



Published in final edited form as:

Transp Res Interdiscip Perspect. 2024 ; 25: . doi:10.1016/j.trip.2024.101109.

Validation of a smartphone telematics algorithm for classifying driver trips

Jeffrey P. Ebert, PhD^{1,2}, Ruiying A. Xiong, MS^{1,2}, Arjun Patel, BA^{1,2}, Dina Abdel-Rahman, BS^{1,2}, Catherine C. McDonald, PhD, RN, FAAN^{1,2,3,4}, M. Kit Delgado, MD, MS^{1,2}

¹Perelman School of Medicine, University of Pennsylvania

²Penn Injury Science Center

³School of Nursing, University of Pennsylvania

⁴Center for Injury Research and Prevention at the Children's Hospital of Philadelphia

Abstract

This study assessed the accuracy of a smartphone telematics algorithm that classifies car trips as driver or non-driver. Participants' trips were measured for 4 weeks by *Way to Drive*, a research telematics application that uses the same data algorithms as leading auto-insurance companies. At the end of each week, participants completed a survey prompting them to review trips within the app and report time and nature of any misclassified trips. Overall accuracy of driver vs. non-driver classification was high (96.5%, SD = 5.1%). Sensitivity, the percentage of actual driver trips classified as such, was also high (97.5%, SD = 4.6%). Specificity, the percentage of non-driver trips classified as such, was slightly lower and more variable (91.2%, SD = 14.8%). The algorithm's accuracy was generally robust to a variety of phone characteristics, vehicle features, and driving habits.

Keywords

Driver classification algorithm; mobile telematics application; driver risk score; naturalistic study; accuracy; usage-based insurance

Introduction

Smartphone telematics applications are increasingly used to measure risky driving behaviors like speeding and phone use while driving. Many usage-based insurance (UBI) programs rely on these apps to create driver risk profiles and price policy premiums accordingly (Kasperowicz, 2022; Reimers & Shiller, 2019). Safety researchers have also deployed telematics apps in observational studies of risky driving behaviors and in controlled field experiments to improve driver safety (Choudhary et al., 2022; Ebert et al., 2022; Stevenson

Corresponding Author: Jeffrey P. Ebert, PhD, Perelman School of Medicine, University of Pennsylvania, Center for Health Care Innovation, 3600 Civic Center Blvd, 8th Floor, Philadelphia, PA, US 19104-5127, jeffrey.ebert@pennmedicine.upenn.edu.

Declaration of interest

None. This research was conducted independently with no editorial input from TrueMotion, Inc./Cambridge Mobile Telematics.

et al., 2021). Telematics apps, which run in the background of a user's smartphone and require minimal user effort once installed, have cost and implementation advantages over traditional plug-in and on-board telematics devices—enabling driver safety programs to scale effectively.

A potential drawback is that telematics apps measure trip activity regardless of whether the smartphone user was driving. To correctly attribute risky driving behaviors to the user, an app must accurately predict whether they were driving on a given trip. This prediction problem is typically solved with deep-learning algorithms that recognize patterns in smartphone sensor (e.g., accelerometer) data or extracted features (e.g., braking) when the user is driving (Park, 2018). A recent review of studies examining the accuracy of driver identification algorithms found that the best ones achieve accuracy of 85% or greater (Zhao et al., 2022). However, these studies looked at algorithms trained on rich on-board sensor data and tasked with identifying which driver out of a small sample (2 to 25 drivers) was driving (Abdenmour et al., 2021; Cai et al., 2018; Choi et al., 2022; Girma et al., 2019; Jeong et al., 2018; Ravi et al., 2022; Van Ly et al., 2013). Research examining the accuracy of algorithms that rely solely on smartphone sensor data to predict whether a given user was driving would advance the field of traffic safety.

The present study set out to test the accuracy of a driver/non-driver classification algorithm widely used in smartphone telematics apps. The primary outcome was overall accuracy. Secondary outcomes were sensitivity (proportion of actual driver trips classified as such) and specificity (non-driver trips classified as such). We also investigated the impact of inaccurate classifications on drivers' risk profiles, as well as smartphone and environmental characteristics that could plausibly affect accuracy.

Materials and Methods

This was a prospective, preregistered field study ([NCT05422586](#)) recruiting from a sample who had previously participated in a national trial of behavioral interventions to encourage safer driving in a UBI program ([NCT04587609](#)) and indicated openness to future research. Recruitment was conducted by email the week of July 11, 2022.

