data that is non-reflective of suitable driving characteristics should be disregarded, hence the significance of this step which has concluded and ranked 18 features in Table 1. Additionally, the dataset shows evidence that identifying demographic differences between different groups of drivers is achievable and helps to better understand the user (driver). This gives an opportunity to customise the car's user experience based on driving patterns generated by in-car sensors but also classify individuals in a small window size starting at only 10s of driving. For the cohort of drivers in our sample, OOB error rates were calculated. Results show that optimisation and variable tuning keeps the OOB error consistently small. Overall, accuracy increases by time and bigger datasets. We also analysed exceptions as shown in Figure 9 because the detection of such anomalies from several cars in accordance to a specific location could inform future innovations in smart cities to work towards a data-informed approach for smart urban transport using the intelligence from the harvested data to create "smart urban roads". For instance, this could help to profile dangerous junctions, turns or faults in a road surface [19] but also to enable safer traffic management by identifying inappropriate speeds around schools, hospitals or other priority areas.

Analysing data from in-car sensors triggers legal and ethical discussions. It can be argued that it could increase the attack landscape while GPS values introduce privacy concerns [14]. However, GPS tracking can also be a key factor to enable a proactive approach through real-time predictive monitoring to resist cyber-hijacking [10]. Instead of disregarding GPS information, regulation and legal context must drive its use such as the General Data Protection Regulation (GDPR) in Europe.

We acknowledge that new disruptive technologies come with risk and that in-car sensor data could recover non-anonymised personal data or data that can be attributed to individuals such as trip related data [1]. It is not clear how this data will be handled and processed, or how the driver's consent will be taken before creating logs related to driving behaviour. Although our study relies on anonymised data the underlying principle of CC deems personal data present within the realm of the vehicle. Therefore, further research in this area is required to protect personally identifiable information through layered security defence, standards, and regulations. It is critical to design solutions where the benefit outweighs risks [25] in addition to conducting further research to perform a risk assessment and evaluate new compliance and regulatory models.

The participants drove a mixture of different vehicles; nevertheless, driving someone else's vehicle could result in unconscious bias, which could influence one's driving style either by additional care taken or through careless driving. Therefore, the influence on results is similar to an unauthorised driver who does not regularly drive the vehicle. In common with self-reported surveys, there can be social-desirability bias when filling in our questionnaire because drivers may have exaggerated or underestimated certain answers. However, throughout this paper, responses from participants were treated as genuine. Further, we have profiled the drivers in our sample and produced results that can suggest a hypothesis for further study, but we acknowledge that generalisation is not possible at this stage and would require other studies to systematically report similar findings in the literature utilising a wider pool of vehicles. Nonetheless, expanding the questionnaire to cover questions such as the type of roads driven, and the main purpose of the vehicle use, could enable deeper analysis between various demographics and driving behaviours. Adding questions on how tech-savvy the drivers are and their use of advanced car features could help address future research questions including the reception of and driving behaviours when combined with higher levels of automation, which will penetrate the consumer market and bring about unprecedented change in the automotive industry.

6. CONCLUSION

Driver Profiling can be utilised in a wide range of applications to support their Authentication, Authorization, and Accounting (AAA) model. This process can be useful where car hijacking can be detected very early in a driving session. In some scenarios, a driver could get their authorisation revoked with immediate effect due to malicious behaviour. Events such as a driver intentionally moving fast towards a crowd could be profiled to mitigate risk to both life and property hence the added value of this research area. We have demonstrated how analysing driving behaviour contributes to driver identification, which could then facilitate incident response and enhance users' experience. We have presented the means to optimise driver detection based on data extracted from in-car sensors and GPS data. Unlike most cited literature, we have produced a car-based dataset which has helped us study driving inconsistencies that crop up during normal driving. In addition, the attention of the driver and their reactions are more risk-averse knowing that street driving has real consequences, unlike a simulated session. Our optimised model ranks the top features we can use to reflect unique driving behaviour and gives evidence that accuracy can go up to 99.7% (95% confidence) given only a 10s driving window. Additionally, drivers can be grouped and there is evidence that ML can cluster drivers to tell us information such as their gender. While our sample (n=19) is larger than other studies found in our literature review, we acknowledge that generalisation is difficult at this stage, and more work is needed to cover various types of people, car models, and environments. However, we argue that our results are linked with existing in-car sensors and that further profiling opportunities would be enabled with the introduction of more IoT sensors to cars.

