Utrecht University

Methodology and Statistics for the Biomedical, Behavioural and Social Sciences

Master Thesis

Automatic Travel Mode Prediction in a National Travel Survey

Author:
Laurent Smeets
6109020

Supervisors:
Peter Lugtig (UU)
Barry Schouten (CBS)

23rd June 2019

Word Count (excl.):
10,987 words

Preferred journal of publication:
Transportation Research Interdisciplinary Perspectives

Key words:
GPS; Travel survey, Automatic transport mode prediction; Big data

Abstract

Goal: Showing the feasibility of automatic travel mode prediction using smartphone location data in a national travel survey. **Data collection:** In the fall of 2018, 1,902 respondents were randomly sampled from the Dutch population to participate in a smartphone-based travel study. A purpose-built app that collected location data and generated a diary of stops and trips was used. For the trips, respondents could label which transport mode they used. Of the respondents, 517 completed data collection for at least 7 days and a total 18,414 trips were collected, of which 5,641 were labelled. **Method:** Every trip consists of a string of chronological ordered GPS points. From these points, trip-level features were engineered, such as average speed. Context-location data, such as the location of public transport stops, was then added and extra features such as how many train stations were passed during a trip were calculated. In addition, the data was enriched with respondent-level characteristics, available through Dutch registries. In total 127 features were engineered. A Random Forest Algorithm was then used to predict transport modes from these features. The transport modes distinguished are: Walking, Bike, E-bike, Car, Bus, Metro, Tram, Scooter, Train, and erroneously recorded trips. This last one is unique to this research, but inherent to app-based studies. **Results:** For 62% of trips the correct transport mode was predicted, when treating trips as independent events. Taking into account how often respondents used a certain transport mode increases the accuracy to 70%. Collapsing similar transport modes, such as bikes and E-bikes, also positively effects the accuracy.

1 Introduction

This paper shows the potential of automatic transport mode prediction in app-based surveys employing mobile device location sensors using machine learning. New relative to other studies is the use of a general population, official context of the survey, the size of the sample, the use of administrative registry data and number and types of features used. Both countries and cities organize regular travel surveys to study travel patterns and transport mode use of its citizens. These studies are necessary to gain insight into societies' mobility patterns, potential bottlenecks of the infrastructure, and are paramount in urban transportation planning. Moreover, accurate estimates of travel behaviour facilitate calculations of carbon footprints, as well as health related statistics such as calories burnt [1, 2]. Traditionally, self-registered diaries were used, in which people kept track of their own travel behaviour for one or more days. Over the last few years there has been an increase in the use of GPS-based (Global Positioning System) travel studies [3–7]. These GPS survey studies have a lower respondent burden and fatigue and greater spatial and temporal precision than the most commonly used alternative, diary-based studies [8]. Instead of letting respondents estimate how far, when, and what route they travelled, GPS-based studies can record these data more precisely. Furthermore, the fact that it is possible to collect data over longer periods makes it feasible to test dynamics of multi-day patterns [9].

At the start of the 21st century, GPS trackers were used in GPS-based studies [10]. A GPS tracker determines its current location on earth by trilateration of signals it receives from three or more satellites that orbit the earth. Every second it will calculate its current location, and save this. However, with the proliferation of GPS-enabled smartphones, phones have become attractive alternatives, as:

1. People are more likely to habitually charge and carry a phone [11].
2. In addition to the GPS, phones can also make use of the Global System for Mobile Communications (GSM) triangulation and the Wi-Fi Positioning System (WPS) to determine location when a GPS fix is unavailable [11].
3. The fact that people already possess

phones decreases the data collection cost.

Research by both Safi et al. [12] and Nahmias-Biran et al. [13] shows that GPS-based studies yield higher quality data than the diary-based alternative. Other authors who compared self-registered diary travel studies with GPS-based studies found a relative underreporting of travel behaviour in the diary-based studies. Respondents are especially likely to forget to report short trips, certain stops *en route*, trips which do not end or start at home, and the exact travel duration [13–15]. Moreover, there is a transport mode correlated bias. Car drivers are likely to underreport their trip duration, while respondents that take public transport are more likely to overestimate their travel time [15–17]. An extra disadvantage of diary-based studies is the non-uniform respondent burden and recall bias. Those respondents who travel more, or use multiple transport modes on a given day, have a harder time recalling exactly when and what distance they travelled and may, therefore, not report certain trips. On the other hand, someone who did not leave the house, theoretically, has it easy. Additionally, it is common practice, at least in the Statistics Netherlands' travel studies, to randomly assign a day to respondents on which they have to report on. It is then left up to the respondents when they report on that day. In the latest diary-based Dutch travel study, 61.8% of respondents waited three or more days to fill in their online diary. Analysis showed that the longer people wait, the fewer trips they report and especially the fewer short trips they report. This biases the results.

Taking these advances and advantages into consideration, Statistics Netherlands (CBS) launched a pilot study into the use of smartphone-based travel studies. It developed its own app and collected data in November and December 2018 [18]. Respondents kept this smartphone app on their phone and were tracked for a week. An automatic diary of stops and trips was generated, and respondents were asked to complete details about the stops (purpose) and trips (mode of transportation). The location tracking ensures accurate travel distance and time are recorded. However, to produce high-quality official statistics (for example, which age groups travel with what transport mode when), respondents still need to label their travel mode in the app. This means that, at the current state, smartphone-based travel studies lead to more precise measurements of distance and time travelled. However, respondents still need to actively label trips, as in a diary-based study. Trips collected in a GPS study can be plotted on a map and presented to respondents. This map can serve as a memory aid to help respondents recall which trips they made on a certain day and help them label these trips. Ideally, future iterations of this app would refrain from asking respondents to manually select transport modes from a long list, but would instead automatically classify the transport mode of different trips. An algorithm able to accurately predict transport modes could thus reduce respondents' burden even more, either by fully predicting transport modes or by giving respondents suggestions they only need to confirm. Additionally, a classification algorithm can be used to predict the transport modes of those trips that were collected, but not labelled. This way, the classification algorithm can be used as an imputation method [7, 9].

This paper shows the feasibility of automatic travel mode prediction using smartphone data in a representative national travel survey. The evaluation comprises a number of research questions. Primarily, what is the prediction accuracy of transport mode prediction? In addition, what is the effect on the prediction accuracy of collapsing different transport modes into broader categories? This allows travel study analysts to decide what is an acceptable balance between accuracy and the number of transport modes that need to be distinguished. What are the most important features in distinguishing between transport modes? Does respondent-level information from national registries add to classification success? This paper adds to the commonly used features in transport mode classification research different respondent level characteristics, such as age and whether someone has a driver's license available to Statistics Netherlands. Finally, how do the results compare to the results of the latest diary-based Dutch Travel Survey [19]?

The remainder of this paper is structured as follows. Section 2 presents a literature review focused on automatic transport mode prediction, stipulates how this research is unique, and what general challenges exist when working with GPS data. Based on Section 2, Section 3 presents the methods that were used in this research to deal with these challenges, as well as which features were engineered from the GPS data and how these were analysed. Section 4 presents the results of this research. This is followed by the Conclusion and the Discussion.

2 Background

This section introduces the main concepts in GPS-based travel surveys, the role of data cleaning and pre-processing, and commonly used classification techniques. It links these to the existing literature, which is extensive. The use of GPS data in national (travel) studies is novel, but not unique. For example, over the past decades (cities in), Australia, Austria, Canada, China, Denmark, France, Israel, Japan, the Netherlands, Singapore, Sweden, Switzerland, Tanzania, the UK, the USA, and Vietnam have all experimented with GPS-based studies [3, 11, 15, 20–34]. However, most of these studies are limited to estimating distances travelled. These studies are not aimed at automatic travel mode prediction and thus not the focus of this literature review.

Table A3 in the Appendix of this paper condenses earlier research (spanning 2003 to 2019) into travel mode classification from either smartphone data or stand-alone GPS trackers. The bulk of this section will summarize the conclusions from this table, but the next section will first further expand and explain some concepts.

2.1 Explaining the Concept(s)

GPS trackers and smartphones alike record lists of location data (measured in longitudes and latitudes with accuracy of this measurement) and corresponding times, typically once per second.[1] This list of points then needs to be split into chunks. Accuracy is measured by smartphones as a radius (in meters) in which it is certain it is currently situated. Thus, a larger (wider) accuracy means that the phone is less sure about its location. Most smartphones are accurate to roughly 5 meters.[2] Table 1 presents an example of the GPS data.

In essence, the goal of travel mode prediction is predicting the transport mode of a trip listed in the final column of Table 1, utilising the information in the first six columns. These different trips of GPS data just contain series of chronologically ordered points without any explicit features. Such features must first be engineered. To engineer these features, in most travel studies, GPS points are split into different trips and stops. A stop is a place a respondent stayed for a longer time (e.g. home, work, school) and a trip is any travel between stops. With some basic trigonometry and physics, it is possible to use longitude, latitude, and time to calculate features such as distance, speed, and acceleration. The location data can also be enriched with context location data, such as the location of bus stops and train sta-

[1] As mentioned above, smartphones have different ways to determine its location. The app requests this information from the phone and does not 'know' which method was used to determine the location. So technically the term "GPS-based" as used in this paper is not completely proper since there are other methods by which the location estimate might have been derived. However, as is the norm in the literature, this term is used throughout.

[2] Different smartphone operating systems use different methods to quantify this uncertainty. For Android, accuracy is *"the radius of 68% confidence"* [35]. iOS does not use a normal distribution for its accuracy measure, but does return some accuracy measure of uncertainty [36].

tions. Finally, the User ID can be used to link certain respondent level characteristics to the data, such as age and whether someone has a driver's license. If the data is structured like in Table 1, the number of respondents is the number of unique Users IDs and the number of trips the number of unique Trips IDs.

User ID	Trip ID	Latitude	Longitude	Time	Accuracy	Mode
15	1	52.370216	4.895168	01-11-2018 10:30:00	20	car
15	1	52.370213	4.895162	01-11-2018 10:30:01	15	car
15	1	52.370208	4.895160	01-11-2018 10:30:02	17	car
15	1	52.370198	4.895175	01-11-2018 10:30:03	18	car
21	4	50.851368	5.690973	02-11-2018 08:56:22	40	bus
21	4	50.851372	5.690984	02-11-2018 08:56:23	35	bus

Table 1: Example of merged GPS data

2.2 On the Importance of Filtering & Smoothing

This research, like much previous research, requires some pre-processing of the raw GPS data before it can be used to engineer features. Unavoidably, raw GPS data are messy and prone to measurements errors. For an explanation of the causes of this noise, refer to Ordonez et al. [37], Ivanović et al. [38], Chelvam et al. [39], or Roddis et al. [14]. Methods to filter likely measurement errors include discarding single points with a too wide accuracy (for example, an accuracy with a radius over 200 metres) or omitting data points that would lead to an unrealistic high speed (for example, over 200 km/h) [7, 40]. In addition to removing unlikely data points, further pre-processing in the form of smoothing the data to remove random noise. Removing this random noise reduces the 'spikiness,' of a trip. This way it gives more realistic estimates of the distance of a trip and the distance between points [3]. One method to do so is the Savitzky-Golay filter. For each location in a trip, this filter fits a polynomial to a set of input samples laid over an odd-sized window centred at the subject point. It then computes a new value for every point by evaluating the resulting polynomial at that point. The attractive property of the Savitzky-Golay filter is that it maintains the original shape of the pattern while reducing noise [40, 41]. The Savitzky-Golay filter has been used successfully by Dabiri et al. [40] in similar travel mode prediction research and implemented in R to smooth trajectories (of animals) by McLean et al. [42]. Another oft-used filter is the Kalman filter. This filter has the attractive property that certain location points can be given more weight, such as location points that are collected with a high degree of spatial accuracy. For information on the implementation and functioning of the Kalman filter, please refer to Hartikainen et al. [43] or Sobrado [44].

2.3 Earlier Research Often Used Small Samples

Summarizing Table A3, we see that most authors use small samples, either by the number of respondents, number of trips measured, or the number of measurements [45–48]. Most authors either collected data themselves or trained a group of participants to collect data. For example, Reddy et al. [1] collected 120 hours of data from 6 individuals, Feng et al. [48] used data from 8 respondents, and Martin et al. [49] used 6 students, while Das et al. [50] analysed 106 trips and Montoya et al. [51] only 87. Studies which use data of more than ten respondents and more than a thousand trips are uncommon. An exception to this was the 2015 study by Geurs et al. [5]. They used 600 respondents from the Dutch Mobile Mobility Panel (a representative sample of the

Dutch population) and recorded 18,000 trips using a smartphone app. Furthermore, most studies were conducted in a single city due to easy access to participants. This ignores the difference in travel patterns between urban and suburban (or rural) spaces, as well as the differences of municipal public transport systems within a country. It is also indicative of the use of selective and non-random samples.

2.4 On the Number of Transport Modes

There is wide variation in the number (3-9, median: 4) and type of transport modes classified in earlier research. Some authors include being stationary as a transport mode, while others do not [7, 48, 52–65]. Some authors combine all motorized transport; others even distinguish between walking and running. Only two authors list electric bicycles (E-bikes) as transport modes [66, 67]. In the Netherlands, this is an important category to distinguish [68].

2.5 On the Different Sensors and Features Used

There is an equally wide variation in the types of sensors used. The GPS (80%) and accelerometer (38%) are the two most common sensors, but authors have also used the barometer (as a proxy for local weather conditions) and the gyroscope [69]. Accelerometer data can be especially useful in distinguishing between transport modes when there is a poor GPS signal, but, of course, cannot estimate the distance travelled [1]. Of those authors using GPS data, almost all of them primarily used speed and acceleration to derive features to distinguish travel modes. Common are the mean, median, minimum, maximum, standard deviation, and different percentiles of speed and acceleration. Assemi et al. [70] and Xiao et al. [71] explored more statistics based on the distributions of speed and acceleration, including skewness and kurtosis. For instance, one can expect that the speed distribution of a car is skewed towards the speed limit. Speed and acceleration are also used to calculate related statistics such as the number of stops and breakpoints in a trip [72]. Heading and heading change direction are also valuable statistics, as a train is less likely to make sharp corners than a car. Measurements of data quality, such as location estimation accuracy and the number of satellites in view, can be used to predict transport modes. Phones in covered or underground transport modes will give less accurate location estimates. This way, accuracy can be used to improve travel mode estimation [70, 73].

21 out of 71 papers included in the literature review also used context location-based data. Most common are the location of public transport systems or road networks. The logic behind the inclusion of this information is that trips near to a train track are more likely to be a train trip. Stenneth et al. [7] and Shah et al. [60] also used real-time bus location data. While these features are informative, they rely on easy and up-to-date accessible information. Only a few authors include data on the respondent level. Exceptions include Feng et al. [58], Bohte et al. [15], Wang et al. [67], and Moiseeva et al. [74], who included different dummy variables for the ownership of different transport modes. Bantis et al. [65] also included personal characteristics, such as age, as a predictor. Someone who does not own a car or driver's license, might be less likely to take a lot of car trips.

2.6 Random Forest is the Most Suitable Method

To be able to classify transport modes, different types of supervised machine learning algorithms have been used. These range from linear models, such as multinomial logistic regressions [75], to convolutional neural networks [40]. The most popular methods are, in order, rule-based algorithms (including decision trees), Random Forests, Support Vector Machines, and Bayesian Networks. Authors comparing the performance of algorithms, specifically within the application of predicting transport modes, often found Random Forest to perform best [2, 7, 47, 48, 76]. These findings concur with Fernández-Delgado et al. [77] findings. Evaluating 179 classifiers arising from 17 families on 121 data sets, they find that: *"the classifiers most likely to be the best are Random Forest versions"*. On average, the classification accuracy is 85% for all reviewed research, while for research using Random Forest it is 92%. For the top ten studies in which most trips are evaluated, the accuracy is 86% on average.

2.7 On Which Trips to Include and How to Calculate Accuracy

There are two factors excluded from the table that are further needed to judge the success of the different study designs. First, authors use different ways to calculate precision in travel mode predictions. There are, broadly, two methods to measure precision. Either as the number of correctly labelled trips as a share of the total number of trips, or as the correctly labelled GPS points, time units, or distance travelled as a share of its respective totals. In this latter strategy, the precision can be dominated by one long correctly or incorrectly labelled trip, while many shorter incorrectly labelled trips are of lesser weight. Considering this, as well as the fact that trips, rather than GPS points or distance travelled is the unit of interest in this study, this paper will use the number of correctly labelled trips as a share of the total number of trips. Secondly, authors employ different selection criteria for the inclusion trips. For example, in Rasmussen et al. [78] *"Trip legs which cannot be map matched properly are classified as non-trips and removed from the data set"*, in Das et al. [79], *"A minimum trip distance is set as 5 km and trip duration as 30 minutes in order to capture activity travel behaviour for sufficient time period"*. In Mäenpää et al. [80] *"If more than ten consecutive corrupt track points were detected, the segment was dismissed as corrupt"*, and Martin et al. [49] *"removed all trips that did not contain at least 120s of data"*.

Strict inclusion criteria are likely to boost the precision of an algorithm. This paper takes a less stringent approach to the deletion of trips. Considering the aim of this study is to test the feasibility of transport mode classification in a large, random sample taken from a country's population, this study omits as little data as possible. This study retains hard to predict trips and trips with missing data.

2.8 On the (Dis)advantages of Using a National Survey

The current research is unparalleled in that a country-wide and representative sample of the country's population was used. In addition, this current research is distinctive because it is closer to a traditional travel survey than earlier research into transport mode classification. This is due to the fact that it suffers from both person and item non-response, in addition to insincere and incorrect answers. Furthermore, respondent burden was of real concern and balanced against the quality of data collected. The app was developed to not drain battery too fast and could operate without an online data connection. Similarly, the app did not save accelerometer data. This decision was made to prevent too much data from being created on a respondent's phone (accelerometer data is sampled at 30-60 hertz). Due to the fact that randomly sampled untrained respondents were used, it was hard to guarantee that they would all complete all data collection and labelling. Other researchers went out of their way to guarantee complete data sets, either by collecting the data themselves or by regularly communicating to respondents about the trips they made that day [15]. For example, Zhao et al. [4]: *"To ensure the quality of the data the volunteers were asked to carry a paper diary and register the start and end of each trip* *in it. The participants were issued travel cards (EZLink cards) to make payments for the train, bus and taxi trips. The transaction logs from the EZLink cards were later used to filter the erroneously annotated trips"*. This current research did not follow up with respondents in such a fashion.

3 Data & Method

This section discusses the way data was collected, edited, and then analysed. After presenting an introduction to the data and the app that was used to collect the data, it describes which of the above-mentioned methods were used to clean and smooth the data. This section also details which data is preserved and which data is omitted. This cleaned data is then used to engineer features from on the trip level. Three different types of features are distinguished and presented. These features are the input of the Random Forest model. Subsection 3.5.2 explains how the importance of features is estimated and subsection 3.6 explains how the results obtained from labelled trips can be used to predict unlabelled trips. Finally, subsection 3.7 explains the way GPS-based results were compared to diary-based results. The annotated code is available on the author's GitHub page [81].