Participants consented and completed an intake survey about their smartphone, vehicle features, and driving habits in Qualtrics. To reduce participant burden, demographic characteristics from the parent trial were used to describe this sample. Participants were invited by text and email to install the University of Pennsylvania's *Way to Drive* app, which measured vehicle trips for 4 weeks. *Way to Drive* is a white-label version of a telematics app developed by TrueMotion, Inc. (now Cambridge Mobile Telematics). It uses the same driver classification and risky behavior detection algorithms (v2.3.0) as leading UBI apps. Driver classification is accomplished with a deep-learning algorithm that recognizes driving patterns characteristic of the user (Park, 2018).

On the Sunday morning after each week, participants were sent a survey to review trips within the app. Tapping on a given trip displayed its start time, a map of the route taken, and

the driver/non-driver classification. For each day of the week (reverse-chronological order), the survey asked about:

- Number of trips the app detected
- Number of trips the app failed to detect
- Whether any trips were misclassified
- Start time and nature (driver as non-driver, non-driver as driver) of misclassified trips

Participants had until midnight the next day (Monday) to complete each survey. Participants were compensated up to \$50 in Amazon gift codes: \$10 for completing the intake survey and activating the app, and \$10 for each weekly survey.

Analysis

Way to Drive JSON trip files were first processed using an R script that summarized key driving metrics and the algorithm's driver/non-driver classification. For each participant, total trips classified as driver and non-driver were computed. Because these totals were used with participant reports of misclassifications to derive true positives and negatives, we only included days for which survey responses were available. For each participant, total trips reported to be misclassified as driver and non-driver were computed. Primary and secondary outcome variables were derived according to Table 1.

If participants unreliably reported false positives and negatives, this could inflate our estimate of accuracy. We derived an objective index of reliability for each participant by computing the Pearson's r between trips detected and reported trips detected across the 28-day measurement period. We also regressed app accuracy onto r^2 to examine how app accuracy varied by participant reliability.

We sought to quantify the impact of misclassifications on driver risk profiles. For each participant, we computed an unadjusted driver risk score by taking the duration-weighted average risk score¹ of all trips classified driver. Then, reportedly misclassified trips were matched to processed trips using trip start time (85.5% matched successfully). These trips were reclassified according to participants' reports, and an adjusted risk score was computed. Adjusted and unadjusted scores were compared using a paired t -test and Pearson's r .

Finally, we investigated whether app accuracy was associated with reported *smartphone characteristics* (iPhone vs. Android, age, quality of cellular data, battery life, frequency of "low power" mode, frequency of cellular data being turned off), *vehicle features* (presence of phone mount, frequency of using phone mount, dashboard touchscreen connectivity), or *driving habits* (percent of trips as driver, variability of trip routes and destinations, frequency of being in slow-moving traffic, frequency of taking public transit, frequency of

¹Trip-level risk scores are computed by Cambridge Mobile Telematics based on phone use, hard braking, mileage, and time of day while driving. Scores range from 0 (most risky) to 100 (least risky); UBI discounts may be warranted for scores of 60 and above (Mosley, 2016)

letting passenger use participant's phone when driving, rural vs. suburban vs. urban driving environment) by running a multiple regression for each of these categories.

Results

Of the 400 solicited, we arrived at an analytic sample of 57 participants (Fig. 1). Demographic characteristics are shown in Table 2.

Self-report data about whether there was a misclassification were available for a mean of 24.9 ($SD = 3.5$) of the 28 days. On these days, the app classified a mean of 120.9 ($SD = 69.8$) trips as driver and 26.5 ($SD = 19.5$) as non-driver. Participants reported a small number of false positives ($M = 2.5$, $SD = 5.6$) and false negatives ($M = 2.3$, $SD = 4.4$). Subtracting false positives from trips classified as driver yielded a mean of 118.4 ($SD = 69.8$) true positives. Subtracting false negatives from trips classified as non-driver yielded a mean of 24.2 ($SD = 18.1$) true negatives. The primary and secondary outcomes are shown in Fig. 2. Overall accuracy was high ($M = 96.5\%$, $SD = 5.1\%$). Sensitivity ($M = 97.5\%$, $SD = 4.6\%$) was greater and less variable than specificity ($M = 91.2\%$, $SD = 14.8\%$).

Participants as a group were high on our reliability index (M Pearson's $r = 0.95$). When app accuracy was regressed onto r^2 , the coefficient for r^2 was small (-0.034) and nonsignificant ($P = 0.292$). Adding this coefficient to the intercept (0.996) projected app accuracy to be 0.962 for a hypothetical participant with perfect reliability, in line with mean accuracy for the sample.