References

[1] C. Gosman, T. Cornea, C. Dobre, F. Pop, and A. Castiglione, “Controlling and filtering users data in intelligent transportation system,” *Future Generation Computer Systems,* vol. 78, pp. 807-816, 2018, https://doi.org/10.1016/j.future.2016.12.014.

[2] G. Castignani, T. Derrmann, R. Frank, and T. Engel, “Driver behavior profiling using smartphones: A low-cost platform for driver monitoring,” *IEEE Intelligent Transportation Systems Magazine,* vol. 7, no. 1, pp. 91-102, 2015, https://doi.org/10.1109/MITS.2014.2328673.

[3] B. I. Kwak, J. Woo, and H. K. Kim, "Know your master: Driver profiling-based anti-theft method." pp. 211-218, https://doi.org/10.1109/PST.2016.7906929.

[4] S. Rizvi, J. Willet, D. Perino, S. Marasco, and C. Condo, “A threat to vehicular cyber security and the urgency for correction,” *Procedia computer science,* vol. 114, pp. 100-105, 2017, https://doi.org/10.1016/j.procs.2017.09.021.

[5] D. Hallac, A. Sharang, R. Stahlmann, A. Lamprecht, M. Huber, M. Roehder, R. Sosič, and J. Leskovec, "Driver identification using automobile sensor data from a single turn." pp. 953-958, https://doi.org/10.1109/ITSC.2016.7795670.

[6] L. Chen, and P. Wang, "Risk factor analysis of traffic accident for different age group based on adaptive boosting." pp. 812-817, https://doi.org/10.1109/ICTIS.2017.8047861.

[7] M. Sawataishi, K. Sato, H. Madokoro, M. Ito, and S. Kadowaki, "Driver internal state estimative model for distracted state detection." pp. 2504-2509, https://doi.org/10.1109/SMC.2017.8123000.

[8] L.-W. Chen, K.-Z. Syue, and Y.-C. Tseng, "A vehicular surveillance and sensing system for car security and tracking applications." pp. 426-427, https://doi.org/10.1145/1791212.1791288.

[9] Z. Liu, A. Zhang, and S. Li, "Vehicle anti-theft tracking system based on Internet of things." pp. 48-52, https://doi.org/10.1109/ICVES.2013.6619601.

[10] H. al-Khateeb, G. Epiphaniou, A. Reviczky, P. Karadimas, and H. Heidari, “Proactive Threat Detection for Connected Cars Using Recursive Bayesian Estimation,” *IEEE Sensors Journal,* vol. 18, no. 12, pp. 4822-4831, 2018, https://doi.org/10.1109/JSEN.2017.2782751.

[11] M. Salemi, "Authenticating drivers based on driving behavior," Rutgers University-Graduate School-New Brunswick, 2015, https://doi.org/10.7282/T3QJ7K4M.

[12] T. Wakita, K. Ozawa, C. Miyajima, K. Igarashi, K. Itou, K. Takeda, and F. Itakura, “Driver identification using driving behavior signals,” *IEICE TRANSACTIONS on Information and Systems,* vol. 89, no. 3, pp. 1188-1194, 2006, https://doi.org/10.1093/ietisy/e89-d.3.1188.

[13] X. Zhang, X. Zhao, and J. Rong, “A study of individual characteristics of driving behavior based on hidden markov model,” *Sensors & Transducers,* vol. 167, no. 3, pp. 194, 2014, Retrieved from:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.641.3718&rep=rep1&type=pdf

[14] A. Burton, T. Parikh, S. Mascarenhas, J. Zhang, J. Voris, N. S. Artan, and W. Li, "Driver identification and authentication with active behavior modeling." pp. 388-393, https://doi.org/10.1109/CNSM.2016.7818453.

[15] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura, "Multimedia data collection of in-car speech communication," Retrieved from:https://isca-speech.org/archive/archive_papers/eurospeech_2001/e01_2027.pdf

[16] J. Engelbrecht, M. J. Booysen, G.-J. van Rooyen, and F. J. Bruwer, “Survey of smartphone-based sensing in vehicles for intelligent transportation system applications,” *IET Intelligent Transport Systems,* vol. 9, no. 10, pp. 924-935, 2015, https://doi.org/10.1049/iet-its.2014.0248.

[17] D. A. Johnson, and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform." pp. 1609-1615, https://doi.org/10.1109/ITSC.2011.6083078.