3.1 Data Collection & Description

Statistics Netherlands specifically developed the "CBS Verplaatsingen" app for Android and iOS for this research project. The app's source code can be found at the app's Gitlab page [82]. On 31 October 2018, 1,902 respondents sampled from the Dutch registry were invited to participate. Respondents who did not register by 14 November, received a reminder that day and respondents who did not participate for 7 days, were sent a reminder on 21 November. The app automatically collects location data based on GPS (most accurate), WPS (second most accurate) or GSM (least accurate). The app splits the location data into trips and stops, with stops classified as locations where a user stayed over a time threshold θ_t within a distance threshold θ_d. Consequently, every location is either part of a trip or stop. A user's trip is a sequence of spatiotemporal points (P) recording the travelled path, i.e.: $trip = (P_0, P_1, \ldots, P_n)$, where a location point $P_i = (x_i, y_i, z_i, t_i)$ is the position $(latitude_i, longitude_i, altitude_i)$, where longitude varies from 0° to ±180° and latitude varies from 0° to ±90° at timestamp $t_i, \forall\, 0 \leq i \leq n$. Considering the geography of the Netherlands, altitude is considered constant. In the app, respondents provide the mode of transport for trips and purpose of stops. Home might be a first stop of the day, riding a bike to the train station the first trip, waiting at the station the second stop, sitting on the train the second trip, *et cetera*. Respondents were asked to label both trips and stops, but doing so was not required. Respondents could label a trip with multiple transport modes. Figure 1 presents the interface of the app that was used.

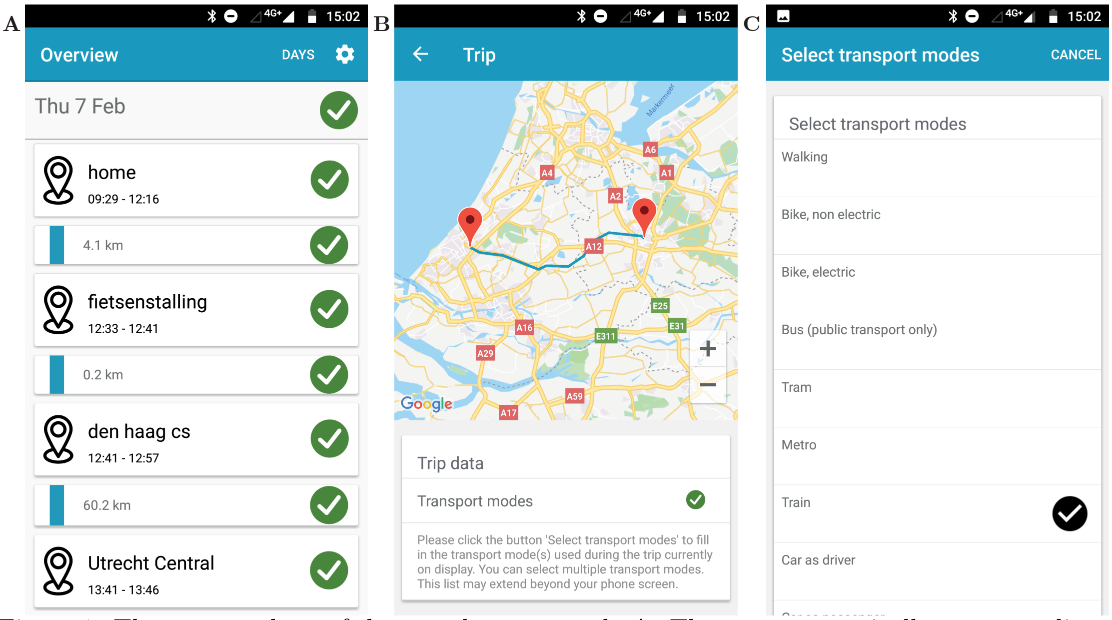

Figure 1: Three screenshots of the app that was used. **A:** The app automatically creates a diary of stops and trips. Respondents can then label these. **B:** When clicking on a trip, a map opens with the trip plotted on a map (in this figure the trip between the stops "Den Haag CS" and "Utrecht CS"). **C:** The respondent can then select one or more transport modes.

Of the 1,902 invitees, 674 downloaded the app, and 517 completed data collection for at least 7 days (the time required to receive a remuneration). In total, 75,649,744 location points were collected, amounting to 796,611km travelled over 18,414 trips, of which 5,641 were labelled. This completion rate is comparable to the study by Roddis et al. [14] in Melbourne, who had 38% of their active participants validate their data during at least two weeks. Figure 2 presents all the trips in the data set. The median number of labelled trips per person is 11 for those respondents that labelled at least one trip. The largest number of labelled trips for a single person is 65 and the minimum 1. 40% of respondents did not label any trip at all, and only one of the respondents labelled all of their trips.

At this stage of the research, multi-mode trips (trips for which a respondent indicated there was more than one mode) are excluded. A total of 144 trips were multi-mode trips and therefore omitted from analyses. Of these, 89 were trips that either started or ended with walking and used a single other transport mode. So, only 55 trips were multi-mode trips with two or more modes that did not include walking. Some uncommon transport modes that users labelled were excluded too. These include: horses, boats, and skippyballs. They were excluded because of their rarity in the data and because there was doubt about the sincerity of labelling. Without these (comparatively) lenient filters, there would be too many different classes of transport modes to distinguish. After omitting these trips there were 5,296 trips left.

In the data there were some erroneously recorded trips. These are trips that are recorded by the app but should not have been. This can happen because the app records a GPS fix where there is none, or because it connects to another Wi-Fi network or GSM tower without the user actually moving. Therefore, this paper adds to the literature by not only creating features to distinguish between transport modes, but also by predicting erroneously recorded trips labelled *"User error"*. In the app, there was no option to label trips as being erroneously recorded, but there was the possibility to leave comments under the *'other'* option. By using regular expression commands [83], comments that indicated a trip was falsely recorded (e.g. *"This trip did not happen"* or *"I did not move"*) were labelled *"User errors."* At this point, regular expressions were also used to combine certain transport modes into categories and to correct for respondents' typos. For example, 'car driver' and 'car passenger' were joined, as were 'bicycle' and 'bycicle' [*"fiets"*, and *"feits"* in Dutch]. Considering travel mode is of interest and the app already splits GPS data into trips and stops, this paper excludes stationary movement as a travel mode. This makes the set of predicted transport modes: mode ∈ {Walking, Bike, E-bike, Car, Bus, Metro, Tram, Scooter, Train, User error}.

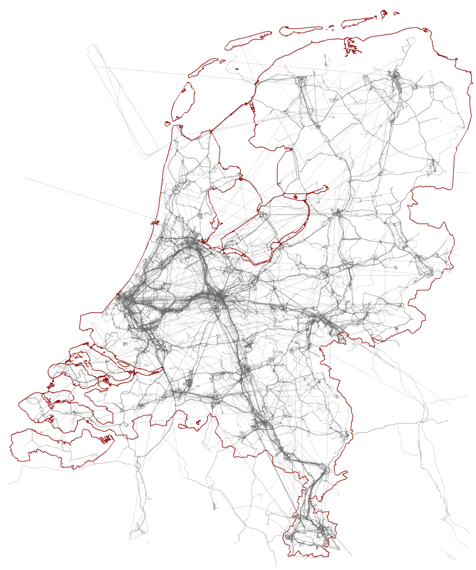

Figure 2: All 18,414 trips in the data set. This research is geographically representative of the Netherlands, with an expected clustering of trips in and around the big cities

3.2 Data Handling

On the back-end of the app, location data, trips, and transport modes were saved separately and then merged into a single data frame (again, see Table 1 for an example). In certain situations, the app struggled with saving data correctly. Both at the level of GPS points and at the trip level, repeated storage occurred. Prior to analysing the data, some cleaning and editing was necessary to correct for these duplication errors. A separation can be made between editing data on a trip level, and filtering and smoothing on the level of GPS data points.

3.2.1 Merging and Deleting Double Entries on Trip Level

To delete erroneously detected stops, trips by same user with the same transport mode and a stop of fewer than 3 minutes in between trips were merged into a single trip. This might, for example, happen because of a longer stop at a train station. The data contained trips by the same person with the same start time that were saved multiple times with different end times and/or different transport modes. This would imply that a location could be part of different trips. To process these anomalies, trips by the same respondent with the same start time, but different end times or transport modes, are considered one trip and merged as such. The latest end time of all registered end times and the transport mode belonging to the trip that was saved last were selected. Using this method, the logic that every movement is either a trip or a stop is preserved. These anomalies exist because the app sometimes struggled to determine the end of a trip and start of a stop; it saved the same trip with slightly different end times many times over. Trips that consisted of 3 unique location points or less were completely omitted. This to ensure all features can actually be calculated.

3.3 Filtering & Smoothing on GPS Level

On the level of the GPS data, similar double data collection problems exist. For GPS points with the exact same latitude, longitude, and timestamp, only one was retained. In consonance with the method employed by Dabiri et al. [40], and Stenneth et al. [7] data points with an accuracy of 200m + and data points that would lead to a speed of +200km/h between adjacent points are discarded. To smooth the location data further, this paper used both a Savitzky-Golay and a Kalman filter. The Savitzky-Golay smooths the data and, akin to the method used by Widhalm et al. [84], the Kalman filter smooths further by giving more weight to points with more precise accuracy.

3.4 Feature Engineering

As explained, this paper does not analyse the data on the level of GPS points, but first engineers features on the level of the trip using the GPS points that, together, make one trip. This paper distinguishes between three different types of features: GPS-based features, (context) location-based, and registry-based. In total, 127 different features were computed and summarized in Table A1 in the Appendix. As a result, this study uses more features than any earlier study. All features were calculated twice; once on the smoothed data and once on the linearly interpolated (smoothed) data, in which a location estimate was calculated for every missing second. For certain features this makes no difference, such as total time travelled, but for others it does, such as proximity to bus stops. Both sets were included in the model as features.

3.4.1 GPS-Based Features

To calculate trip-level statistic of the GPS-based features, such as the mean, median, min, max, 0.05^{th} percentile, 0.95^{th} percentile, skewness, and sd of speed, acceleration and angles between points, one first has to calculate the Euclidean distance between consecutive points. To calculate the great-circle distance between two sets of location estimates on a sphere, this research uses the spherical trigonometry's Haversine formula [85]. This information was also used to calculate the number of turns above a certain threshold (90°and 150°). Section A.1 of the Appendix gives the formulas that were used to calculate these statistics. In this set of features, the mean accuracy of the location measurements of a trip, and whether or not a trip was during rush hour or not are included.

3.4.2 Context Location Based Features

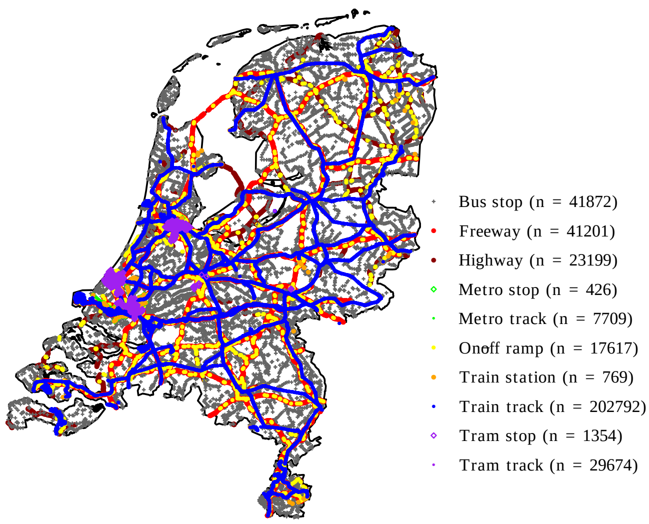

Figure 3: All OSM points of interest that were included in the research.

While GPS-based features are likely to be sufficient to separate certain transport modes from each other, such as cars and walking, extra context location information can help in classifying some other modes. A car and a bus might have similar speed patterns, but a bus will drive past and stop close to bus stops more often. The context location-based features included in this research are the location of all bus stops, train stations, tram stops, metro stops, highway and freeway on-off ramps, and the OpenStreetMap (OSM) nodes that are part of highways, freeways, train tracks, metro tracks and tram tracks [86]. Exact locations were scraped from the OSM using the Overpass API implementation in R [87]. Through this, 366,613 points of interests were collected (see Figure 3). Using the location of these features, the distance to them can be calculated for every location point. Per trip, the average distance to any of these features, as well as which ratio of points in a trip are in the vicinity (less than 25 meters) to the features, can be calculated. For the public transport features, one can, additionally, check if both the start and end point of a trip are near to one. See Biljecki et al. [88] for a brief discussion on the data quality of OSM features in the Netherlands. Exact real-time locations of buses and other public transport modes are not available for all public transport mode operators in the Netherlands and were therefore not used.

Furthermore, using the Statistics Netherlands data on local population density, this study checks whether a GPS point lies in an urban area, and what the population density of the area surrounding this location is [89, 90]. With this information, one can estimate which ratio of location points of a trip fell within urban centres, how many neighbourhoods were visited per trip, and what was the average population density of the locations of that trip. This way, the interaction between speed and urbanicity can be included in the algorithm. One can, for example, expect that cars have a higher mean speed outside, rather than inside, of urban areas.

3.4.3 Registry Based Features

As mentioned above, little earlier research included respondent-level information; something this current research is particularly well-equipped to do. Statistics Netherlands has access to approximately 30 variables stemming from administrative records for every person in the Dutch population. Respondent level statistics aiding in the differentiating of transport modes may include car ownership, age, and income. For example, one hypothesis that can be tested is if the interaction between speed and age can help differentiate between bikes and E-bikes. For all respondents, their age, gender, income percentile, urbanicity of their home address, car ownership, scooter ownership, and category of driver's license were collected from the relevant Dutch registries. These respondent level features are added as features to the prediction model.

3.5 Models

The supervised machine-learning algorithm of choice is the Random Forest algorithm [91]. The Random Forest algorithm was implemented using the CARET package [92] in R [93] with the RANGER implementation of the Random Forest algorithm as implemented by Breiman [91]. For an introduction to the Random Forest algorithm, the author refers to Banerjee et al. [94]. Random Forest models can be expected to work well on transport mode classification as compared to other classification algorithms such as multi-class SVMs. This is because, inherently, many features will correlate strongly (e.g. the 95^{th} and 99^{th} percentile of speed) and are measured on different scales. As recommended, a tuning strategy to find the optimal number of random features per tree was used and this optimum was implemented [95]. The number of trees was held constant at 1000. As recommended [96–98], to assess the performance of the model, a 5-fold 10 times repeated cross-validation will be used. In addition to overall model accuracy:

$$\text{accuracy} = \frac{\text{correctly classified trip}}{\text{total trips}},$$

class level sensitivity and the macro-averaged F1 statistic are reported. The F1 statistic is the harmonic mean of sensitivity (also called recall) (ρ) and precision (π). ρ and π are defined as:

$$\rho_i = \frac{TP_i}{TP_i + FN_i}, \pi_i = \frac{TP_i}{TP_i + FP_i}$$

Here, TP_i (True Positives) is the number of trips assigned correctly to transport mode i; FP_i (False Positives) is the number of trips that do not belong to transport mode i, but are assigned to transport mode i incorrectly by the model; and FN_i (False Negatives) is the number of trips that are not assigned to transport mode i by the model, but which actually belong to transport mode i. In macro-averaging, the F1 statistic is computed locally over each class and then the average of these measures is taken.

$$F1_i = \frac{2 \cdot \pi_i \cdot \rho_i}{\pi_i + \rho_i}$$

$$F1_{macro\ averaged} = \frac{\sum_{i=1}^{M} F1_i}{M},$$

where M is total number of classes (transport modes in this research).

3.5.1 Order of Analysis

To test whether context and respondent-level features increases the performance of the model, first, only the features related to the GPS data and measurement information were included (e.g. speed, acceleration, turning, accuracy, etc.). In a second step, context location related features were added (e.g. location of bus stops and train tracks and urbanicity). Thirdly, the respondent level information was added (age, income, driver's license) in a third model. In the fourth to the eight model, certain transport mode classes are combined. E-bike and bikes are collapsed into one category, as are cars and scooters, and then trams, buses, metros, and trains. Finally, the User error category is dropped in the ninth model. This collapsing of transport mode is done to test whether combining similar, and therefore potentially hard to distinguish, transport modes improves the accuracy.

Due to class imbalance in the data (there are only 19 tram trips compared to 2,172 car trips), a downsampling strategy was adopted. This strategy counters the tendency of the Random Forest algorithm to focus on the prediction accuracy of the largest classes, leading to poor local accuracy of smaller classes [99, 100]. Additionally, models were optimized for a high Cohen's Kappa, instead of overall high model accuracy. Cohen's Kappa is a normalization of the accuracy metric that takes the baseline random chance of correct classification due to class imbalances of the dataset into account. It does so by comparing observed accuracy with the expected accuracy under random chance [101]. For model three and nine, confusion matrices are presented.

3.5.2 Feature Importance

Not all features are created equal. Some features are more useful in classifying transport modes than others. To investigate which features are most influential in the different models, feature importance metrics are calculated. Instead of using the default 'decrease in impurity mechanism' of features, this paper adopts the permutation importance as a metric of feature importance as proposed by Strobl et al. [102]. Simply summarized, this method passes out-of-bag (OOB) samples (those that are not used in the training set) through the Random Forest model, records this baseline accuracy, then permutes a single feature (predictor) into random noise by shuffling the values of that feature around before rerunning the Random Forest model. The importance of a feature is then calculated as the difference between the original accuracy and the accuracy of the model with the permuted column. Simply removing the feature instead of permuting it is not possible since the Random Forest needs all features in the test set to fit the model. This is more computationally intensive than the impurity mechanism, but proven to yield less biased estimates of feature importance [102, 103]. In the impurity mechanism, for every feature the sum of Gini decreases of the splits across all the trees of the Random Forest in which that specific feature was included, is calculated and weighted. This method is almost 'computationally free' since the decrease in node impurity for every split in every tree is recorded anyway, but is *"not reliable in situations where potential predictor feature vary in their scale of measurement or their number of categories"* [104]. Features used in this paper vary widely in their scales of measurement. Permutation importance can also be argued to be a purer metric of feature importance, since this metric actually estimates how much a feature contributes to the model predictive performance. The impurity measure uses decrease in impurity as a proxy, but not direct measurement [105].

3.6 Predicting Unlabelled Trips

Since the data was collected in an actual travel survey, as opposed purposefully collected to later classify, there is considerable item non-response in the form of unlabelled trips. Fortunately, the fitted Random Forest model does allow for prediction of these 12,773 unlabelled trips. The proportions of how often the different transport modes are used can then be compared to labelled trips and to the data collected in earlier, diary-based, studies. Prediction only works if we assume that the labelled and unlabelled trips come from the same data generating mechanism and that there are sufficient similarities between them. If, for example, all unlabelled trips are horse riding trips, prediction will be futile with a trained algorithm which does not include a horse as a transport mode. Similarly, if all unlabelled car trips would be long-distance trips, the algorithm would fail to predict correctly.