Unadjusted ($M = 74.5$, $SD = 5.1$) and adjusted ($M = 74.4$, $SD = 5.2$) driver risk scores were very similar ($P = 0.166$). Across participants the correlation between unadjusted and adjusted risk scores was essentially perfect ($r = 1.00$). These results suggest that occasional trip misclassifications do not invalidate drivers' overall risk profiles.

No vehicle features or driving habits significantly predicted accuracy (all P s $> .26$). Having a phone in "low power" mode while driving 1–2 days in the prior 2 weeks was significantly associated with lower app accuracy (5.3 percentage points, $P < .005$), and having a phone at least 3 years old was marginally associated with lower accuracy (5.2 percentage points, $P = .081$). No other smartphone characteristics predicted accuracy (all P s $> .12$).

Discussion

The driver/non-driver classification algorithm used by *Way to Drive* and other smartphone telematics apps achieved accuracy comparable to the best-performing driver identification algorithms that take on-board sensor data as input (Zhao et al., 2022). Accuracy was lower for phones that occasionally went into low power mode while driving; otherwise, the algorithm was robust to varying phone characteristics, vehicle features, and driving habits. The relatively small number of misclassifications did not significantly affect drivers' overall risk profiles, which should give insurers, researchers, and users greater confidence when using apps that rely on this classification algorithm to attribute risky behaviors to the user.

Sensitivity was greater than specificity, indicating a bias toward classifying trips as driver. When the algorithm is used with populations that have a smaller proportion of driver trips, we can expect overall accuracy to be lower. Based on the present study's sensitivity and specificity, a user with a 50/50 driver/non-driver split may see accuracy of 94.4%; a user with a 25/75 split may see accuracy of 92.8%. This decreased accuracy may not substantially alter a user's overall risk profile, but it could lead to lower user acceptance. For apps that let users correct misclassifications, encouraging infrequent drivers to make corrections could improve acceptance and classification accuracy.

A limitation of this study is its reliance on participants' reports of misclassifications instead of "ground truth" data (e.g., coded video footage from participants' vehicles). We chose this method because it was less invasive, costly, and effortful. Allaying some concern about inaccurate or incomplete self-reports, participants reliably reported the number of trips their apps detected, and the most reliable participants reported a high level of app accuracy.

Conclusions

Smartphone telematics apps hold great promise for measuring and improving driving behavior at scale. To be trusted, these apps must attribute behaviors to a user when the user is driving, and not attribute behaviors when the user is not driving. Correct attribution depends on the accuracy of driver/non-driver classification algorithms. Our results confirm that this classification can be done with minimal error using smartphone telematics data, and without user input or additional in-vehicle measurement devices.

Funding acknowledgments

This research was supported by CDC grant R49CEE003083 (PI: Z. Meisel) and funding from the Abramson Family Foundation (M.K. Delgado). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Centers for Disease Control and Prevention or the Abramson Family Foundation. The funders had no role in the study design or in the collection, analysis, and interpretation of data.

Data availability statement

Data will be made available on request.

References

- Abdenmour N, Ouni T, & Amor NB (2021). Driver identification using only the CAN-Bus vehicle data through an RCN deep learning approach. *Robotics and Autonomous Systems*, 136. ARTN 103707 10.1016/j.robot.2020.103707
- Cai H, Hu Z, Chen Z, & Zhu D (2018). A Driving Fingerprint Map Method of Driving Characteristic Representation for Driver Identification. *IEEE Access*, 6, 71012–71019. 10.1109/access.2018.2881722
- Choi G, Lim K, & Pan SB (2022). Driver Identification System Using 2D ECG and EMG Based on Multistream CNN for Intelligent Vehicle. *IEEE Sensors Letters*, 6(6), 1–4. 10.1109/lSENS.2022.3175787
- Choudhary V, Shunko M, Netessine S, & Koo S (2022). Nudging Drivers to Safety: Evidence from a Field Experiment. *Management Science*, 68(6), 4196–4214. 10.1287/mnsc.2021.4063
- Ebert J, Xiong A, Halpern S, Winston F, McDonald C, Rosin R, Volpp K, Barnett I, Small D, Wiebe D, Abdel-Rahman D, Hemmons J, Finegold R, Kotrc B, Radford E, Fisher W, Gaba K, Everett W,