[18] S. Ezzini, I. Berrada, and M. Ghogho, "Who is behind the wheel? Driver identification and fingerprinting," *Journal of Big Data,* vol. 5, no. 1, pp. 9, 2018, https://doi.org/10.1186/s40537-018-0118-7.

[19] G. A. M. Meiring, and H. C. Myburgh, "A review of intelligent driving style analysis systems and related artificial intelligence algorithms," *Sensors,* vol. 15, no. 12, pp. 30653-30682, 2015, https://doi.org/10.3390/s151229822.

[20] R. Kalsoom, and Z. Halim, "Clustering the driving features based on data streams." pp. 89-94, https://doi.org/10.1109/INMIC.2013.6731330.

[21] A. Aljaafreh, N. Alshabatat, and M. S. N. Al-Din, "Driving style recognition using fuzzy logic." pp. 460-463, https://doi.org/10.1109/ICVES.2012.6294318.

[22] M. Van Ly, S. Martin, and M. M. Trivedi, "Driver classification and driving style recognition using inertial sensors." pp. 1040-1045, https://doi.org/10.1109/IVS.2013.6629603.

[23] H. Hooft van Huysduynen, J. Terken, J.-b. Martens, and B. Eggen, *Measuring driving styles: a validation of the multidimensional driving style inventory*, 2015, https://doi.org/10.1145/2799250.2799266.

[24] H. Al-Khateeb, G. Epiphaniou, and H. Daly, "Blockchain for Modern Digital Forensics: The Chain-of-Custody as a Distributed Ledger," *Blockchain and Clinical Trial: Securing Patient Data*, H. Jahankhani, S. Kendzierskyj, A. Jamal, G. Epiphaniou and H. Al-Khateeb, eds., pp. 149-168, Cham: Springer International Publishing, 2019, https://doi.org/10.1007/978-3-030-11289-9_7.

[25] G. Ahmadi-Assalemi, H. M. al-Khateeb, G. Epiphaniou, J. Cosson, H. Jahankhani, and P. Pillai, "Federated Blockchain-Based Tracking and Liability Attribution Framework for Employees and Cyber-Physical Objects in a Smart Workplace." pp. 1-9, https://doi.org/10.1109/ICGS3.2019.8688297.

Gabriela Ahmadi-Assalemi is a Ph.D. researcher at the Wolverhampton Cyber Research Institute (WCRI), University of Wolverhampton. She is a Deputy Chief Information Security officer at the University of Cambridge. Gabriela's research interests include cyber-physical systems, driver profiling through behavioural modelling in connected cars and data acquisition for digital forensics.

Haider M. al-Khateeb received the Ph.D. in cyber security from the University of Bedfordshire, U.K., in 2011. He is a Senior Lecturer in Cyber Security, a Consultant, and a Senior Fellow of the Higher Education Academy (SFHEA). His research interests include User Authentication Methods, Distributed Digital Forensics, Remote Incident Response, and Security for Cyber-Physical Systems (CPS).

Carsten Maple received the Ph.D. from the University of Leicester, U.K., in 1998. He is currently a Professor of Cyber Systems Engineering with WMG's Cyber Security Centre. He is the Director of research in Cyber Security working with organizations in key sectors such as manufacturing, healthcare, and the broader public sector to address the challenges presented by today's cyber environment.

Gregory Epiphaniou received the Ph.D. degree from the University of Bedfordshire, U.K., in 2010. He is currently an Associate Professor of security engineering at the University of Warwick. His role involves bid support, applied research, publications and teaching. He led and contributed to several research projects funded by EPSRC, IUK and local authorities totalling over £3M.

Mohammad Hammoudeh received the Ph.D. degree from the University of Wolverhampton, U.K., in 2008. He is currently the Head of the MMU IoT Laboratory and a Senior Lecturer in Computer Networks and Security with the School of Computing, Math, and Digital Technology, Manchester Metropolitan University, Manchester, U.K.

Hamid Jahankhani received his Ph.D. from the Queen Mary College, University of London. His principal research area has been in the field of cyber security, information security and digital forensics. In partnership with the key industrial sectors, he has examined and established several innovative research projects that are of direct relevance to the needs of UK and European industries.

Prashant Pillai received the Ph.D. degree from the University of Bradford, U.K., in 2007. He is currently the Director of the Wolverhampton Cyber Research Institute (WCRI), U.K., which consists of 24 academics. He has more than 15 years of research experience and specializes in the area of communication protocols and cyber security.