3.7 Comparing Results to Diary-based Study

For the comparison with diary-based studies, this research uses the 2018 diary-based Dutch travel survey (*"Onderzoek Onderweg in Nederland"* or ODiN) [19]. In this research 56,858 people participated, who kept a travel diary for a single day [19]. In the sampling of respondents for this smartphone GPS-based survey, roughly half was sampled from the respondents from this latest (diary-based) travel survey, which was held 2 months prior. This facilitates some further comparison between results of the two survey methods, because we have respondents who participated in both studies. Of the 674 respondents who participated, 422 came from this OdIN group. Together they made up 60.6% of the labelled trips and 52.9% of the unlabelled trips.

4 Results

In this section, the research questions on accuracy, collapsing transport modes, feature importance, influence of respondent-level information, and comparison with diary-based survey. In addition, the paper explores what factors help explain why transport modes are correctly predicted or not in Subsection 4.3. This paper will also present a strategy to improve prediction accuracy. It does so by including information on the previous travel behaviour of respondents.

4.1 What is the Accuracy?

Figure 4 presents the results of the nine different models. As is to be expected, adding extra information to the model in the form of more features yields higher accuracies; so too does combining classes of similar transport modes. The accuracy of the model without registry information or context data is 52%. Adding the context location data clearly improves the overall classification sensitivity of public transport modes (metro goes from .48 to .88 and train from .73 to .80, for example), but has little influence on modes such as walking or scooter for which there are no related location features. The total accuracy goes up 60% when adding the context location features.

4.1.1 Effect of Adding Registry Information?

Adding registry data further improves the model accuracy (62%) and macro-averaged F1 statistic slightly, mostly due to a better separation of bikes and E-bikes.

4.1.2 Effect of Merging Transport Modes?

Progressively merging different transport modes also leads to higher accuracy, sensitivity, and F1 statistic, at the cost of being able to distinguish between fewer transport modes. Collapsing E-bikes and bikes, scooters and car, and tram and metro lead to the highest improvements, while further adding the train, and then bus, to the collapsed category of public transport only improves accuracy slightly (up to 77%), but does lead to a higher F1 statistic. When modes are collapsed into one, the sensitivities are combined too (but not necessarily averaged). This can lead to a decrease in sensitivity for certain modes, for example for the sensitivity of scooter when it is combined with car.

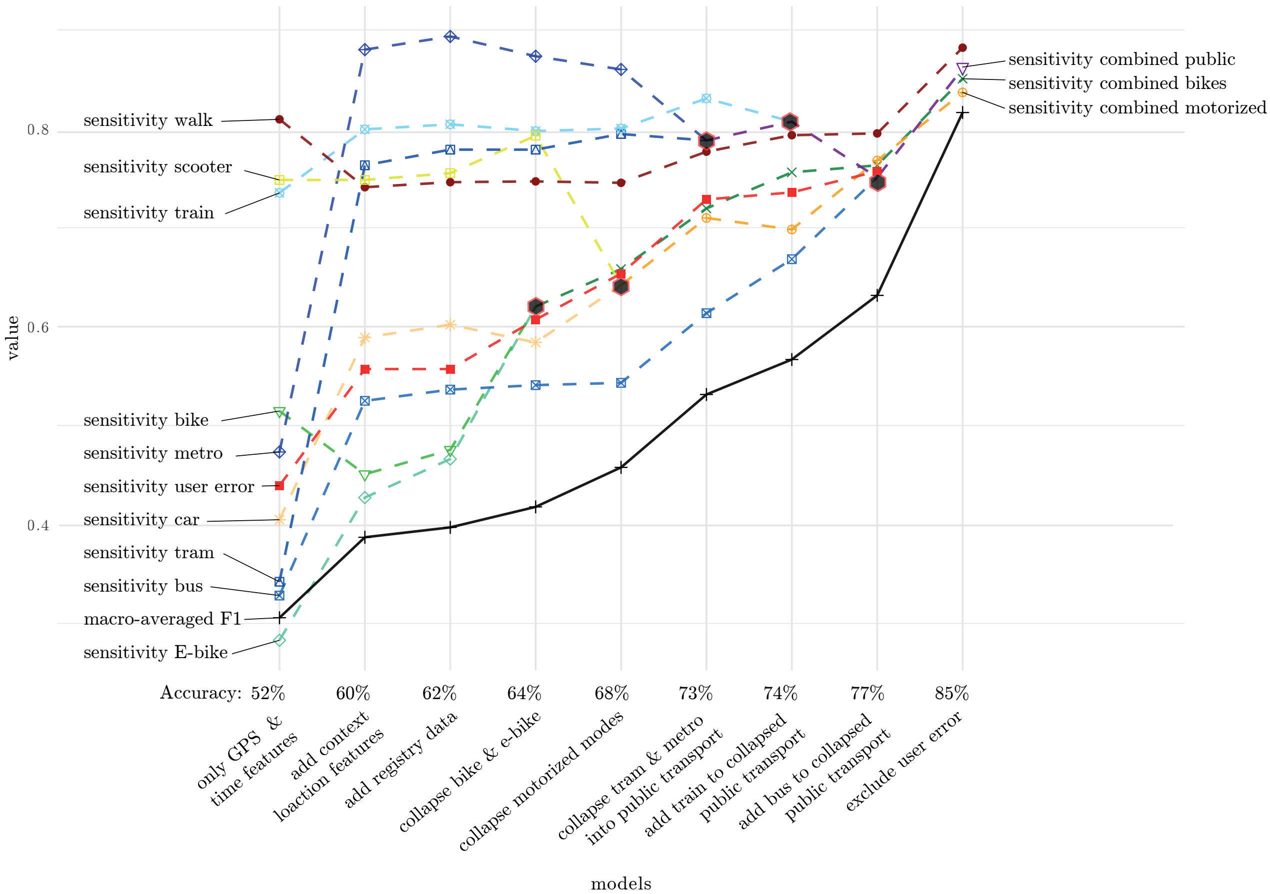

Figure 4: The black line represents macro-averaged F1 statistic of the nine different models represented on the x-axis. The coloured dashed lines represent the sensitivity of the model to different transport modes. The accuracy of the models is listed on the x-axis. In the first three models, additional features are added. In models four to eight different classes of transport modes are collapsed (⬢ indicate when which modes are collapsed). The final model represents the model with collapsed modes, without the User errors.

Dropping the User error category further improves the predictive power of the model to an accuracy of 85%. However, this means that one would have to develop an app that does not record erroneous trips.[3]. The confusion matrices of model 3 and 9 are presented in respectively Table 2 and 3.

[3]If the User error category was already dropped after the third model (not presented here), the accuracy of this model would be 64%

predicted	observed									
	$E-bike$	$bike$	car	$metro$	bus	$scooter$	$train$	$tram$	$User\ error$	$walk$
$E-bike$	70	164	96	0	2	1	1	0	0	22
$bike$	29	361	51	0	2	0	0	0	1	35
car	8	20	1308	0	8	3	5	0	3	18
$metro$	0	11	24	13	0	1	7	2	0	9
bus	4	20	199	0	24	1	4	0	1	5
$scooter$	13	14	195	0	0	22	0	0	0	4
$train$	2	4	74	0	2	0	142	0	1	10
$tram$	2	53	35	1	4	0	7	15	2	35
$User\ error$	10	54	109	0	1	1	8	2	16	91
$walk$	10	59	82	0	2	0	1	1	3	671

Table 2: Confusion Matrix for the third model with all the features and no collapsed transport modes yet, accuracy: 62%

predicted	observed			
	bike collapsed	motorized	public transport	walk
bike collapsed	772	144	8	58
motorized	50	1842	19	30
public transport	12	108	219	18
walk	73	108	8	793

Table 3: Confusion Matrix for the ninth model with all the features and collapsed transport modes and no User errors, accuracy: 85%

4.2 Feature Importance?

Table 4 presents the most important features in models one, three, and nine. Clearly, speed-related features, such as the mean, median, and different percentiles, contribute considerably in the potential to distinguish different transport modes. Other features, such as the skew of the speed distribution throughout the trip, acceleration, heading chance, and the features related to the number of turns are only of marginal importance. For the models in which context location information is included, especially the vicinity to train tracks and other public transport locations are of importance. This is especially true in the models in which different types of public transport need to be distinguished (model 3,4, and 5). Once the public transport modes are collapsed, these features are no longer among the most influential ones, as seen in the last column of Table 4.

In consonance with the results presented in Figure 4 (that adding registry data only adds little to the total accuracy), we do not see any of the registry related features among the most important features. In model 3, scooter ownership was the 18^{th} most important feature (30.3%). For visual aid, Figure 5 presents the marginal probabilities of the different transport modes on three of the most influential features in model three. Section A.3 in the Appendix explains how these plots are constructed. They can be interpreted as the probability that a trip belongs to a certain transportation mode given the different levels of the feature the x-axis and given that all the other features stay constant. From this it can be concluded that, trips that are relatively close to train and subway tracks are respectively more likely to be trips using trains and subway as transport modes. Similarly, trips with a high 90^{th} percentile of speed are more likely to be car trips, while trips with a low $.90^{\text{th}}$ percentile are more likely to be walking or User errors. When the Random Forest models are retrained with only the most important features presented in Table 4, the accuracy is 51% for the first model (1% lower than the model with all the features), 59% in the third (3% lower) and 84% in the ninth model (1% lower). These results might be unsurprising, but the fact the model picks up on understandable and interpretable features and not on noise is reassuring.

	model 1		model 3		model 9	
	features	%	features	%	features	%
1	speed $P_{95\%}$	100	rel. prox. train tracks	100	speed $P_{75\%}$	100
2	distance straight line	57.0	rel. prox. subway stops	92.0	speed $P_{90\%}$	80.2
3	speed $P_{75\%}$	52.9	speed $P_{90\%}$	79.3	median speed	49.2
4	speed $P_{90\%}$	51.7	rel. prox. subway tracks	74.6	speed $P_{95\%}$	42.9
5	mean speed	49.3	speed $P_{95\%}$	72.4	mean speed	37.4
6	median speed	47.6	rel. prox. tram stops	61.1	rel. prox. train tracks	29.4
7	speed $P_{25\%}$	36.8	distance straight line	45.7	σ speed	20.6
8	speed $P_{5\%}$	36.3	speed $P_{5\%}$	42.9	n point near train track	15.3
9	acceleration $P_{10\%}$	30.2	speed $P_{75\%}$	41.1	speed $P_{25\%}$	14.8
10	ratio speed + 120 km/h	29.1	median speed	37.7	unique train stations	11.4

Table 4: Top 10 most important features for model 1 (only location), model 3 (context location and registry information added), and model 9 (transport modes collapsed and User errors omitted). Feature importance is measured using the permutation method and all features were checked for significance using the method proposed by Altmann et al. [106] (all p-values of presented features $< .01$). Importance values are all scaled, such that the most important feature in the model gets 100%. Values can only be compared within in models and not between models. Table A2 presents the summary statistics of 10 important features for the different modes (apologies to the reader for having the interpret such things as a median of the mean of speed per trip per transport mode).

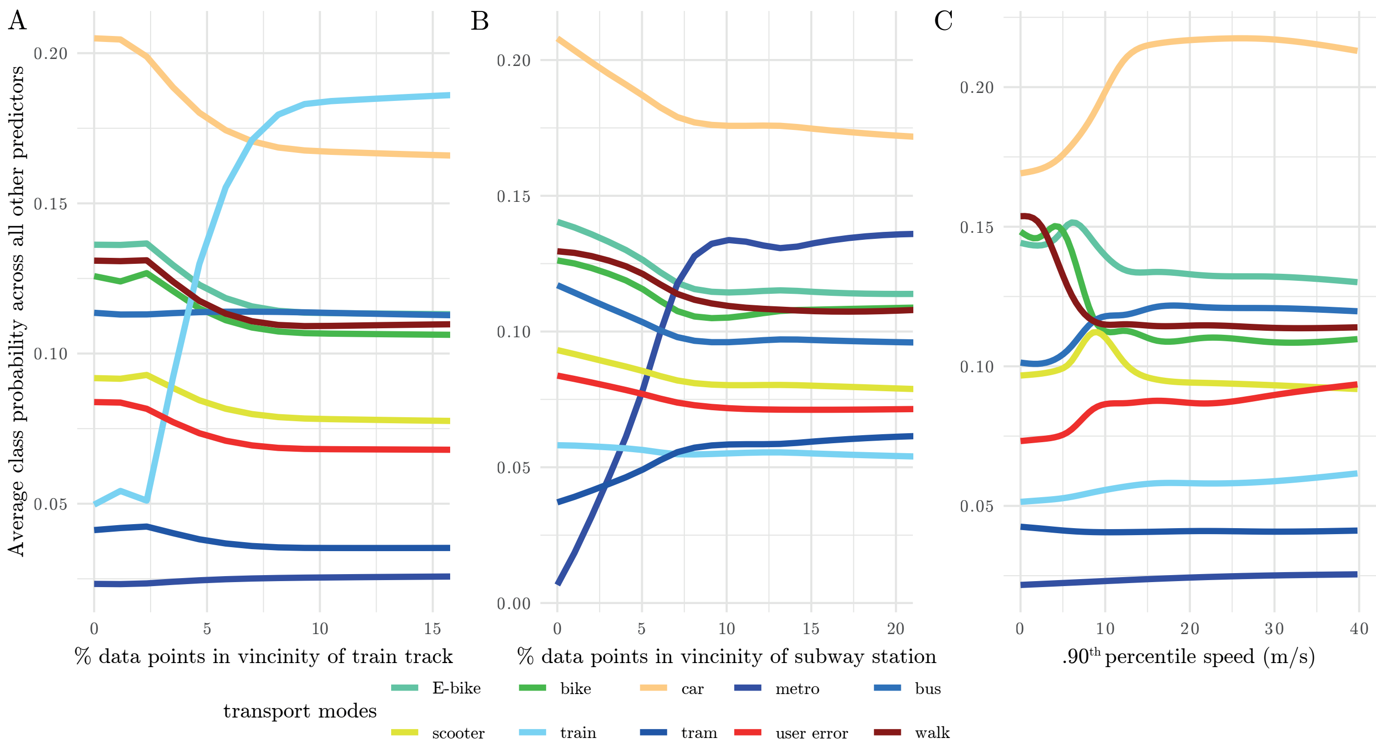

Figure 5: Marginal probability plots of the top three features of model three. **A** is the marginal probability plot of the ratio of location points near railway tracks, **B** is the marginal probability plot of the ratio of location points near subway stations, **C** represents the marginal probability plot of the .90$^{\text{th}}$ percentile of speed. These plots clearly show the class imbalance in the data, but also how these features do really influence the probability of a class, given that all other features stay constant.

4.3 On What Trips are Classified Correctly

A logistic regression with as binary outcome variable whether or not a trip was correctly classified, is presented in Table 5.[4] This regression shows that indeed the number of data points a trip consists of predicts whether a trip was correctly classified. This finding supports the conclusion of Byon et al. [73], who found that longer monitoring leads to higher accuracy. Other variables important in predicting whether a trip is correctly classified include the average accuracy of GPS points in a trip, the ratio of points in an urban area, and the transport mode. A smaller, as in more precise, accuracy radius leads to a better probability of correct classification. With a wider accuracy range, most features, such as speed or proximity to a bus stop, are less precisely calculated. Trips in urban areas are easier to predict, most likely because more context location features are present in urban areas. Finally, not all modes are equally likely to be classified correctly. This can also be deduced from the confusion matrices. Variables related to the smartphone used by the respondent, such as the brand, model, and the operating system (iOS or Android), were not related to correct clas-

[4] For the interested reader, a Bayesian implementation of this analysis can be found in the supplemental materials hosted on the author's GitHub page [81].

sification and not presented here. One reason that correctly classifying trips is hard is because short trips simply have very little information in them. If only trips longer than 10 minutes are included, the accuracy of model 3 increases to 65% percentage points and the accuracy of model 9 rises to 87%.[5]

	Dependent variable:		
	classified (ref = incorrect)		
	coef	S.E.	O.R. [profile CIs]
n data points	0.0003***	.00005	1.0003 [1.0002; 1.0004]
mean accuracy (meter)	−0.003***	.0003	.997 [.997; .998]
ratio of trip in urban area	1.352***	.122	3.866 [3.050; 4.920]
Transport mode (ref = walk)			
mode (E-bike)	−1.116***	.186	.328 [.227; .472]
mode (bike)	−1.208***	.108	.299 [.242; .369]
mode (car)	−1.067***	.095	.344 [.285; .414]
mode (metro)	1.213	.786	3.363 [.874; 22.267]
mode (bus)	−1.041***	.326	.353 [.186; .675]
mode (scooter)	0.467	.502	1.595 [.645; 4.811]
mode (train)	0.501**	.235	1.650 [1.055; 2.656]
mode (tram)	0.263	.541	1.301 [.479; 4.167]
mode (User error)	−0.472	.413	.624 [.277; 1.422]
Constant	0.825***	.079	2.281 [1.956; 2.670]
Note:			**p<0.05; ***p<0.01

Table 5: Logistic regression on the correct classification of transport modes.

4.4 Using Earlier Trips to Improve Accuracy

Thus far, all trips have been analysed as separate and independent events, without taking into consideration the nested structure of the trips within persons. This way, the Random Forest algorithm merely uses all the different features to classify a trip. This is useful if you want to use the trained model from this research on new respondents. In reality, knowledge on previous transport modes used by a respondent is likely to help predict the transport modes of trips by the same respondent. For example, a person who uses an E-bike regularly might be less likely to use a bike for a given trip. To test this hypothesis, dummy variables for all transport modes were calculated per respondent, indicating whether a respondent labelled this mode at any time during the research period. Due to the relatively large sample size of respondents, this paper can test whether the inclusion of such dummy variables improves the accuracy of the model. To prevent the labelling of the current trip from influencing the classification, the trip that is being classified is not included in the dummy variable calculation. The accuracy of a model with these dummy variables, gives an idea of how the algorithm would perform if respondents would indicate which transport modes they regularly use, prior to using the app for the first time.

[5] In addition to the so-far presented Random Forest models, this paper also applied other supervised machine learning techniques to classify transport modes, including Support Vector Machines, Naïve Bayes, Neural Networks, and Gradient Boosting. In consonance with earlier research, these techniques did not outperform the Random Forest models in terms of accuracy. Support Vector Machines did score comparably to the Random Forest models (accuracy 60% on model 3), but do require more data handling.

Such a question could be included in future versions of the app.

The nine different models with the inclusion of the dummy variables are summarized in Figure 6 and Table 6. Compared to the model presented in Figure 4, similar trends exist. Adding features improves accuracy and public transport modes are better distinguished with the inclusion of context location data. Additionally, the inclusion of these dummy variables leads to higher overall accuracy, and better sensitivities in all models. With the inclusion of these dummy variables, the mean accuracy of the nine different models, on average, rises by 4.8%. In the first three models the increased accuracy is the highest and, for the third model it is 6%. Interestingly, adding extra features actually decreases the sensitivity of the E-bike class. The macro-averaged F1 statistic, on average, rises by 0.045 points (0.059 for the third model). We furthermore see that especially the sensitivity of E-bikes (+.24), bikes (+.11), and scooter (+.15) and tram (+.04) in the third model improves when compared to the same model without these dummy variables. These might be modes that respondents either use regularly or not at all. Few respondents will use both a bike and an E-bike regularly within the data collection period: someone might either use a tram to commute to work or not at all. For the current respondents, we see that the ratio of bike to E-bike trips is 5.3 for all labelled trips, but 19.1 for respondents who used a bike at least once, and 0.4 for respondents who used an E-bike for at least one trip. Similarly, only 9 respondents were responsible for all tram trips, and for these respondents, tram trips made up 12% of their trips, compared to .4% for all respondents. This also explains why the increase of accuracy (1% in the final model) and increased F1 statistic (+ 0.015) compared to the model without these dummy variables diminishes when modes are collapsed.