- & Delgado MK (2022). Summary Report: Comparative Effectiveness of Alternative Smartphone-Based Nudges to Reduce Cellphone Use While Driving: Final Report (FHWA-HRT-22-057).
- Girma A, Yan X, & Homaifar A (2019). Driver Identification Based on Vehicle Telematics Data using LSTM-Recurrent Neural Network. arXiv pre-print server. None arxiv:1911.08030
- Jeong D, Kim M, Kim K, Kim T, Jin J, Lee C, & Lim S (2018). Real-time Driver Identification using Vehicular Big Data and Deep Learning.
- Kasperowicz L (2022). Usage-Based Auto Insurance and Telematic Systems. <https://www.autoinsurance.org/usage-based-car-insurance/>
- Mosley RC (2016). TrueMotion UBI Score Development. Pinnacle Actuarial Resources.
- Park J. g. (2018). Using telematics data to identify a type of a trip (U.S. Patent No. U. S. P. a. T. Office.
- Ravi C, Tigga A, Reddy GT, Hakak S, & Alazab M (2022). Driver Identification Using Optimized Deep Learning Model in Smart Transportation. *Acm Transactions on Internet Technology*, 22(4). Artn 84 10.1145/3412353
- Reimers I, & Shiller BR (2019). The Impacts of Telematics on Competition and Consumer Behavior in Insurance. *The Journal of Law and Economics*, 62(4), 613–632. 10.1086/705119
- Stevenson M, Harris A, Wijnands JS, & Mortimer D (2021). The effect of telematic based feedback and financial incentives on driving behaviour: A randomised trial. *Accident Analysis and Prevention*, 159. ARTN 106278 10.1016/j.aap.2021.106278 [PubMed: 34218195]
- Van Ly M, Martin S, & Trivedi MM (2013). Driver classification and driving style recognition using inertial sensors.
- Zhao D, Hou J, Zhong Y, He W, Fu Z, & Zhou F (2022). Driver Identification Methods in Electric Vehicles, a Review. *World Electric Vehicle Journal*, 13(11), 207. 10.3390/wevj13110207

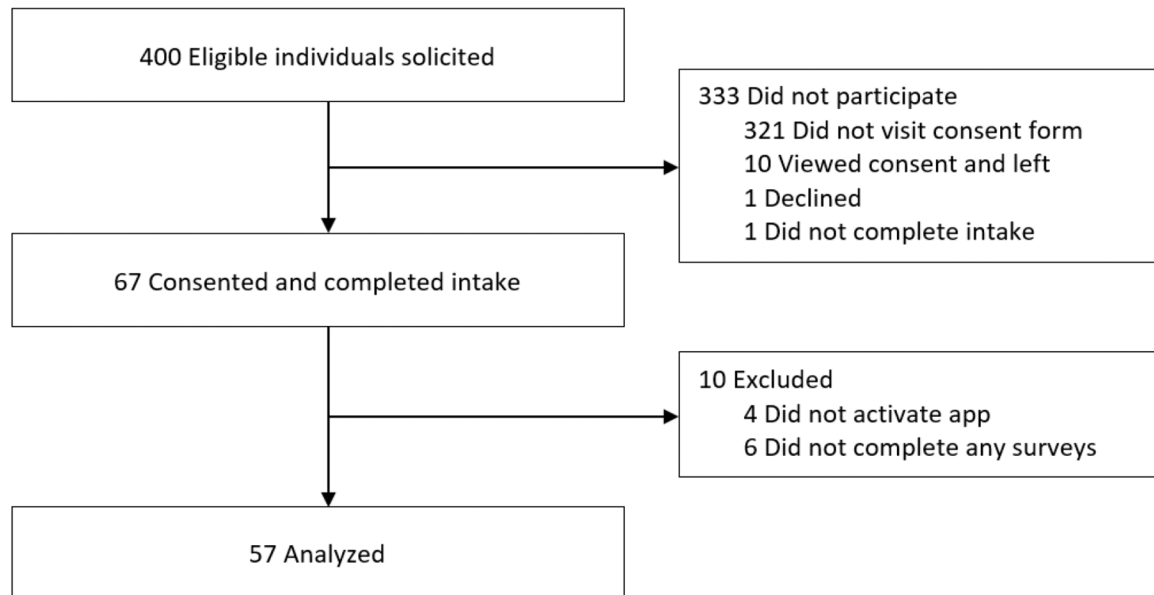


Fig. 1.
CONSORT diagram.