If, instead of dummy variables on whether a mode was used, the proportion by which a mode was used by a respondent are included as features, the accuracy of the third model rises to 70% and to 87% in the ninth model. To reach these accuracies one does need to observe respondents for a longer time and, therefore, this model is not directly applicable in travel mode prediction studies with new respondents. However, it represents the accuracy of the model once it has observed all trips (travel behaviour) of a respondent. This creates valuable information in the prediction of unlabelled trips.

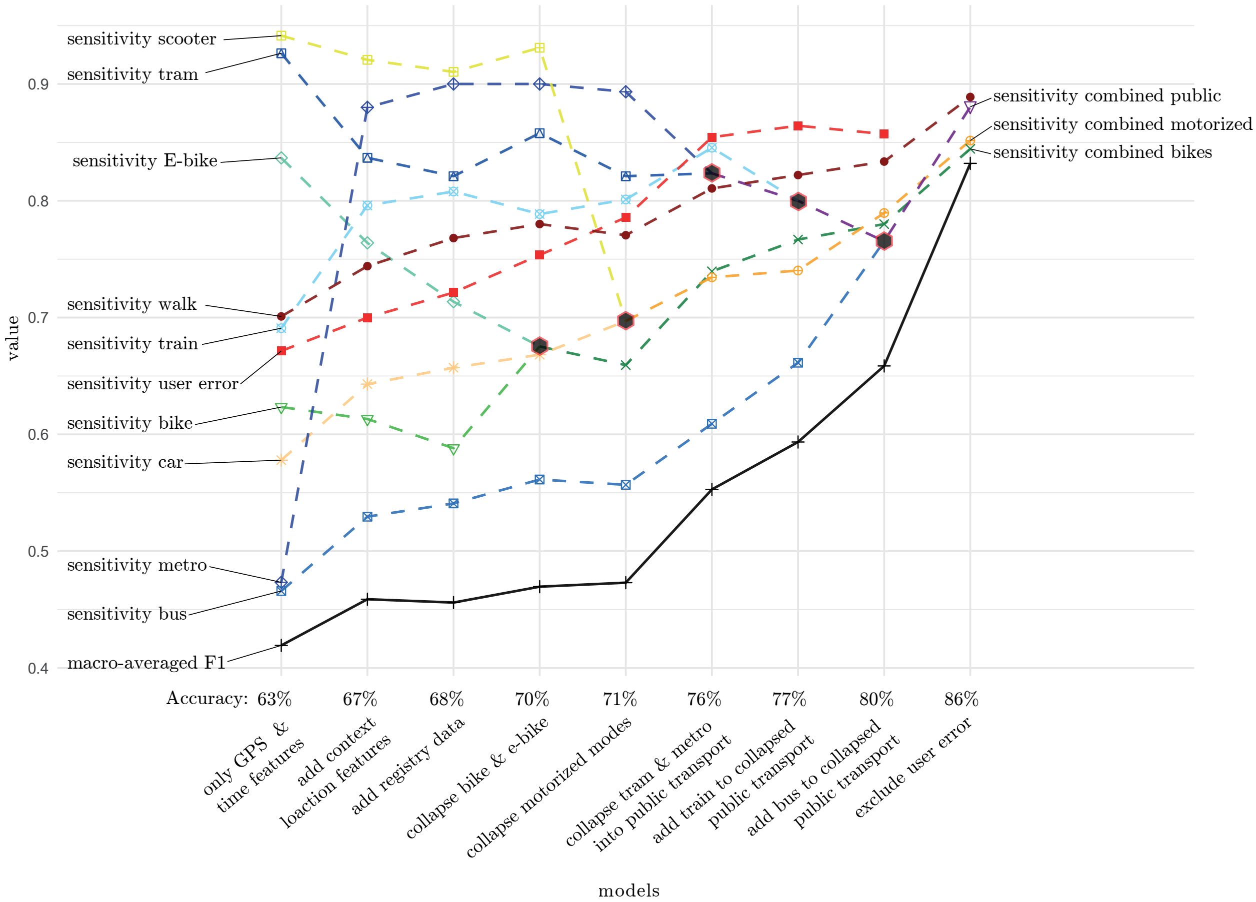

Figure 6: Models with dummy variables of modes used by respondents. The black line represents the macro-averaged F1 statistic of the nine different models represented on the x-axis. The coloured dashed lines represent the sensitivity of the model to different transport modes. The accuracy of the models is listed on the x-axis. In the first three models, additional features are added to the model. In models four to eight different classes of transport modes are collapsed (⬢ indicate when which modes are collapsed). The final model represents the model with collapsed modes, without User errors.

predicted	observed									
	E − bike	*bike*	*car*	*metro*	*bus*	*scooter*	*train*	*tram*	*User error*	*walk*
E − bike	105	90	97	0	1	0	1	0	0	24
bike	14	447	85	0	3	0	0	0	1	42
car	6	27	1427	0	8	1	6	0	2	21
metro	0	13	26	14	0	0	9	2	0	11
bus	2	26	178	0	24	0	4	0	1	7
scooter	2	10	83	0	0	26	0	0	0	4
train	1	4	73	0	2	0	142	0	1	11
tram	2	43	34	1	4	0	6	16	1	29
User error	6	32	70	0	2	1	6	1	20	62
walk	10	69	100	0	2	0	1	1	1	690

Table 6: Confusion Matrix for the third model with all the features and the dummy variables of earlier used modes and no collapsed transport modes yet, accuracy: 68%

4.5 Predicting Unlabelled Trips

The final presented model was used to predict the transport modes for trips where these were missing. In this model the proportions of how often respondents used different transport modes are included as features. 77% of the unlabelled trips came from respondents who labelled at least one trip.

For these respondents the proportions of how often they used different transport modes can be calculated from the trips they did label. For the other 23%, the proportions of used transport modes were predicted using bagged decision trees using the CARET package. Leaving these missing cases out did not yield different results. The results are presented in Table 7. The first column presents the statistics of the labelled trips from the GPS-based study. The second column contains the predicted proportions for the unlabelled trips, and the third column combines both the labelled and unlabelled trips. The unlabelled trips, on average, are shorter in both time and distance. Partly for this reason, there are predicted to be more walking trips in the unlabelled data and fewer car trips. Furthermore, in the labelled trips, there is a consistently lower proportion of public transport modes compared to the predicted unlabelled transport modes. There are expected to be more User errors in the unlabelled data, which is coherent with the more cumbersome method of labelling these erroneously recorded trips. Respondents might be more likely to just leave a trip unlabelled, instead of reporting the error.

Two notes of caution are warranted. First, the model only has a 70% prediction accuracy. Adding to this, the logistic regression presented in Table 5 shows that wrong prediction of transport modes is not equal among transport modes. Compared to walking, E-bike, bike, car, and bus trips are less likely to be predicted correctly, while train trips are more likely to be predicted correctly. Secondly, the prediction model was trained on data from where multi-mode trips and trips with rare transport modes were omitted (5.8% of trips). In the unlabelled data, we can expect these trips to still be present. Unavoidably, these trips are now 'forced' into one of the ten predicted transport modes. One way to quantify, how certain the Random Forest algorithm is about the predictions, is by looking at the average prediction certainty of the most likely transport mode per trip. The Random Forest gives a prediction probability between 0 and 1 for all ten transport modes for every trip. These prediction probabilities sum to one and the model picks the transport mode with the highest probability for a given trip. So, if the model is a sure that a trip was a bike trip it would give the transport mode bike a score of 1 and the 9 other modes a score of 0. If it is not sure at all, it would give all transport modes a score of .1. For the labelled trips, the prediction certainty of the most likely transport mode on average is .51 and .48 for the unlabelled trips. This means that, on average, the algorithm is a bit surer about the predicted transport modes from the labelled trips.

	Labelled trips	Unlabelled trips	Combined labelled and unlabelled trips
median data points	351	233	264
median duration	553 sec	507 sec	525 sec
median distance	2299 m	1758 m	1927 m
% trips $\leq$ 500m	20.4%	26.9%	24.9%
% E-bike trips	3.45%	7.01%*	5.92%*
% bike trips	17.84%	9.31%*	11.91%*
% car trips	50.98%	35.09%*	39.95%*
% metro trips	.35%	2.48 %*	1.83%*
% bus trips	1.04%	8.16%*	5.98%*
% scooter trips	.67%	1.24%*	1.07%*
% train trips	4.13%	6.95%*	6.09%*
% tram trips	.44%	2.47%*	1.86%*
% walk trips	21.10%	27.27%*	25.38%*
% User errors	.65%	2.57%*	1.99%*

Table 7: Summary statistics and proportions of used transport modes of the labelled and unlabelled trips in the dataset. For the statistics and the transport mode proportions, the User error category is excluded from the proportion calculation. This facilitates better comparison with the proportions of used transport modes in the diary-based study presented in Table 8 *Note:** proportions (partly) based on predicted transport modes.

4.6 How do Results Compare to Results of Diary-based Study?

Table 8 shows the comparison of the diary-based ODiN study statistics with the GPS-based study. The first column gives the GPS-based results of those respondents who participated in both the GPS-based study and the diary-based ODiN study. The second column gives the combined labelled and unlabelled trips from them. The third column gives the diary-based results of these same respondents.

Research by Roth [107] shows that there is non-random nonresponse in this app-based travel study and that, particularly, older and lower educated people have relatively lower response rates. In contrast, younger people are often harder to reach with traditional survey methods [108]. These groups of people might have different travel patterns, which would influence the results. Additionally, the studies were conducted at different times of the year, with this app-based survey collecting data in the colder time of the year.

Nonetheless, the results show that transport modes are reported in different proportions in the two study designs. As expected [13–15], the GPS-based study records shorter trips and a larger percentage of short trips (500m or less) compared to the diary-based study. In the GPS-based study, there are more walking and train trips, and considerably fewer bike trips compared to the diary-based study. It can be speculated that people are less likely to forget their bike trips in a diary-based study than their short walks, although is also possible that people used their bikes less and instead used the bus more in the (colder) period of the GPS-based travel study. More research on these dynamics is needed.

	Labelled GPS-based trips from 422 respondents who participated in both studies	Combined (un)labelled GPS-based trips from respondents who participated in both studies	Diary-based trips from respondents who participated in both studies
median data points	344	260	NA
median duration	549 sec	521 sec	1200 sec
median distance	2281 m	2006 m	3000 m
% trips $\leq$ 500m	20.3%	24.1%	12.7%
% E-bike trips	3.35%	7.15 %*	4.72%
% bike trips	17.22%	11.85 %*	22.12%
% car trips	52.23%	41.98 %*	44.44%
% metro trips	.35%	.99 %*	.93%
% bus trips	1.37%	5.96 %*	2.34%
% scooter trips	.93%	1.52 %*	1.24%
% train trips	3.62%	4.46 %*	3.09%
% tram trips	.47%	2.02 %*	.59%
% walk trips	20.45%	24.04 %*	18.47%
% User errors	.65%	1.98%*	NA

Table 8: Summary statistics and proportion for trips by respondents who participated in both the GPS-based study and the latest (diary-based) Dutch Travel Survey [19]. For both studies, multi-mode trips and trips with rare transport modes are excluded and for GPS-based trips the User error category is excluded from the statistics and the proportions calculation. *Note:** proportions partly based on predicted transport modes.

5 Conclusion

In summary, the prediction accuracy of the ten classes of transportation modes within a large-scale travel survey is fruitful. In the most complete model, a 70% accuracy was achieved. Features from GPS data on their own do a reasonable job in separating transport modes when using a Random Forest model (accuracy 52%). However, such a model is not very sensitive in separating similar modes, such as trams and metros. Adding context location data to the model increases accuracy with 8 percentage points, and it especially increases the sensitivity of detection of public transport modes. Including additional registry data improves the model by another 2%. In the model without context location, speed and travel distance are the most important features in distinguishing transport modes. Context information on public transport infrastructure is more important in models where these features are included. When dropping the User error category and merging similar transport modes, this research achieves a prediction accuracy equal to the average prediction accuracy of earlier research which used large samples, of 85%. Nevertheless, to create high-quality official statistics, the distinction of just four transport modes is insufficient. Using nine of the most commonly used categories of transport modes, both in this research and in traditional Dutch Travel Surveys, as well as an extra class of erroneously recorded trips, the accuracy is 62% when treating trips as independent events. Taking into account the transport that respondents used previously, the accuracy of the predictions increases to 68%. Taking into account how often respondents used a certain transport mode increases the accuracy to 70%. Inherent to these GPS-based smartphone studies is the fact that there are erroneously recorded trips. This research, for the first time, aimed to classify these measurement errors. From the increase of accuracy and macro-average F1 statistic in models without these errors, it is evident that User errors are hard to predict with the data collected in the current version of the app. However, simply leaving these out of the model would be disingenuous. Within the GPS-based study, there is a similar trend as exists in diary-based studies of respondents not labelling (reporting) on shorter trips [13–15]. However, GPS-based studies have the advantage that these shorter unlabelled trips are, at least, recorded. The transport modes of these trips can then be imputed. If the goal is to reduce the respondent burden, one could also give a set of the three most likely transport modes belonging to a trip, from which a respondent then needs to pick the correct one. The fact that the Random Forest model gives a prediction probability for all transport modes and not only the most likely transport mode, can be used to check how often the labelled transport mode falls within the top three most likely predicted transport modes. For the model three without the dummy variables the observed transport mode falls in the top three predicted transport modes 86% of the time. For the same model with the dummy variables for used transport modes 89% and for the model with proportions of used transport modes 91% of the time.

6 Discussion

This study showed the feasibility of, at least partly, automatic transport modes prediction in national travel surveys. Even if full automatization of transport mode prediction is not feasible, the results of this research could lead to a reduction in response burden in the future. For example, you could use the fitted model to calculate expected probabilities for different modes and, based on that, only present the top three to the respondent. Alternatively, you could only ask respondents to label trips for which the algorithm is less than 70% certain about the prediction.

However, the current model suggests that it is likely not accurate enough to distinguish between the nine most commonly used transport modes and to also flag erroneously recorded trips. The best results of this research, in which there are only four classes of transport modes, is still 7% lower than earlier research also employing the Random Forest algorithm. There are a few reasons that likely contributed to this discrepancy. The majority of earlier research aimed to answer the question of whether smartphone GPS data could be used to accurately predict transport modes; this research aimed to answer whether smartphone GPS data can be used to accurately predict transport modes in a national travel survey. This is a subtle difference, but an important one. Instead of purposely collecting trips of different transport modes, a large number of respondents for this app was randomly sampled from the Dutch population. The respondents were not trained in the use of the app and only the minimum of data screening was done. The app was also not purposely developed to predict transport modes, but instead to collect data on travel patterns.

More data on the rarer modes, especially scooter, tram, and metro, would likely increase prediction accuracy, as would further improvement of the app. Furthermore, this research assumed labelled trips to be the ground truth with no mislabelling, an assumption that is unlikely to completely hold. Some respondents might have (accidentally) mislabelled trips. Notably, the category User error is likely to be underestimated. Respondents could not delete trips or label them as erroneously recorded, but had to make the effort to leave a comment in the 'other transport modes' text field. Better separation of trips into different segments with a single transport mode is also likely to improve the model accuracy. Future analyses might employ the splitting algorithms designed by Xiao et al. [66] or Tsui et al. [10] to split the multi-mode trips into uni-mode trips. Furthermore, it might be possible to use context location features to split trips, as a train station is a likely point to change transport modes. It is not unlikely that some respondents only reported the dominant transport mode in a trip where multiple modes were used. The app that was used for this research is still in active development and did crash repeatedly for some respondents. Especially for users with older phones and phones with the newest version of iOS, the app was unstable. This might have prevented them from changing a mislabelled trip. Furthermore, the user interface for labelling trips was not yet optimized. For instance, transport mode was only labelled for 45.5% of the trips in the 7 days that the respondents were asked to participate, while stop purposes were labelled 69.4% of the time. The only difference between the two was the interface in the app. As mentioned above, the current version of the app did not (yet) store accelerometer data, which is information that could pick up on more subtle differences in acceleration and vibrations between different modes. Location data from current-day smartphones on its own is most likely not precise enough to calculate these types of features. Reddy et al. [1] found a drop of 20% in accuracy when they either only used the GPS sensor or only the accelerometer sensor compared to using both.

Readers who are better versed in 'traditional' statistical methods than machine learning approaches might take issue with the lack of both a power calculation and discounting of the hierarchical structure of data. Surely, 5,641 trips from 5,641 different individuals would have given more information than the current nesting of data of trips within respondents. Due to nesting, specific respondent level peculiarities, such as an extreme walking speed of a few individuals who labelled a lot of trips, could dominate the data. For certain transport modes, a larger influence of respondent can be expected. For example, a respondent has no influence on the speed and cornering behaviour of public transport modes but does so while walking, riding bikes or E-bikes, and, to a lesser extent, driving a car or scooter. In a hypothetical and extreme case, in which one individual collected all the data, the trained algorithm would not generalize well to trips collected by other respondents, due to individual differences in moving patterns. Little research exists on necessary sample size calculations for classifying transport modes. One exception includes Bolbol et al. [109]. They conclude that, for a study period of two weeks, 11 respondents are needed to classify walking trips and up to 81 respondents are needed to classify bus and car trips. These same authors [110] argue that *"methods using labelled data have achieved accuracies of 90% or more, yet sample sizes and durations were often not adequate to give full accreditation to the result."* However, they only use speed as a feature to distinguish transport modes and not any of the other highly informative features that are often used. They also do not take the nested structure of data into consideration. In the current research, every transport mode was labelled by between 7 (scooter) and 327 (car) different respondents. Only 9 people labelled tram trips, 14 metro trips, and 21 User errors while walking (301) and bikes (186) were the second and third most labelled transport modes. Exact power calculations for classification of nested data using Random Forest are not available and rarely discussed in the literature. Nonetheless, considering the large sample size, both in trips and respondents, this research expects that there was enough information in the data to successfully classify the most common transport modes. However, it can be expected that scooter and User error classification is likely to generalize poorly to other (new) respondents.

Considering this nesting of the data, future research might try to implement recent multilevel extensions of the Random Forest algorithm. This way, in addition to which transport modes someone used before, the algorithm can pick up on respondent specific features, such as someone's walking and biking speed. While Karpievitch et al. [111], showed that, in a simulation study, prediction accuracy was equal for multilevel and single level data for Random Forest, more recent research argues in favour of multilevel extensions of the Random Forest algorithm for both regression and classification. These include stochastic semi-parametric mixed-effects models for longitudinal data [112] and a mixed-effects Random Forest extension of the Random Forest which estimates random part within the framework of the Expectation–Maximization algorithm [113]. The latter is now available as a Python package, but only for regression and not yet for classification (in development) [114, 115].