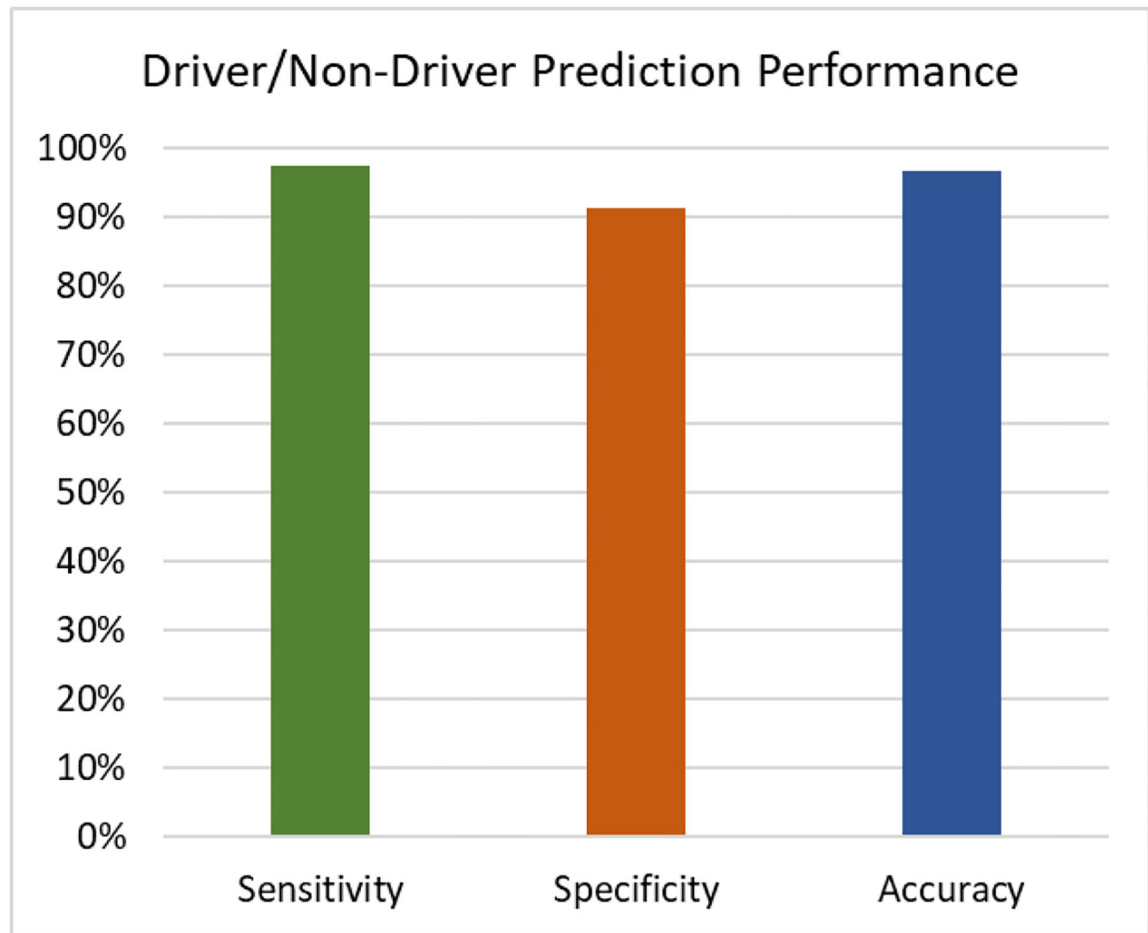


Fig 2.
Mean sensitivity, specificity, and overall accuracy.

Table 1.

Derivation of primary and secondary outcomes.

Variable	Computation
False positives (FP)	Total reported misclassified as driver
False negatives (FN)	Total reported misclassified as non-driver
True positives (TP)	Total app classified as driver – FP
True negatives (TN)	Total app classified as non-driver – FN
Sensitivity	$TP / (TP + FN)$
Specificity	$TN / (TN + FP)$
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$

Table 2.

Participant demographic characteristics as of the March 2021 launch of the earlier RCT. The present study launched July 2022, so mean age at time of participation was more than 1 year older.

Age (mean (SD))	33.2 (6.9)
Sex (%)	
Female	37 (64.9)
Male	20 (35.1)
Marital status (%)	
Married	13 (22.8)
Single	44 (77.2)
Race (%)	
White or Caucasian	40 (70.2)
Black or African American	13 (22.8)
American Indian or Alaska Native	1 (1.8)
Asian or Asian American	3 (5.3)
Undisclosed	1 (1.8)
Hispanic ethnicity (%)	4 (7.0)
Education completed (%)	
High school degree or equivalent	6 (10.5)
Some college	13 (22.8)
College degree	22 (38.6)
Some graduate	2 (3.5)
Post-graduate degree	14 (24.6)
Years of licensure (%)	
0 to 4	2 (3.5)
5 to 9	16 (28.1)
10 to 14	15 (26.3)
15 or more	24 (42.1)
Crashes in prior 5 years (%)	
0	36 (63.2)
1	16 (28.1)
2	3 (5.3)
3 or more	2 (3.5)