Zooming out a bit from the issue at hand, it is worth pondering a bit on both the potential avenues automatic transport mode prediction opens and what challenges lie ahead. First of all, both technical obstacles and new opportunities need to be addressed in future research. For example, the app used for this research feuded with different operating systems to keep the flow of location data streaming in. All mobile operating systems use methods to 'kill' apps and background processes. This is done for both privacy reasons and to preserve hard needed battery life. At busy moments or when a respondent did not use its phone for a period of time, phones would simply decide to stop collecting and saving data. Partly due to these restrictions, on average only 12.8 hours of location data per respondent per day (instead of the ideal 24) were collected. This is just one of the many issues researchers in the field of sensor data encounter. This means that instead of, or maybe in addition to, being experts on question design and sampling strategies, future travel survey researchers will also need be acquainted with technical issues related to app development and sensor data. These challenges notwithstanding, an increasingly higher smartphone penetration in most countries and the fast technological development of smartphones offers many new possibilities. The first phones with dual-frequency GPS receivers are currently hitting the market [116]. In combination with the now operational multi-frequency Galileo satellite system, these phones should be able to determine accuracy to 30cm, even indoors [117]. More precise location estimates will likely improve transport mode prediction, especially when paired with other sensor data. Additionally, smartphones' Near Field Communication (NFC) chips are increasingly used for payments, to enter public transport infrastructure, and to unlock ride sharing modes of transport. Knowing when someone enters the metro network or when someone unlocks a shared E-scooter will most definitely help with transport mode prediction. Even if precise and privacy-sensitive data is not available, smartphones logs on when the phone's NFC chip was used, can give researchers import information for prediction models. These more precise location data can not only help to predict transport modes, but also predict the purpose of a stop, especially if the data can be linked to up-to-date land use and OSM data. In a sense, the current use of trips and stops as units of measurement is an artefact from diary-based studies. To get perfect estimates of transport mode use, you need to be able to correctly place a stop between every change of transport modes and you need to define, a priori, how you define a stop. Is, for example, a visit to both a baker and a grocer next to each other, one or two stops? Future research could abandon the paradigm of first separating stops and trips, then engineering features, and then employing a machine learning algorithm. With sufficient data, a convolutional neural network adapted for structured data should be able to pick up on features such as speed and points of interest from the raw location data itself and predict transport modes. Then instead of estimating how many trips a person average made per day and how long those trips were, even more precise statistics could be calculated. For example, how many minutes people walk from home to a bus stop or car, what percentage of time people spend at work, and how many different shops they visit on a day. In such a paradigm, a respondent's day consists of 86,400 seconds that are all spend either travelling or stopping in a certain manner. For every second a transport mode or a stopping activity can be assigned. If one then, afterwards, wants to define trips and stops, that is still possible. As a first step, the results of this study can help reduce respondent burden in app-based travel studies, while at the same time serving as a lead to improve transport mode prediction even further and reconsider how travel surveys are designed and interpreted.

References

[1] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen and Mani Srivastava. 'Using mobile phones to determine transportation modes'. In: *ACM Transactions on Sensor Networks (TOSN)* 6.2 (2010), p. 13 (cit. on pp. 1, 4, 5, 26, 44).

[2] Arash Jahangiri and Hesham A. Rakha. 'Applying machine learning techniques to transportation mode recognition using mobile phone sensor data'. In: *IEEE Transactions on Intelligent Transportation Systems* 16.5 (2015), pp. 2406–2417 (cit. on pp. 1, 6, 47).

[3] Linlin Wu, Biao Yang and Peng Jing. 'Travel mode detection based on gps raw data collected by smartphones: a systematic review of the existing methodologies'. In: *Information* 7.4 (2016), p. 67 (cit. on pp. 1, 3, 4).

[4] Fang Zhao, Ajinkya Ghorpade, Francisco Câmara Pereira, Christopher Zegras and Moshe Ben-Akiva. 'Stop detection in smartphone-based travel surveys'. In: *Transportation Research Procedia* 11 (2015), pp. 218–226 (cit. on pp. 1, 6).

[5] Karst Geurs, Tom Thomas, Marcel Bijlsma and Salima Douhou. 'Automatic trip and mode detection with move smarter: First results from the Dutch mobile mobility panel'. In: *Transportation research procedia* 11 (2015), pp. 247–262 (cit. on pp. 1, 4, 46).

[6] Christopher Zegras, Menghan Li, Talip Kilic, Nancy Lozano-Gracia, Ajinkya Ghorpade, Marco Tiberti, Ana I. Aguilera and Fang Zhao. 'Assessing the representativeness of a smartphone-based household travel survey in Dar es Salaam, Tanzania'. In: *Transportation* 45.2 (2018), pp. 335–363 (cit. on p. 1).

[7] Leon Stenneth, Ouri Wolfson, Philip S. Yu and Bo Xu. 'Transportation mode detection using mobile phones and GIS information'. In: *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM. 2011, pp. 54–63 (cit. on pp. 1, 2, 4–6, 10, 44).

[8] Akshay Vij and Kalyanaraman Shankari. 'When is big data big enough? Implications of using GPS-based surveys for travel demand analysis'. In: *Transportation Research Part C: Emerging Technologies* 56 (2015), pp. 446–462 (cit. on p. 1).

[9] Hongmian Gong, Cynthia Chen, Evan Bialostozky and Catherine T. Lawson. 'A GPS/GIS method for travel mode detection in New York City'. In: *Computers, Environment and Urban Systems* 36.2 (2012), pp. 131–139 (cit. on pp. 1, 2, 45).

[10] Sheung Y.A. Tsui and Amer S. Shalaby. 'Enhanced system for link and mode identification for personal travel surveys based on global positioning systems'. In: *Transportation Research Record* 1972.1 (2006), pp. 38–45 (cit. on pp. 1, 25, 43).

[11] Stephen Greaves, Adrian Ellison, Richard Ellison, Dean Rance, Chris Standen, Chris Rissel and Melanie Crane. 'A web-based diary and companion smartphone app for travel/activity surveys'. In: *Transportation Research Procedia* 11 (2015), pp. 297–310 (cit. on pp. 1, 3).

[12] Hamid Safi, Behrang Assemi, Mahmoud Mesbah and Luis Ferreira. 'An empirical comparison of four technology-mediated travel survey methods'. In: *Journal of traffic and transportation engineering (English edition)* 4.1 (2017), pp. 80–87 (cit. on p. 2).

[13] Bat-hen Nahmias-Biran, Yafei Han, Shlomo Bekhor, Fang Zhao, Christopher Zegras and Moshe Ben-Akiva. 'Enriching Activity-Based Models using

Smartphone-Based Travel Surveys'. In: *Transportation Research Record* 2672.42 (2018), pp. 280–291 (cit. on pp. 2, 23, 25).

[14] Stephen Roddis, Stephan Winter, Ronny Kutadinata, Rahul D. Das, Farzad Khodadadi, Maria Vasardani, Fang Zhao, W. Ko and Nicole Ronald. 'Victorian future mobility sensing (FMS) trial'. In: *Project Report. Department of Infrastructure Engineering, The University of Melbourne* (2016) (cit. on pp. 2, 4, 8, 23, 25).

[15] Wendy Bohte and Kees Maat. 'Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands'. In: *Transportation Research Part C: Emerging Technologies* 17.3 (2009), pp. 285–297 (cit. on pp. 2, 3, 5, 6, 23, 25, 43).

[16] Dick Ettema, Harry J.P. Timmermans and Leo van Veghel. *Effects of data collection methods in travel and activity research*. Tech. rep. 00740871. Institute for Road Safety Research, SWOV, 1996, p. 109. URL: `https://trid.trb.org/view/573629` (cit. on p. 2).

[17] Peter R. Stopher. 'Use of an activity-based diary to collect household travel data'. In: *Transportation* 19.2 (1992), pp. 159–176 (cit. on p. 2).

[18] *Tabi*. 2018. URL: `https://tabiapp.eu` (visited on 2018-09-15) (cit. on p. 2).

[19] Statistics Netherlands (CBS). *Onderzoek Onderweg in Nederland (ODiN) 2018 Concept eindrapportage*. Tech. rep. 2019 (cit. on pp. 3, 13, 24).

[20] Stacey Bricka, Johanna Zmud, Jean Wolf and Joel Freedman. 'Household travel surveys with GPS: An experiment'. In: *Transportation Research Record* 2105.1 (2009), pp. 51–56 (cit. on p. 3).

[21] Shinji Itsubo and Eiji Hato. *Effectiveness of household travel survey using GPS-equipped cell phones and Web diary: Comparative study with paper-based travel survey*. Tech. rep. 06-0701. Transportation Research Board, 2006, p. 13. URL: `https://trid.trb.org/view/776475` (cit. on p. 3).

[22] Paul Kelly, Patricia Krenn, Sylvia Titze, Peter R. Stopher and Charlie Foster. 'Quantifying the difference between self-reported and global positioning systems-measured journey durations: a systematic review'. In: *Transport Reviews* 33.4 (2013), pp. 443–459 (cit. on p. 3).

[23] Stephan Krygsman and Johanna Nel. 'The use of global positioning devices in travel surveys-a developing country application'. In: *Proceedings of the 28th Southern African Transport Conference (SATC 2009)*. Vol. 6. 2009, p. 9 (cit. on p. 3).

[24] Philippe Marchal, Sophie Roux, Shuning Yuan, Jean-Paul Hubert, Jimmy Armoogum, Jean-Loup Madre and Martin Lee-Gosselin. 'A study of non-response in the GPS sub-sample of the French national travel survey 2007–08'. In: *Proceedings of the 8th International Conference on Survey Methods in Transport, France*. 2008, pp. 25–31 (cit. on p. 3).

[25] Marcelo G. Simas Oliveira, Peter Vovsha, Jean Wolf, Yehoshua Birotker, Danny Givon and Julie Paasche. 'Global Positioning System–Assisted Prompted Recall Household Travel Survey to Support Development of Advanced Travel Model in Jerusalem, Israel'. In: *Transportation Research Record* 2246.1 (2011), pp. 16–23 (cit. on p. 3).

[26] Dominik Papinski, Darren M. Scott and Sean T. Doherty. 'Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 12.4 (2009), pp. 347–358 (cit. on p. 3).

[27] Thomas K. Rasmussen, Jesper B. Ingvardson, Katrín Halldórsdóttir and Otto A. Nielsen. 'Using wearable GPS devices in travel surveys: A case study in the Greater Copenhagen Area'. In: *Proceedings from the Annual Transport Conference at Aalborg University*. 2013, pp. 1603–9696 (cit. on p. 3).

[28] Stefan Schönfelder, Kay W. Axhausen, Nicolas Antille and Michel Bierlaire. 'Exploring the potentials of automatically collected GPS data for travel behaviour analysis: A Swedish data source'. In: *Arbeitsberichte Verkehrs- und Raumplanung* 124 (2002) (cit. on p. 3).

[29] Nadine Schüssler and Kay W. Axhausen. 'Processing raw data from global positioning systems without additional information'. In: *Transportation Research Record* 2105.1 (2009), pp. 28–36 (cit. on p. 3).

[30] Peter R. Stopher, Claudine J. Moutou and Wen Liu. *Sustainability of voluntary travel behaviour change initiatives: a 5-year study*. Tech. rep. Working Paper ITLS-WP-13-12. Institute of Transport and Logistics Studies, 2013. URL: `https://ses.library.usyd.edu.au/bitstream/2123/19201/1/ITLS-WP-13-12.pdf` (cit. on p. 3).

[31] Peter R. Stopher and L. Wargelin. 'Conducting a household travel survey with GPS: Reports on a pilot study'. In: *World Congress on Transport Research, 12th, 2010, Lisbon, Portugal*. 03323. 2010 (cit. on p. 3).

[32] Xianyu Jianchuan, Juan Zhicai, Xiao Guangnian and Fu Xuemei. 'Smartphone-based travel survey: A pilot study in China'. In: *Mobile technologies for activity-travel data collection and analysis*. IGI Global, 2014, pp. 209–223 (cit. on p. 3).

[33] Philippe Marchal and Thao Pham. 'Comparison of conventional versus GPS-based data collection in the French National Travel Survey'. In: *Proceedings of the NTTS (New Techniques and Technologies for Statistics)* (2013) (cit. on p. 3).

[34] Jean Wolf, Randall Guensler and William Bachman. 'Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data, 2001'. In: *Transportation Research Board 80th Annual Meeting*, pp. 7–11 (cit. on p. 3).

[35] Google. *Android API reference*. URL: `https://developer.android.com/reference/android/location/Location` (visited on 2019-06-07) (cit. on p. 3).

[36] *Location accuracy defined - iOS Stack Exchange forum*. URL: `https://stackoverflow.com/questions/30673627/location-accuracy-defined-ios` (visited on 2019-06-07) (cit. on p. 3).

[37] Celestino Ordonez, Javier Martínez, José Ramón Rodriguez-Pérez and A. Reyes. 'Detection of outliers in GPS measurements by using functional-data analysis'. In: *Journal of Surveying Engineering* 137.4 (2011), pp. 150–155 (cit. on p. 4).

[38] Stefan S. Ivanović, Ana-Maria O. Raimond, Sébastien Mustière and Thomas Devogele. 'Detection of outliers in crowdsourced GPS traces'. In: *Spatial Accuracy 2016 Symposium*. 2016 (cit. on p. 4).

[39] Thirugnanam T. Chelvam and Vijaya Samundeeswari. 'Processing Outliers in GPS Trajectory Data Set'. In: *Proceedings of International Conference on Electronics*. Vol. 96. 2012 (cit. on p. 4).

[40] Sina Dabiri and Kevin Heaslip. 'Inferring transportation modes from GPS trajectories using a convolutional neural network'. In: *Transportation Research Part C: Emerging Technologies* 86 (2018), pp. 360–371 (cit. on pp. 4, 5, 10, 49).

[41] Ronald W. Schafer. 'What is a Savitzky-Golay filter'. In: *IEEE Signal processing magazine* 28.4 (2011), pp. 111–117 (cit. on p. 4).

[42] Donald J. McLean and Marta A. V. Skowron. 'Trajr: an R package for characterisation of animal trajectories'. In: *Ethology* 124.6 (2018), pp. 440–448 (cit. on p. 4).

[43] Jouni Hartikainen, Arno Solin and Simo Särkkä. 'Optimal filtering with Kalman filters and smoothers'. In: *Department of Biomedica Engineering and Computational Sciences, Aalto University School of Science, 16th August* (2011) (cit. on p. 4).

[44] Boaz Sobrado. *Implementing a Kalman filter in R for Android GPS measurements*. 2017-08-29. URL: `http://boazsobrado.com/blog/2017/08/29/implementing-a-kalman-filter-in-r-for-android-gps-measurements/` (visited on 2018-11-01) (cit. on p. 4).

[45] Donald J. Patterson, Lin Liao, Dieter Fox and Henry Kautz. 'Inferring high-level behavior from low-level sensors'. In: *International Conference on Ubiquitous Computing*. Springer. 2003, pp. 73–89 (cit. on pp. 4, 43).

[46] Philip J. Troped, Marcelo S. Oliveira, Charles E. Matthews, Ellen K. Cromley, Steven J. Melly and Bruce A. Craig. 'Prediction of activity mode with global positioning system and accelerometer data.' In: *Medicine and science in sports and exercise* 40.5 (2008), pp. 972–978 (cit. on pp. 4, 43).

[47] Luca Bedogni, Marco Di Felice and Luciano Bononi. 'By train or by car? Detecting the user's motion type through smartphone sensors data'. In: *2012 IFIP Wireless Days*. IEEE. 2012, pp. 1–6 (cit. on pp. 4, 6, 45).

[48] Tao Feng and Harry J.P. Timmermans. 'Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using GPS data'. In: *Transportation Planning and Technology* 39.2 (2016), pp. 180–194 (cit. on pp. 4–6, 48).

[49] Bryan Martin, Vittorio Addona, Julian Wolfson, Gediminas Adomavicius and Yingling Fan. 'Methods for real-time prediction of the mode of travel using smartphone-based GPS and accelerometer data'. In: *Sensors* 17.9 (2017), p. 2058 (cit. on pp. 4, 6, 49).

[50] Rahul D. Das and Stephan Winter. 'Detecting urban transport modes using a hybrid knowledge driven framework from GPS trajectory'. In: *ISPRS International Journal of Geo-Information* 5.11 (2016), p. 207 (cit. on pp. 4, 48).

[51] David Montoya, Serge Abiteboul and Pierre Senellart. 'Hup-me: Inferring and reconciling a timeline of user activity from rich smartphone data'. In: *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM. 2015, p. 62 (cit. on pp. 4, 46).

[52] Ian A.H. Muller. 'Practical activity recognition using gsm data'. In: *Proceedings of the 5th International Semantic Web Conference (ISWC). Athens*. Vol. 1. 8. Citeseer. 2006 (cit. on pp. 5, 43).

[53] Min Mun, Deborah Estrin, Jeff Burke and Mark Hansen. 'Parsimonious mobility classification using GSM and WiFi traces'. In: *Proceedings of the Fifth Workshop on Embedded Networked Sensors (HotEmNets)*. Citeseer. 2008 (cit. on pp. 5, 43).

[54] Lin Liao, Donald J. Patterson, Dieter Fox and Henry Kautz. 'Learning and inferring transportation routines'. In: *Artificial Intelligence* 171.5-6 (2007), pp. 311–331 (cit. on pp. 5, 43).

[55] Timothy Sohn, Alex Varshavsky, Anthony LaMarca, Mike Y. Chen, Tanzeem Choudhury, Ian Smith, Sunny Consolvo, Jeffrey Hightower, William G. Griswold and Eyal De Lara. 'Mobility detection using everyday GSM traces'. In: *International Conference on Ubiquitous Computing*. Springer. 2006, pp. 212–224 (cit. on pp. 5, 43).

[56] Shuangquan Wang, Canfeng Chen and Jian Ma. 'Accelerometer based transportation mode recognition on mobile phones'. In: *2010 Asia-Pacific Conference on Wearable Computing Systems*. IEEE. 2010, pp. 44–46 (cit. on pp. 5, 44).

[57] Apichon Witayangkurn, Teerayut Horanont, Natsumi Ono, Yoshihide Sekimoto and Ryosuke Shibasaki. 'Trip reconstruction and transportation mode extraction on low data rate GPS data from mobile phone'. In: *Proceedings of the international conference on computers in urban planning and urban management (CUPUM 2013)*. 2013, pp. 1–19 (cit. on pp. 5, 45).

[58] Tao Feng and Harry J.P. Timmermans. 'Transportation mode recognition using GPS and accelerometer data'. In: *Transportation Research Part C: Emerging Technologies* 37 (2013), pp. 118–130 (cit. on pp. 5, 45).

[59] Meng-Chieh Yu, Tong Yu, Shao-Chen Wang, Chih-Jen Lin and Edward Y. Chang. 'Big data small footprint: The design of a low-power classifier for detecting transportation modes'. In: *Proceedings of the VLDB Endowment* 7.13 (2014), pp. 1429–1440 (cit. on pp. 5, 46).

[60] Rahul C. Shah, Chieh-yih Wan, Hong Lu and Lama Nachman. 'Classifying the mode of transportation on mobile phones using GIS information'. In: *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. ACM. 2014, pp. 225–229 (cit. on pp. 5, 46).

[61] Samuli Hemminki, Petteri Nurmi and Sasu Tarkoma. 'Accelerometer-based transportation mode detection on smartphones'. In: *Proceedings of the 11th ACM conference on embedded networked sensor systems*. ACM. 2013, p. 13 (cit. on pp. 5, 45).

[62] Shih-Hau Fang, Hao-Hsiang Liao, Yu-Xiang Fei, Kai-Hsiang Chen, Jen-Wei Huang, Yu-Ding Lu and Yu Tsao. 'Transportation modes classification using sensors on smartphones'. In: *Sensors* 16.8 (2016), p. 1324 (cit. on pp. 5, 48).

[63] Xiaolu Zhou, Wei Yu and William C. Sullivan. 'Making pervasive sensing possible: Effective travel mode sensing based on smartphones'. In: *Computers, Environment and Urban Systems* 58 (2016), pp. 52–59 (cit. on pp. 5, 48).

[64] Mehmet Guvensan, Burak Dusun, Baris Can and Irem Turkmen. 'A Novel Segment-Based Approach for Improving Classification Performance of Transport Mode Detection'. In: *Sensors* 18.1 (2018), p. 87 (cit. on pp. 5, 49).

[65] Thanos Bantis and James Haworth. 'Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics'. In: *Transportation Research Part C: Emerging Technologies* 80 (2017), pp. 286–309 (cit. on pp. 5, 49).

[66] Guangnian Xiao, Zhicai Juan and Chunqin Zhang. 'Travel mode detection based on GPS track data and Bayesian networks'. In: *Computers, Environment and Urban Systems* 54 (2015), pp. 14–22 (cit. on pp. 5, 25, 47).

[67] Bao Wang, Linjie Gao and Zhicai Juan. 'Travel mode detection using gps data and socioeconomic attributes based on a random forest classifier'. In: *IEEE Transactions on Intelligent Transportation Systems* 19.5 (2018), pp. 1547–1558 (cit. on pp. 5, 49).

[68] *Zonnig 2018 stuwt omzet fietsbranche naar record*. https://raivereniging.nl/artikel/persberichten/2019-q1/0301-zonnig-2018-stuwt-omzet-fietsbranche-naar-record.html. Accessed: 2019-03-11 (cit. on p. 5).

[69] Xing Su, Hernan Caceres, Hanghang Tong and Qing He. 'Online travel mode identification using smartphones with battery saving considerations'. In: *IEEE Transactions on Intelligent Transportation Systems* 17.10 (2016), pp. 2921–2934 (cit. on pp. 5, 48).

[70] Behrang Assemi, Hamid Safi, Mahmoud Mesbah and Luis Ferreira. 'Developing and validating a statistical model for travel mode identification on smartphones'. In: *IEEE Transactions on Intelligent Transportation Systems* 17.7 (2016), pp. 1920–1931 (cit. on pp. 5, 48).

[71] Zhibin Xiao, Yang Wang, Kun Fu and Fan Wu. 'Identifying different transportation modes from trajectory data using tree-based ensemble classifiers'. In: *ISPRS International Journal of Geo-Information* 6.2 (2017), p. 57 (cit. on pp. 5, 49).

[72] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie and Wei-Ying Ma. 'Understanding mobility based on GPS data'. In: *Proceedings of the 10th International Conference on Ubiquitous Computing*. ACM. 2008, pp. 312–321 (cit. on pp. 5, 43).

[73] Young-Ji Byon, Baher Abdulhai and Amer S. Shalaby. 'Real-time transportation mode detection via tracking global positioning system mobile devices'. In: *Journal of Intelligent Transportation Systems* 13.4 (2009), pp. 161–170 (cit. on pp. 5, 18, 43).

[74] Anastasia Moiseeva, Joran Jessurun and Harry J.P. Timmermans. 'Semi-automatic imputation of activity-travel diaries using GPS traces, prompted recall and context-sensitive learning algorithms'. In: *Proceedings of the 89th TRB Annual Meeting, Washington, DC*. 2010 (cit. on pp. 5, 44).

[75] Muhammad A. Shafique and Eiji Hato. 'Modelling of accelerometer data for travel mode detection by hierarchical application of binomial logistic regression'. In: *Transportation Research Procedia* 10 (2015), pp. 236–244 (cit. on pp. 5, 47).

[76] Mohammad Etemad, Amilcar S. Júnior and Stan Matwin. 'On feature selection and evaluation of transportation mode prediction strategies'. In: *arXiv preprint arXiv:1808.03096* (2018) (cit. on pp. 6, 50).

[77] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro and Dinani Amorim. 'Do we need hundreds of classifiers to solve real world classification problems?' In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 3133–3181 (cit. on p. 6).

[78] Thomas K. Rasmussen, Jesper B. Ingvardson, Katrín Halldórsdóttir and Otto A. Nielsen. 'Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area'. In: *Computers, Environment and Urban Systems* 54 (2015), pp. 301–313 (cit. on pp. 6, 47).

[79] Rahul D. Das, Nicole Ronald and Stephan Winter. 'Clustering based transfer detection with fuzzy activity recognition from smart-phone GPS trajectories'. In: *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 2014, pp. 3138–3143 (cit. on p. 6).

[80] Heikki Mäenpää, Andrei Lobov and Jose L. M. Lastra. 'Travel mode estimation for multi-modal journey planner'. In: *Transportation Research Part C: Emerging Technologies* 82 (2017), pp. 273–289 (cit. on pp. 6, 49).

[81] Laurent Smeets. URL: https://github.com/LaurentSmeets/Master-Thesis (cit. on pp. 7, 18).

[82] Tabi. URL: https://gitlab.com/tabi/tabi-app (cit. on p. 7).

[83] Ken Thompson. 'Programming techniques: Regular expression search algorithm'. In: *Communications of the ACM* 11.6 (1968), pp. 419–422 (cit. on p. 9).

[84] Peter Widhalm, Philippe Nitsche and Norbert Brändie. 'Transport mode detection with realistic smartphone sensor data'. In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE. 2012, pp. 573–576 (cit. on pp. 10, 45).

[85] Roger W. Sinnott. 'Virtues of the Haversine'. In: *Sky Telesc.* 68 (1984), p. 159 (cit. on p. 10).

[86] Eindhoven University of Technology. *Tutorial: Using OpenStreetMap data.* URL: https://www.win.tue.nl/aga/tutorials/openstreetmap/ (cit. on p. 11).

[87] OpenStreetMap. *Planet dump retrieved from https://planet.osm.org.* https://www.openstreetmap.org. 2018 (cit. on p. 11).

[88] Filip Biljecki, Hugo Ledoux and Peter Van Oosterom. 'Transportation mode-based segmentation and classification of movement trajectories'. In: *International Journal of Geographical Information Science* 27.2 (2013), pp. 385–407 (cit. on pp. 11, 45).

[89] Statistics Netherlands. *Bevolkingskernen in Nederland 2011.* Dutch. 2017. URL: https://www.cbs.nl/nl-nl/achtergrond/2014/13/bevolkingskernen-in-nederland-2011 (visited on 2019-04-01) (cit. on p. 11).

[90] Statistics Netherlands. *Wijk- en buurtkaart 2017.* Dutch. 2017. URL: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische\%20data/wijk-en-buurtkaart-2017 (visited on 2019-04-01) (cit. on p. 11).

[91] Leo Breiman. 'Random Forests'. In: *Machine learning* 45.1 (2001), pp. 5–32 (cit. on p. 11).

[92] Max Kuhn. 'Building predictive models in R using the caret package'. In: *Journal of statistical software* 28.5 (2008), pp. 1–26 (cit. on p. 11).

[93] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria, 2019. URL: https://www.R-project.org (cit. on p. 11).

[94] Mousumi Banerjee, Ying Ding and Anne-Michelle Noone. 'Identifying representative trees from ensembles'. In: *Statistics in medicine* 31.15 (2012), pp. 1601–1616 (cit. on p. 11).

[95] Philipp Probst, Marvin N. Wright and Anne-Laure Boulesteix. 'Hyperparameters and tuning strategies for random forest'. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (2019), p. 1301 (cit. on p. 11).

[96] Matthew W. Mitchell. 'Bias of the Random Forest out-of-bag (OOB) error for certain input parameters'. In: *Open Journal of Statistics* 1.03 (2011), p. 205 (cit. on p. 11).

[97] Max Kunhn and Johnson Kjell. *Comparing the Bootstrap and Cross-Validation.* 2014. URL: http://appliedpredictivemodeling.com/blog/2014/11/27/08ks7leh0zof45zpf5vqe56d1sahb0 (cit. on p. 11).

[98] Silke Janitza and Roman Hornung. 'On the overestimation of random forest's out-of-bag error'. In: *PloS one* 13.8 (2018), pp. 1–31 (cit. on p. 11).

[99] Max Kuhn and Kjell Johnson. *Applied predictive modeling.* New York: Springer, 2013 (cit. on p. 12).

[100] Chao Chen, Andy Liaw and Leo Breiman. 'Using random forest to learn imbalanced data'. In: *University of California, Berkeley* 110 (2004), pp. 1–12 (cit. on p. 12).

[101] Arie Ben-David. 'Comparison of classification accuracy using Cohen's Weighted Kappa'. In: *Expert Systems with Applications* 34.2 (2008), pp. 825–832 (cit. on p. 12).

[102] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin and Achim Zeileis. 'Conditional variable importance for random forests'. In: *BMC bioinformatics* 9.1 (2008), p. 307 (cit. on pp. 12, 13).

[103] Carolin Strobl, Anne-Laure Boulesteix and Thomas Augustin. 'Unbiased split selection for classification trees based on the Gini Index'. In: *Computational Statistics & Data Analysis* 52.1 (2007), pp. 483 –501. ISSN: 0167-9473. URL: `http://www.sciencedirect.com/science/article/pii/S0167947306005093` (cit. on p. 13).

[104] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis and Torsten Hothorn. 'Bias in random forest variable importance measures: Illustrations, sources and a solution'. In: *BMC Bioinformatics* 8.1 (2007), p. 25 (cit. on p. 13).

[105] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin and Achim Zeileis. 'Conditional variable importance for random forests'. In: *BMC Bioinformatics* 9.1 (2008), p. 307 (cit. on p. 13).

[106] André Altmann, Laura Toloşi, Oliver Sander and Thomas Lengauer. 'Permutation importance: a corrected feature importance measure'. In: *Bioinformatics* 26.10 (2010), pp. 1340–1347 (cit. on p. 17).

[107] Katalin Roth. 'Analysis of missing data in a longitudinal smartphone survey'. In: *Unpublished Master Thesis* (2019) (cit. on p. 23).

[108] Stacey Bricka. 'Non-response challenges in GPS-based surveys'. In: *8th International Conference on Survey Methods in Transport.* 2008, pp. 25–31 (cit. on p. 23).

[109] Adel Bolbol, Tao Cheng, Ioannis Tsapakis and Andy Chow. 'Sample size calculation for studying transportation modes from GPS data'. In: *Procedia-Social and Behavioral Sciences* 48 (2012), pp. 3040–3050 (cit. on p. 26).

[110] Adel Bolbol, Tao Cheng, Ioannis Tsapakis and James Haworth. 'Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification'. In: *Computers, Environment and Urban Systems* 36.6 (2012), pp. 526–537 (cit. on pp. 26, 45).

[111] Yuliya V. Karpievitch, Elizabeth G. Hill, Anthony P. Leclerc, Alan R. Dabney and Jonas S. Almeida. 'An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++'. In: *PloS one* 4.9 (2009), p. 7087 (cit. on p. 26).

[112] Louis Capitaine, Robin Genuer and Rodolphe Thiébaut. 'Random forests for high-dimensional longitudinal data'. In: *arXiv preprint arXiv:1901.11279* (2019) (cit. on p. 26).

[113] Ahlem Hajjem, François Bellavance and Denis Larocque. 'Mixed-effects random forest for clustered data'. In: *Journal of Statistical Computation and Simulation* 84.6 (2014), pp. 1313–1328 (cit. on p. 27).

[114] Sourav Dey. *Mixed Effects Random Forests in Python.* 2017-10-29. URL: `https://towardsdatascience.com/mixed-effects-random-forests-6ecbb85cb177` (visited on 2019-06-01) (cit. on p. 27).

[115] merf 0.3. *Mixed Effects Random Forest.* 2019-03-28. URL: https://pypi.org/project/merf/ (visited on 2019-06-01) (cit. on p. 27).

[116] Samuel K. Moore. *Superaccurate GPS Chips Coming to Smartphones in 2018.* 2017. URL: https://spectrum.ieee.org/tech-talk/semiconductors/design/superaccurate-gps-chips-coming-to-smartphones-in-2018 (visited on 2019-06-17) (cit. on p. 27).

[117] European Global Navigation Satellite System Agency (GSA). *GNSS chip manufacturers gear up for Galileo roll-out in U.S.* 2019-03-12. URL: https://www.gsa.europa.eu/newsroom/news/gnss-chip-manufacturers-gear-galileo-roll-out-us (visited on 2019-06-17) (cit. on p. 27).

[118] Przemyslaw Biecek. *Ceteris Paribus Plots for classification.* 2018. URL: https://pbiecek.github.io/ceterisParibus/articles/ceteris_paribus_HR.html (visited on 2019-05-01) (cit. on p. 42).

[119] Alex Hayes. *Understanding Multinomial Regression With Partial Dependence Plots.* 2018-10-23. URL: https://www.alexpghayes.com/blog/understanding-multinomial-regression-with-partial-dependence-plots/ (visited on 2019-05-23) (cit. on p. 42).

[120] Eui-Hwan Chung and Amer S. Shalaby. 'A trip reconstruction tool for GPS-based personal travel surveys'. In: *Transportation Planning and Technology* 28.5 (2005), pp. 381–401 (cit. on p. 43).

[121] Yu Zheng, Like Liu, Longhao Wang and Xing Xie. 'Learning transportation mode from raw gps data for geographic applications on the web'. In: *Proceedings of the 17th International Conference on World Wide Web.* ACM. 2008, pp. 247–256 (cit. on p. 43).

[122] Christian Rudloff and Markus Ray. *Detecting travel modes and profiling commuter habits solely based on GPS data.* Tech. rep. 10-0634. Transportation Research Board, 2010, p. 15. URL: https://trid.trb.org/view/909563 (cit. on p. 44).

[123] Vincenzo Manzoni, Diego Maniloff, Kristian Kloeckl and Carlo Ratti. 'Transportation mode identification and real-time CO2 emission estimation using smartphones'. In: *SENSEable City Lab, Massachusetts Institute of Technology, nd* (2010) (cit. on p. 44).

[124] Theresa Nick, Edmund Coersmeier, Jan Geldmacher and Juergen Goetze. 'Classifying means of transportation using mobile sensor data'. In: *The 2010 International Joint Conference on Neural Networks (IJCNN).* IEEE. 2010, pp. 1–6 (cit. on p. 44).

[125] Paola A. Gonzalez, Jeremy S. Weinstein, Sean J. Barbeau, Miguel A. Labrador, Philip L. Winters, Nevine L. Georggi and Roxana Perez. 'Automating mode detection for travel behaviour analysis by using global positioning systems-enabled mobile phones and neural networks'. In: *IET Intelligent Transport Systems* 4.1 (2010), pp. 37–49 (cit. on p. 44).

[126] Sasank Reddy, Jeff Burke, Deborah Estrin, Mark Hansen and Mani Srivastava. 'Determining transportation mode on mobile phones'. In: *2008 12th IEEE International Symposium on Wearable Computers.* IEEE. 2008, pp. 25–28 (cit. on p. 44).

[127] Lijuan Zhang, Sagi Dalyot, Daniel Eggert and Monika Sester. 'Multi-stage approach to travel-mode segmentation and classification of gps traces'. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII-4/W25 (2011), pp. 87–93 (cit. on p. 44).

[128] Nadine Schüssler, Lara Montini and Christoph Dobler. 'Improving post-processing routines for GPS observations using prompted-recall data'. In: *Working paper Transport and Spatial Planning* 724 (2011) (cit. on p. 44).

[129] Dafeng Xu, Guojie Song, Peng Gao, Rongzeng Cao, Xinwei Nie and Kunqing Xie. 'Transportation Modes Identification from Mobile Phone Data Using Probabilistic Models'. In: *Advanced Data Mining and Applications*. Ed. by Jie Tang, Irwin King, Ling Chen and Jianyong Wang. Berlin, Heidelberg: Springer, 2011, pp. 359–371. ISBN: 978-3-642-25855-8 (cit. on p. 45).

[130] Philippe Nitsche, Peter Widhalm, Simon Breuss and Peter Maurer. 'A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys'. In: *Procedia-Social and Behavioral Sciences* 48 (2012), pp. 1033–1046 (cit. on p. 45).

[131] Richard Brunauer, Michael Hufnagl, Karl Rehrl and Andreas Wagner. 'Motion pattern analysis enabling accurate travel mode detection from GPS data only'. In: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE. 2013, pp. 404–411 (cit. on p. 46).

[132] Xun Li. 'Using Complexity Measures of Movement for Automatically Detecting Movement Types of Unknown GPS Trajectories'. In: *American Journal of Geographic Information System* 3 (2014-01), pp. 63–74 (cit. on p. 46).

[133] Lara Montini, Nadine Rieser-Schüssler and Kay W. Axhausen. 'Personalisation in multi-day GPS and accelerometer data processing'. In: *14th Swiss Transport Research Conference (STRC)*. 2014 (cit. on p. 46).

[134] Adrian C. Prelipcean, Győző Gidófalvi and Yusak O. Susilo. 'Mobility collector'. In: *Journal of Location Based Services* 8.4 (2014), pp. 229–255 (cit. on p. 46).

[135] Young-Ji Byon and Steve Liang. 'Real-time transportation mode detection using smartphones and artificial neural networks: Performance comparisons between smartphones and conventional global positioning system sensors'. In: *Journal of Intelligent Transportation Systems* 18.3 (2014), pp. 264–272 (cit. on p. 46).

[136] Jingmin Chen and Michel Bierlaire. 'Probabilistic multimodal map matching with rich smartphone data'. In: *Journal of Intelligent Transportation Systems* 19.2 (2015), pp. 134–148 (cit. on p. 47).

[137] Michel Bierlaire, Jingmin Chen and Jeffrey Newman. 'A probabilistic map matching method for smartphone GPS data'. In: *Transportation Research Part C: Emerging Technologies* 26 (2013), pp. 78–98 (cit. on p. 47).

[138] Zahra A. Lari and Amir Golroo. 'Automated transportation mode detection using smart phone applications via machine learning: Case study mega city of Tehran'. In: *Proceedings of the Transportation Research Board 94th Annual Meeting, Washington, DC, USA*. 2015, pp. 11–15 (cit. on p. 47).

[139] Guangnian Xiao, Zhicai Juan and Jingxin Gao. 'Travel mode detection based on neural networks and particle swarm optimization'. In: *Information* 6.3 (2015b), pp. 522–535 (cit. on p. 47).

[140] Muhammad A. Shafique and Eiji Hato. 'Use of acceleration data for transportation mode prediction'. In: *Transportation* 42.1 (2015), pp. 163–188 (cit. on p. 47).

[141] Xiaolu Zhu, Jinglin Li, Zhihan Liu, Shangguang Wang and Fangchun Yang. 'Learning transportation annotated mobility profiles from gps data for context-aware mobile services'. In: *2016 IEEE International Conference on Services*

Computing (SCC). IEEE. 2016, pp. 475–482 (cit. on p. 47).

[142] Tim Sonderen. 'Detection of transportation mode solely using smartphones'. In: *Proc. 20th Student Conf. IT*. 2016, pp. 1–7 (cit. on p. 47).

[143] Adrian C. Prelipcean, Gyözö Gidofalvi and Yusak O. Susilo. 'Measures of transport mode segmentation of trajectories'. In: *International Journal of Geographical Information Science* 30.9 (2016), pp. 1763–1784 (cit. on p. 48).

[144] Yuki Endo, Hiroyuki Toda, Kyosuke Nishida and Akihisa Kawanobe. 'Deep feature extraction from trajectories for transportation mode estimation'. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer. 2016, pp. 54–66 (cit. on p. 48).

[145] Hao Wang, GaoJun Liu, Jianyong Duan and Lei Zhang. 'Detecting transportation modes using deep neural network'. In: *IEICE Transactions on Information and Systems* 100.5 (2017), pp. 1132–1135 (cit. on p. 49).

[146] Rahul D. Das and Stephan Winter. 'A fuzzy logic based transport mode detection framework in urban environment'. In: *Journal of Intelligent Transportation Systems* 22.6 (2018), pp. 478–489 (cit. on p. 50).

[147] Qiuhui Zhu, Min Zhu, Mingzhao Li, Min Fu, Zhibiao Huang, Qihong Gan and Zhenghao Zhou. 'Transportation modes behaviour analysis based on raw GPS dataset.' In: *IJES* 10.2 (2018), pp. 126–136 (cit. on p. 50).

[148] Mohammad Etemad, Amilcar S. Júnior and Stan Matwin. 'Predicting transportation modes of GPS trajectories using feature engineering and noise removal'. In: *Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence, Canadian AI 2018, Toronto, ON, Canada, May 8–11, 2018, Proceedings 31*. Springer. 2018, pp. 259–264 (cit. on p. 50).

[149] Alessio D. Marra, Henrik Becker, Francesco Corman and Kay W. Axhausen. 'Developing a passive GPS tracking system to study long-term travel behavior'. In: *Arbeitsberichte Verkehrs-und Raumplanung* 1393 (2018) (cit. on p. 50).

[150] Tao Feng and Harry J.P. Timmermans. 'Integrated imputation of activity-travel diaries incorporating the measurement of uncertainty'. In: *Transportation Planning and Technology* (2019), pp. 1–19 (cit. on p. 50).

[151] Zhanbo Sun and Xuegang Jeff Ban. 'Vehicle classification using GPS data'. In: *Transportation Research Part C: Emerging Technologies* 37 (2013), pp. 102–117 (cit. on p. 50).

A Appendix

A.1 Formulas for GPS-based Features

Total distance is calculated as:

$$total\ distance = \sum_{i=trip.s}^{trip.e-1} dist(P_i, P_{i+1}),$$

where $trip.s$ and $trip.e$ represents the start and end of a trip and $dist(P_i, P_{i+1})$ represents the distance between P_i and P_{i+1}.

The straight-line distance between the start and end point is just the $dist(P_{trip.s}, P_{trip.e})$. One can also take the ratio between straight line distance and total distance as a statistic for the straightness of a trip. The speed ($\bar{v}$) between two locations can be expressed as:

$$\bar{v}_i = \frac{dist(P_i, P_{i+1})}{t_{i+1} - t_i},$$

in which P_i and t_i are the location and time at moment i. Consequently, acceleration (or deceleration for negative acceleration) is expressed as:

$$\bar{a}_i = \frac{\bar{v}_{i+1} - \bar{v}_i}{t_{i+1} - t_i}$$

Heading direction change (hdc) of three consecutive points, such as $P_1.hdc$, between P_1, P_2, and P_3 in Figure 7, is calculated as:

$$P_i.hdc = |P_i.head - P_{i+1}.head|$$

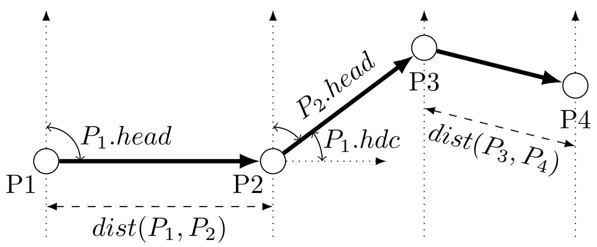

Figure 7: Visual representation of how heading change (turns) and distance are calculated.

A.2 Table with Features and Table with Summaries of Statistics per Transport Mode

model 1 (raw location features)		model 2 (add context location features)		model 3+ (add registry features)
1 total distance	31 mean accuracy	58 mean distance train station	88 near subway stop at start	118 gender
2 distance straight line	32 median accuracy	59 median distance train station	89 near subway stop at end	119 age
3 ratio total dist. dist. straight line	33 % filtered data points	60 mean distance train track	90 near subway stop at start & end	120 car ownership
4 mean distance between points	34 mean heading change	61 median distance train track	91 mean distance tram stop	121 scooter ownership
5 total time	35 median heading change	62 unique train station passed	92 median distance tram stop	122 driver's license
6 median speed	36 max heading change	63 rel. prox. train station	93 mean distance tram track	123 leaseauto
7 mean speed	37 min heading change	64 rel. prox. train track	94 median distance tram track	124 income percentile
8 max speed	38 heading change $P_{5\%}$	65 n points near train station	95 unique tram stops passed	125 urbanicity home address
9 min speed	39 heading change $P_{25\%}$	66 n points near train track	96 rel. prox. tram stop	126 urbanicity locality
10 speed $P_{5\%}$	40 heading change $P_{75\%}$	67 near train station at start	97 rel. prox. tram track	127 social economic status
11 speed $P_{25\%}$	41 heading change $P_{90\%}$	68 near train station at end	98 n points near tram stop	
12 speed $P_{75\%}$	42 heading change $P_{95\%}$	69 near train station at start & end	99 n points near tram track	
13 speed $P_{90\%}$	43 heading change $P_{99\%}$	70 mean distance highway	100 near tram stop at start	
14 speed $P_{95\%}$	44 σ heading change	71 median distance highway	101 near tram stop at end	
15 speed $P_{99\%}$	45 mean turn	72 mean distance on-off ramps	102 near tram stop at start & end	
16 σ speed	46 median turn	73 median distance on-off ramps	103 mean distance bus stop	
17 skew speed	47 n turns + $90°$	74 unique on-off ramps passed	104 median distance bus stop	
18 ratio points + 80km/h	48 n turns + $150°$	75 rel. prox. highway	105 unique bus stops passed	
19 ratio points + 120km/h	49 n turns + $90°$(lag 5 point)	76 rel. prox. on-off ramp	106 rel. prox. bus stop	
20 ratio points - 5km/h	50 n turns + $150°$(lag 5 point)	77 n points near highway	107 n points near bus stop	
21 mean acceleration	51 turns + $90°$p/s	78 n points near on-off ramp	108 near bus stop at start	
22 max acceleration	52 turns + $150°$(lag 5 point) p/s	79 mean distance subway stop	109 near bus stop at end	
23 min acceleration	53 turns + $90°$(lag 5 point) p/s	80 median distance subway stop	110 near tram stop at start & end	
24 acceleration $P_{5\%}$	54 n segments	81 mean distance subway track	111 ratio in urban area	
25 acceleration $P_{25\%}$	55 time at mid point	82 median distance subway track	112 unique urban area passed	
26 acceleration $P_{75\%}$	56 during rush hour (Y/N)	83 unique subway stops passed	113 mean pop. density	
27 acceleration $P_{90\%}$	57 n data points	84 rel. prox. subway stop	114 median pop. density	
28 acceleration $P_{95\%}$		85 rel. prox. subway track	115 mean urbanicity	
29 acceleration $P_{99\%}$		86 n points near subway stop	116 median urbanicity	
30 σ acceleration		87 n points near subway track	117 σ urbanicity	

Table A1: All 127 features used in this research. All variables were calculated twice, once on the data and on the linearly interpolated data in which for every missing second a location estimate was calculated. Rel. prox. stands for relative proximity, which is defined as the number of points near a certain feature divided by all number of location points in a trip.

	statistic	E-bike	bike	Car	metro	bus	scooter	train	tram	User error	walk
1	rel. prox. train tracks*	0 [0; 0]	0 [0; 0]	0 [0; 0]	0 [0; 0.01]	0 [0; 0.01]	0 [0; 0]	**0.1** **[0.03; 0.15]**	0 [0; 0.02]	0 [0; 0]	0 [0; 0]
2	rel. prox. subway stops*	0 [0; 0]	0 [0; 0]	0 [0; 0]	**0.36** **[0.18; 0.45]**	0 [0; 0]	0 [0; 0]	0 [0; 0]	0 [0; 0.19]	0 [0; 0]	0 [0; 0]
3	speed $P_{90\%}$	6.75 [5.41; 7.84]	5.41 [4.56; 6.49]	16.9 [11.3; 26.7]	24.1 [14.2; 50.4]	13.8 [9.66; 20.3]	8.9 [7.46; 10.3]	**39** **[32.7; 56.6]**	12.4 [8.03; 18.9]	2.77 [0; 21.2]	2.27 [1.42; 4.24]
4	rel. prox. subway tracks*	0 [0; 0]	0 [0; 0]	0 [0; 0]	**0.05** **[0.02; 0.26]**	0 [0; 0]	0 [0; 0]	0 [0; 0]	0 [0; 0]	0 [0; 0]	0 [0; 0]
5	speed $P_{95\%}$	7.37 [5.91; 9.82]	5.89 [4.88; 7.63]	19.9 [13; 29.8]	28.3 [18.5; 91.6]	19 [11.5; 24.9]	9.76 [7.84; 12.1]	**45** **[35.4; 90.4]**	15.6 [9.05; 21.2]	6.46 [0; 31.7]	2.8 [1.53; 5.86]
6	rel. prox. tram stops*	0 [0; 0]	0 [0; 0]	0 [0; 0]	0.03 [0; 0.2]	0 [0; 0]	0 [0; 0]	0 [0; 0.01]	**0.64** **[0.5; 0.76]**	0 [0; 0]	0 [0; 0]
7	distance straight line	1173 [441; 2871]	872 [366; 1903]	3235 [860; 11015]	3847 [1310; 10680]	3778 [2333; 5028]	1486 [1133; 3054]	19269 [7018; 37440]	1829 [1011; 3809]	**71.8** **[0; 304]**	88.4 [11.7; 229]
8	speed $P_{5\%}$	0.99 [0.3; 2.82]	1.31 [0.38; 2.58]	1.68 [0.47; 3.87]	0.38 [0.01; 1.15]	0.97 [0.52; 1.8]	2.43 [1.11; 3.83]	1 [0.35; 3.63]	0.47 [0.3; 0.94]	**0** **[0; 0.09]**	0.41 [0.1; 1.23]
9	speed $P_{75\%}$	5.56 [4.61; 6.59]	4.74 [3.91; 5.58]	13.2 [8.65; 21.1]	17.5 [9.46; 29.4]	9.32 [7.12; 15.6]	7.87 [6.8; 9.26]	**35.6** **[21; 38.6]**	8.69 [5.76; 12.5]	1.2 [0; 8.4]	1.81 [1.15; 2.86]
10	median speed	4.66 [3.57; 5.54]	4.02 [3.1; 4.7]	9.34 [5.89; 13.6]	11.1 [2.91; 13.7]	5.47 [4.47; 8.32]	6.78 [5.87; 7.44]	25.3 [9.36; 33.1]	5.17 [3.49; 6.36]	**0.6** **[0; 4.62]**	1.47 [0.83; 2.04]

Table A2: Median [1^{th} quartile; 3^{rd} quartile] of different statistics per trip. In **bold** the most extreme value of the statitsic of that feature. *note: *Mean instead of median presented as summary statistic.*

A.3 Marginal Dependence Plots

Marginal dependence plots can be used understand and visualize the marginal effect of a feature x_s on the class probabilities. Suppose:

$$X = [x_s x_c] \in \mathbb{R}^{n \times p},$$

where x_s is a a vector of the feature we want to know the marginal dependence for and x_c is a matrix with the remaining features. Suppose we estimate some fit $\hat{f}$. Then $\hat{f}_s(x)$, the marginal dependence of $\hat{f}$ at x, is defined as:

$$\hat{f}_s(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}\left(x, x_{c_i}\right)$$

We keep x constant for the feature we want to know the marginal effect for and take the average class probability prediction over all other combinations of other features. Computationally this is a heavy task, but luckily computers are patient. This approach is closely related to what Przemyslaw Biecek calls Ceteris Paribus Plots for classification [118]. For a longer explanation of this approach, but for a simpler model, please refer to this helpful blog post by Hayes [119].

A.4 Literature Review Table

Year & reference	Data source	Modes	Features	Classifier (Precision)
2003 [45]	12 h of GPS logs from 1 author	Walking; Bus; Car	**GPS based features:** Velocity; Sd of the velocity previous 60 seconds **Location based features:** Bus stops; Bus routes **Other features:** -	Graph-based Bayes filter (84%)
2005 [120]	60 trips in 1 city	Walking; Bicycle; Car; Transit	**GPS based features:** Speed; Distance **Location based features:** Bus stops **Other features:** Transport mode previous trip (limits possibilities next trip)	Rule based algorithm (91.7%)
2006 [52]	6 h of GSM data over 3 weeks	Walking; Stationary; Car	**GPS based features:** - **Location based features:** - **Other features:** GSM signal strength; Fluctuation cell towers	Hidden Markov Model (80%)
2006 [55]	GSM traces of 3 collectors for 1 month (420 trips)	Walking; Stationary; Car	**GPS based features:** Distance between points; Variance distances; Distance first to last point **Location based features:** - **Other features:** Common cell towers consecutive measurements; variance in signal strengths	Boosted logistic regression (85%)
2006 [10]	237 segments	Walking; Bicycle; Car; Bus	**GPS based features:** Mean speed; 95^{th} percentile maximum speed; Acceleration **Location based features:** - **Other features:** Data quality	Fuzzy logic rules (91%)
2007 [54]	60 days of data of 1 person wearing a GPS unit	Walking; Stationary; Car; Bus	**GPS based features:** Speed **Location based features:** Car location; Bus stops **Other features:** Labelled transportation goals; Edge transition	Rao-Blackwellized particle filters within hierarchical Markov model (84%)
2008 [46]	29 segments (total 286 min)	Walking; Running; Bicycling; Inline skating; Car	**GPS based features:** Speed **Location based features:** - **Other features:** N steps; Accelerometer data	Discriminant function analysis (93%)
2008 [53]	GSM data by 1 author for 2 days (12.5 hours)	Walking; Stationary; Car	**GPS based features:** - **Location based features:** - **Other features:** Unique cell ID; WiFi beacon signal strength variance; Duration dominant WiFi beacon	Decision Tree (88%)
2008 [72, 121]	GPS logs of 65 people over 10 months	Walking; Bicycle; Car; Bus	**GPS based features:** Heading change rate; Stop Rate; Velocity Change Rate; Segment distance; Speed (max, mean, top 3); Acceleration (max, mean, top 3, variance) **Location based features:** - **Other features:** -	Decision Tree (76.2%)
2009 [73]	60 h of GPS data in one city	Walking; Car; Streetcar; Bus	**GPS based features:** Average speed; Acceleration **Location based features:** - **Other features:** Average number of satellites in view	Multilayer perceptron neural networks (82%)
2009 [15]	GPS logger data of 1104 respondents for 2 weeks	Walking; Bicycle; Car ; Train; Bus/Tram/ Metro; Others	**GPS based features:** Speed (mean, max); Distance; Trip length **Location based features:** Shops; Railway stations; Railways; Schools; Cultural services **Other features:** Dummy variable car ownership; Home and work addresses	Rule based algorithm (70%)

2010 [122]	792 trips in 1 city	Walking; Bicycle; Tram; Bus; Underground; Commuter Train	**GPS based features:** Speed (95^{th} , median, sd); Acceleration (95^{th}); Maximum change of orientation;% Time over 16 km/h; % Time below 2 km/h;% time 4-40km/h; Ratio distance from start to end-point/travelled distance **Location based features:** **Other features:** Max signal loss time greater 60 s; Distance during maximal signal loss;	Support Vector Machine (86.9%)
2010 [123]	4 users collected several hours of traces for each transportation mode	Bus; Metro; Train; Car; Motorcycle; Walking; Bicycle	**GPS based features:** Speed **Location based features: -** **Other features:** Accelerometer data (FFT coefficients of different frequencies); Accuracy	Decision tree (82.1%)
2010 [74]	1554 trips	Walking; Running; Bicycle; Motorbike; Bus; Car; Train	**GPS based features:** Deviation mean speed; Mean speed; Max speed; Mean acceleration; Max acceleration; Distance every 4 Minutes **Location based features:** Distance to the Railway **Other features:** Accuracy (2D and 3D): N used satellites; N viewed satellites: Dummy car ownership; Dummy Bike ownership	Bayesian Belief Network (95.4%)
2010 [56]	5544 samples by 7 volunteers	Bicycle; Bus; Car; Stationary; Subway; Walking	**GPS based features: -** **Location based features: -** **Other features:** Accelerometer data (3 axis and different frequencies)	Decision Tree (70.7%)
2010 [124]	1568 trips	Car; Train; Walking	**GPS based features: -** **Location based features: -** **Other features:** Accelerometer data (SD, max, min, norm, n sign changes for normalized, and cumulated sensor data over 3 axis); Rolling average of these features	Support Vector Machines (97%)
2010 [125]	114 trips	Bus; Car; Walking	**GPS based features:** Max speed; Mean speed; Max Acceleration; Acceleration; Total distance **Location based features: -** **Other features:** Critical points (CP: a minimum set of GPS fixes required to accurately reconstruct the user's path) over total distance; CP over time; Mean distance between CP	Neural network-learning with 300 epochs (91.2%)
2010 [1, 126]	Total of 120 h of data by 6 individuals, carrying 5 of the same phones.	Walking; Stationary; Biking; Running; Motorized	**GPS based features:** Speed **Location based features: -** **Other features:** Accelerometer data(force, mean, variance, energy, and the DFT)	Two-stages Decision Tree & Hidden Markov Model (93.6%)
2011 [7]	Smartphone data of 6 individuals over 3 weeks in 1 city	Walking; Bus; Car; Train; Stationary; Bicycle; Other	**GPS based features:** Speed; Heading; Acceleration **Location based features:** Real time bus locations; Railway locations; Bus stop location. **Other features:** Historical information used bus stops; Parking location vehicle; Accuracy	Random Forest (93.5%)
2011 [127]	125 segments in 1 city	Walking; Bicycle; Motorized	**GPS based features:** Mean speed; Max speed (mean of the top 5 values segment); Mean acceleration (mean of the top 5 values segment); Max acceleration; Heading change **Location based features: -** **Other features: -**	Two stage. first Decision Tree followed by Support Vector Machines (aprox. 90%)
2011 [128]	322 trips in 1 city	Walking; Bicycle; Public Transport; Car; Rail	**GPS based features:** Median speed; 95^{th} of speed: Median acceleration variation **Location based features:** Distance nearest public transport stop **Other features: -**	Fuzzy logic (83%)

2011 [129]	500 sequences split in a normal and congested data set	Bicycle; Car; Walking	**GPS based features:** Speed **Location based features: -** **Other features:** User's historical data	Hidden Markov Model (92.5%)
2012 [110]	81 participants for 2 weeks; measurement once per minute in 1 city	Walking; Bicycle; Car; Train; Bus; Metro	**GPS based features:** Distance; Speed; Acceleration; Heading **Location based features:** Proximity to tube stations **Other features: -**	Support Vector Machines (88%)
2012 [9]	GPS logger data 35 employees for 1 weekday an 28 volunteers for 5 days, 49 days selected for analysis in 1 city	Walking; Bus; Car; Subway; Commuter rail	**GPS based features:** Speed, 85th percentile Speed; Duration; Average speed; 95th percentile of acceleration. **Location based features:** Streets, Bus routes; Bus stops; Subway lines: Subway stations (and entrances): Commuter rail lines; Commuter Rail Stations; Aboveground railroad lines **Other features: -**	Rule based algorithm (82.6%)
2012 [130]	266 h in 1 city	Bus; Car; Bicycle; Walk; Motorcycle; Tram; Train; Metro	**GPS based features:** Speed; Acceleration (median, max, min for both); Max angle **Location based features: -** **Other features:** Accelerometer data (64 features)	Random Subspace Method (71.2%)
2012 [47]	2 h of data at high frequency	Walking; Car; Train	**GPS based features: -** **Location based features: -** **Other features:** Accelerometer data (magnitude, min, max, mean, sd); Gyroscope data (min, max, mean, sd); Historical data of user	Random Forest (97.7%)
2012 [84]	355 h of travel data were collected by 15 volunteers	Bus; Car; Bicycle; Tram; Train; Subway; Walking; Motorcycle	**GPS based features:** Speed; Acceleration; Angular velocity (mean, percentiles, and sd for all) **Location based features: -** **Other features:** Accelerometer data (Power spectrum and sd)	Classifier ensemble followed by a Discrete Hidden Markov Model (76.4%)
2013 [88]	17 million GPS data point	Walking; Bicycle; Motorized; Train; Metro; Aircraft	**GPS based features:** Speed (95th percentile, mean speed, mean moving speed) **Location based features:** Location of water surfaces; Proximity to: railway, tram lines, roads, bus lines, metro lines **Other features: -**	Hierarichal Fuzzy logic (91.6%)
2013 [57]	data from 100 users	Stationary; Walking; Bicycle; Car; Train	**GPS based features:** Distance; Speed (min, max, mean); Acceleration; Velocity change rate; Average speed over segment **Location based features:** Road network; Train network **Other features:** Duration	Random Forest (87.8%)
2013 [58]	80,670 data points	Stationary; Walking; Running; Bicycle; Motorcycle; Bus; Car; Tram; Metro; Light rail	**GPS based features:** Speed; Distance; Mean speed in every 3 min; Max speed; Average time duration of the device not moving in 1 min **Location based features:** Distance to road; Distance to metro line; Distance to light rail **Other features:** Dummy car ownership; Dummy bicycle ownership; Dummy motorcycle ownership; Accuracy; N satellites; Accelerometer data (mean and sd of x, y, and z)	Bayesian Belief Network (91.7%)
2013 [61]	150 h of data from 16 individuals	Stationary; Walking; Bus; Train; Metro; Tram	**GPS based features: -** **Location based features: -** **Other features:** Accelerometer data (range, RMSE, entropy, max coeff, variance, vertical variance, peak volume, peak frequency, intensity, peak length, and variation of peak lengths)	Three-stage hierarchical classification framework (80.1%)

2013 [131]	400 GPS trajectories in different cities	Car; Train; Bus; Bicycle; Walking	**GPS based features:** Speed: Acceleration (max, min, mean, and sd of both); Discretized average speed; Discretized average acceleration; Discretized average horizontal angular speed; Bendiness **Location based features: -** **Other features: -**	C4.5 Decision Trees (92%)
2014 [132]	400 GPS trajectories	Walking; Running; Bicycle; Car	**GPS based features:** Speed; Acceleration; Angle; Ratio direct line to distance travelled (mean, sd, skewness and ApEn of all of these) **Location based features:** **Other features:** Complexity measurement ApEn (a higher ApEn value suggests that the sequential data is a random series, while a smaller value implies less complexity and more regularity); Geometric complexity measures (fractal dimensions)	Support Vector Machine (85.4%)
2014 [59]	8,311 h of data from 224 respondents	Stationary; Walk; Run; Bike; Motorized	**GPS based features: -** **Location based features: -** **Other features:** Accelerometer data (sd, mean, FFT peak, FFT ratio largest and second largest); Magnetometer (sd); Gyroscope data (mean, sd)	Support Vector Machines in two tiers (93.5%)
2014 [60]	50 h of smartphone data, 129 trips from 15 individuals in 2 cities	Stationary; Motorized; Non-motorized; Random movements	**GPS based features: -** **Location based features:** Distance to transit line **Other features:** Accelerometer data; Real time public transport data and schedule (and deviation thereof); Speed of travel for current segment relative to public transport schedule; Time interval between stops	Decision tree with Markov model smoother (92.9%)
2014 [133]	56 users, 6,990 segments 1 week in 1 city	Car; Walking; Bicyle; Train; Tram; Bus; Other	**GPS based features:** Speed (median, mean, sd, 95 percentile) **Location based features:** Distance to public transport stop **Other features:** Start time; Person data on average walk speed and bike speed; N GPS points per second; Accelerometer data (median, 95^{th} percentile)	Random Forest (85.8%)
2014 [134]	10 days of 11 users, total 14,203 GPS points	Car; Train ; Walking; Subway; Bus; Bicycle; Ferry	**GPS based features:** Speed (mean, max, min and rolling window of 5); Distance; **Location based features: -** **Other features:** Accelerometer data (mean, max, sd, and rolling window of 5); N steps (and rolling window of 5); User ID	Random Forest (90.8%)
2014 [135]	50 hours of GPS data in 1 city	Bicycle; Car; Run; streetcar; Walking	**GPS based features:** Speed **Location based features: -** **Other features:** Accelerometer data; Magnetometer data; N satellites in view	Neural Network (74%)
2015 [5]	560 respondents, 18,189 trips in the MoveSmarter app	Walking; Bicycle; Car; Train; Bus/Tram/ Metro; Others	**GPS based features:** Speed **Location based features:** Location of road, rail, water and air infrastructure, Location of public transport stops **Other features:** Sensor data characteristics; Personal trip history	Bayesian probability statistics (75%)
2015 [51]	87 journeys (42.5 h) in the 1 city	Walking; Bicycle; Car; Bus; Train; Tram	**GPS based features:** Acceleration **Location based features:** Geographic data (roads, railways); Public transportation data (routes and schedules) **Other features:** Accelerometer data; Bluetooth data (number of discoverable Bluetooth networks); Wi-Fi emmiters count	Two-phase algorithm: dynamic Bayesian network and smoothing algorithm (75.8%, by time)

2015 [136, 137]	36 trips	Walking; Biking; Car; Bus; Metro	**GPS based features:** Speed; Acceleration **Location based features:** OSM network data (roads and public transport data) **Other features:** Accelerometer data; Bluetooth data (number of discoverable Bluetooth networks);	Probabilistic multi-modal map-matching (no ground truth test)
2015 [75]	172 trips in 1 city	Walking; Bus; Bicycle; Car; Subway; Train	**GPS based features:** - **Location based features:** - **Other features:** Accelerometer data (mean, standard deviation, skewness and kurtosis); Dummy variable trip in weekend or a weekday	Binomial logistic regression (NA)
2015 [2]	25 h total phone data from 10 participants.	Walking; Running; Bicycle; Car; Bus	**GPS based features:** - **Location based features:** - **Other features:** Accelerometer data (mean speed, mean acceleration max accelerometer, heading change rate, spectral entropy acceleration; range acceleration, variance acceleration); Gyroscope sensor information (min, max; mean); Rotation vector	Random Forest (95.1%)
2015 [66]	4,685 segments from 202 respondents in 1 city	Walking; Bicycle; E-bike; Bus; Car	**GPS based features:** Low speed rate (ratio of points with a speed of less than 1 m/s); Mean speed; 95 th percentile speed, Mean acceleration; Distance; Heading change rate **Location based features:** - **Other features:** -	Bayesian network with K2 estimator (92.7%)
2015 [138]	49,293 segments collected by 35 individuals in 1 city	Bus; Car; Walking	**GPS based features:** Bearing; Speed; Change Speed; Acceleration; Change acceleration (for all mean, min, max, and sd) **Location based features:** - **Other features:** Accuracy (mean, min, max, sd)	Random Forest with 160 trees (96.9%)
2015 [139]	1,654 segments in 1 city	Walking; Bicycle; Car ; Bus	**GPS based features:** Low speed rate (ratio of points with a speed of less than 1 m/s); Mean speed; Median speed; 95th percentile speed, Mean acceleration; Distance **Location based features:** - **Other features:** -	Particle swarm optimization (94.4%)
2015 [140]	3,105 trips in 3 cities	walking; Bicycle; Car; Train	**GPS based features:** - **Location based features:** - **Other features:** Accelerometer data (moving average of max, min and mean in 3 directions and resultant)	Random Forest (99.8%)
2015 [78]	644 person days of travel in 1 city	Walking; Bicycle; Bus; Rail; Car	**GPS based features:** Speed (median, percentiles, max); Acceleration (median, percentiles, max) **Location based features:** Rail lines; Bus lines **Other features:** -	Three-step fuzzy logic (90.6%)
2016 [141]	10,890 segments in 1 city from the Geolife Project	Bicycle; Bus; Car; Subway; Walking	**GPS based features:** Speed; Distance; Acceleration; Heading **Location based features:** Bus line closeness; Subway stop closeness **Other features:** -	Deep neural network with Stacked Autoencoders (93.6%)
2016 [142]	24 h of GPS data total by 2 individuals	Walking; Running; Bicycle; Car	**GPS based features:** - **Location based features:** - **Other features:** Accelerometer data; Gyroscope data; Magnetometer data (min, max, mean, IQR over different frequencies for all)	Random Forest (97%)

2016 [62]	476 h of data from the HTC company	Stationary; Walking; Running; Bicycle; Motorized	**GPS based features:** - **Location based features:** - **Other features:** Accelerometer data (mean in 3 directions, sd, instant changes, max FFT, The ratio highest and second-highest FFT,); Magnetometer data (mean, sd, instant changes); Gyroscope data (mean, sd)	Support Vector Machines (86.9%)
2016 [48]	1,554 trips from 8 individuals in 1 city	Stationary; Train; Walking; Bicycle; Car; Bus; Motorbike; Running; Tram; Metro	**GPS based features:** Speed (mean, max, sd); Acceleration (mean, max); Distance **Location based features:** Distance to road line; Distance to tram line; Distance to metro line **Other features:** N used satellites; N viewed satellites; GPS fix type; Accuracy (2F and 3D); Dummy car ownership; dummy bike ownership; dummy motorbike ownership	Bayesian network (99.4%)
2016 [63]	12 people for 6 days in 1 city	Bicycle; Car; Run; Stationary; Walk	**GPS based features:** Speed **Location based features:** - **Other features:** Accelerometer data (mean, max, percentile, peaks over 3 axis)	Random Forest with 500 trees (93.8%)
2016 [69]	196 min of phone data	Bus; Subway; Car; Bicycle; Walking; Jogging	**GPS based features:** - **Location based features:** - **Other features:** Accelerometer data (max, min, mean, sd over 3 axis); Gravity sensor; Barometer (estimate of current weather); Magnetometer data (mean, sd, instant changes)	Support Vector Machines (97.1%)
2016 [143]	1,307 segments from 26 users in 1 city	Car; Train; Walking; Subway; Bus; Bicycle; Ferry	**GPS based features:** Speed (mean, max, min and rolling window of 5); Distance; **Location based features:** - **Other features:** Accelerometer data (mean, max, sd, and rolling window of 5); N steps (and rolling window of 5);	Random Forest (80.1, by traditional method%)
2016 [70]	1,323 segments from 75 participants	Walking; Bicycle; Bus	**GPS based features:** Speed (max, 95^{th} percentile, median, mean, sd, skewness, kurtosis); Ratio of points less than 2 m/s, between 2 and 8 m/s, between 8 and 15 m/s and above 15 m/s; Acceleration (max, 95^{th} percentile, median, mean, sd, skewness, kurtosis); Mean deceleration; Max change heading; Mean change heading; Total distance; Ratio direct line to distance travelled; N stops **Location based features:** **Other features:** Accuracy; Ratio invalid GPS points; Duration signal loss, distance traveled in signal loss	Nested logit model (98.4%)
2016 [50]	106 trips in 1 city split in segments of 60 s	Walking; Bus; Train; Tram	**GPS based features:** Speed (mean, 95^{th} percentile) **Location based features:** Mean proximity to bus network; Mean proximity to train network; Mean proximity to tram network **Other features:** -	Artificial neural network (ANN) combined with Sugeno-type fuzzy logic (83%)
2016 [144]	9,043 from the GeoLife project; 30 users' trajectories for 20 days	walking; Bicycle; Car; Bus; Taxi; Motorcycle; Train	**GPS based features:** Speed; Acceleration (13 features in total) **Location based features:** - **Other features:** Raw GPS data represented as 2D image data, followed by Deep Feature Extraction	Deep Neural Network (75.5%)

2017 [71]	7,985 labelled trajectories from the GeoLife project	Walk; Bus/Taxi; Bicycle; Car; Subway; Train	**GPS based features:** Speed; Acceleration; Turn angle; Sinuosity (for all these three the mean, sd, mode, top 3 values, min 3 values, range, percentiles, interquartile range, skewness, kurtosis, coefficient of variation, and autocorrelation); Heading change rate; Stop rate; Velocity change rate; Distance **Location based features: -** **Other features: -**	XGBoost (90.8%)
2017 [65]	2 disabled participants in 1 city	Walking; Rail; Stationary; Bus	**GPS based features:** Speed **Location based features:** Bus stops; Rail stations; Index of Multiple Deprivation (IMD, proxy for level of attractiveness area) **Other features:** Socio-demographic factors and personal characteristics of travellers (such as age and disability)	Dynamic Bayesian network (90%)
2017 [80]	9,205 GPS segments from of buses in 1 city, and of walking, biking and driving from OSM.	Walking; Bicycle; Car; Bus	**GPS based features:** Speed; Acceleration for both (min, max, median, mean, var, skewness and kurtosis); Auto- and cross correlations of these features **Location based features: -** **Other features: -**	Random Forest (88.7%)
2017 [49]	96.59 h of GPS data and 98.62 h of accelerometer data from 6 students	Walking; Bicycle; Car; Bus; Rail	**GPS based features:** Speed (mean change, mean, median, variance, min, max, interquartile range, 20^{th} percentile and 80^{th} percentile) **Location based features: -** **Other features:** Accelerometer data (mean change, mean, median, variance, min, max, interquartile range, 20^{th} percentile and 80^{th} percentile, correlation, sum of the first 6 components of the Fourier transformation)	Random Forest over movelets (96.8%)
2017 [145]	GPS logs of 65 people over 10 months	Not mentioned	**GPS based features:** Heading change rate; Stop Rate; Velocity Change Rate; Segment distance; Max speed; Max acceleration; Mean speed; Variance acceleration; **Location based features: -** **Other features: -**	Deep neural network (74.1%)
2018 [40]	GPS trajectories collected by 69 users in the GeoLife project	Walking; Train; Car; Bicycle; Bus	**GPS based features:** Speed, Acceleration, Jerk (Difference acceleration); Bearing **Location based features: -** **Other features:** Raw GPS data	Convolutional Neural Network (84.8%)
2018 [64]	79 h of data in 1 city	Bus; Car; Ferry; Metro; Train; Tram; Walking; Stationary	**GPS based features: -** **Location based features: -** **Other features:** Accelerometer; Gyroscope; Magnetometer data (max, min, max increase, max decrease, min increase, min decrease, range, mean, median, mode, quadratic mean, sd); Mean instant exchange; Quadratic mean instant exchange; Covariance; Freq sd above median; Freq sd above mean; Freq sd below median; Freq sd below mean; Freq sd between mean; Freq sd between median; Max consecutive above median; Max consecutive above mean; Max consecutive below median; Max consecutive sd between median; Max consecutive below mean; Max consecutive sd between mean	segment-based Healing algorithm (94.5%)
2018 [67]	2,740 segments from 125 respondents in 1 city	Subway; Walking; Bicycle; E-bike; Bus; Car	**GPS based features:** Distance; Speed (mean; median; 75^{th} percentile speed, 95^{th} percentile speed); mean change in orientation; Skewness of speed distribution **Location based features: -** **Other features:** Dummy having a bus card; N household bicycles; N household E-bikes; N household cars	Random Forest combined with rule-based system (93.1%)

2018 [146]	85 h of data in 1 city	Walking; Bus; Tram; Train	**GPS based features:** Speed (mean, max) **Location based features:** Bus network Train network; Tram network **Other features:**	Fuzzy Logic (81%)
2018 [147]	44,831 h of data from 100 users	Walking; Car; Bus; Bicycle; Train; Plane	**GPS based features:** Speed (85^{th} percentile, median, min, max, mean, covariance, expectation, change rate); Acceleration (85^{th} percentile, median, min, max, mean, covariance, expectation, change rate); Stop rate; Heading change **Location based features: -** **Other features:** time of day (morning, afternoon, or evening)	Random Forest (82.3%)
2018 [76, 148]	GPS logs of 65 people over 10 months	Walking; Bicycle; Bus; Car; Train	**GPS based features:** Distance, Speed; Acceleration Jerk; Bearing; Bearing change rate (mean, max, min, median, sd for all) **Location based features: -** **Other features: -**	Random Forest with noise removal algorithm (93.3%)
2018 [149]	1,053 days of data from 41 respondents, 1,906 trips selected for use	Walking; Bus/Tram; Train; Car	**GPS based features:** Distance; Speed (mean, median, max); Acceleration (mean, median, max); Mean angle, Median angle **Location based features:** Public transport schedule **Other features:** N points; Past data of user; Duration trip; Mean distance between two consecutive points	Random Forest (87%)
2019 [150]	41,516 trips from 314 respondents in 1 city using GPS trackers	Walking; Bus; Bicycle; Car; Train	**GPS based features:** Speed (mean, max, sd, 85^{th} percentile); Acceleration (mean, max); Distance **Location based features:** N of bus stops detected in the trip and whether the route is consistent with a bus line; Distance to road line; Distance to tram line; Distance to metro line: percentage of points near to 4 different lines **Other features:** N used satellites; N viewed satellites; GPS fix type; Accuracy (2F and 3D); Dummy car ownership; Dummy bike ownership; Dummy motorbike ownership	Bayesian network (88.1%)

Table A3: Literature review table: Only projects that use either smart phones as a measuring device, GPS data or both are included. Equally, only studies that aim to distinguish transport modes regularly found in travel studies are included, studies such as [151], that only distinguish between trucks and cars, are excluded. Work that only used car mounted GPS receivers was also excluded. If multiple classifiers were tested, only the best preforming one is reported. If only precision per transport mode was reported, the average was calculated. Please note also, that different authors use vastly different strategies and methods to calculate precision, so comparison in hard.